Human Genetics (2024) 143:1481–1498
https://doi.org/10.1007/s00439-024-02707-9

ORIGINAL INVESTIGATION

Germline copy number variants and endometrial cancer risk

Cassie E. Stylianou[1] · George A. R. Wiggins[1] · Vanessa L. Lau[1] · Joe Dennis[2] · Andrew N. Shelling[3] · Michelle Wilson[4] · Peter Sykes[5] · Frederic Amant[6,7] · Daniela Annibali[7] · Wout De Wispelaere[7] · Douglas F. Easton[2,8] · Peter A. Fasching[9] · Dylan M. Glubb[10] · Ellen L. Goode[11] · Diether Lambrechts[12,13] · Paul D. P. Pharoah[14] · Rodney J. Scott[15,16,17] · Emma Tham[18,19] · Ian Tomlinson[20] · Manjeet K. Bolla[2] · Fergus J. Couch[21] · Kamila Czene[22] · Thilo Dörk[23] · Alison M. Dunning[8] · Olivia Fletcher[24] · Montserrat García-Closas[25] · Reiner Hoppe[26,27] · ABCTB Investigators[28] · Helena Jernström[29] · Rudolf Kaaks[30] · Kyriaki Michailidou[2,31] · Nadia Obi[32,33] · Melissa C. Southey[34,35,36] · Jennifer Stone[37,38] · Qin Wang[2] · Amanda B. Spurdle[39] · Tracy A. O'Mara[10] · John Pearson[40] · Logan C. Walker[1]

Received: 24 July 2024 / Accepted: 30 September 2024 / Published online: 4 November 2024

Abstract

Known risk loci for endometrial cancer explain approximately one third of familial endometrial cancer. However, the association of germline copy number variants (CNVs) with endometrial cancer risk remains relatively unknown. We conducted a genome-wide analysis of rare CNVs overlapping gene regions in 4115 endometrial cancer cases and 17,818 controls to identify functionally relevant variants associated with disease. We identified a 1.22-fold greater number of CNVs in DNA samples from cases compared to DNA samples from controls ($p = 4.4 \times 10^{-63}$). Under three models of putative CNV impact (deletion, duplication, and loss of function), genome-wide association studies identified 141 candidate gene loci associated ($p < 0.01$) with endometrial cancer risk. Pathway analysis of the candidate loci revealed an enrichment of genes involved in the 16p11.2 proximal deletion syndrome, driven by a large recurrent deletion (chr16:29,595,483-30,159,693) identified in 0.15% of endometrial cancer cases and 0.02% of control participants. Together, these data provide evidence that rare copy number variants have a role in endometrial cancer susceptibility and that the proximal 16p11.2 BP4-BP5 region contains 25 candidate risk gene(s) that warrant further analysis to better understand their role in human disease.

Introduction

Endometrial cancer is it the most commonly diagnosed gynaecological cancer in developed countries (Rodríguez-Palacios et al. 2022). The incidence of endometrial cancer has been increasing, and a key contributor to this trend is the rising prevalence of obesity, a major risk factor for this disease. Other risk factors include reproductive risk factors such as early menarche, late menopause and nulliparity, exogenous oestrogen use, and a family history of endometrial or colorectal cancer (Lortet-Tieulent et al. 2018). While much progress has been made to understand the biology of endometrial cancer, the genetic risk factors underlying this disease have not been fully elucidated.

Genetic risk factors for endometrial cancer include inherited pathogenic variants DNA mismatch repair (MMR) genes associated with Lynch Syndrome (*MLH1, MSH2*, *MSH6* and *PMS2*) and the tumour suppressor *PTEN*. Genome-wide technologies, such as single nucleotide polymorphisms (SNP)-arrays have identified common risk loci associated with endometrial cancer that confer levels of risk (odds ratio [OR] < 2), and in aggregate explain less than a third of the estimated familial relative risk for endometrial cancer (Chen et al. 2016; O'Mara et al. 2018; Wang et al. 2022).

Copy number variants (CNVs) are a form of structural variation that are pervasive in the human genome and can disrupt gene function by altering gene dosage, coding sequence or regulation. The de novo mutation rate of CNVs is several orders of magnitude higher than the mutation rate of single nucleotide variants (Zhang et al. 2009). However, CNVs are typically rare which is consistent with the

Cassie E. Stylianou, George A. R. Wiggins have contributed equally to this work.

The members of the ABCTB Investigators are mentioned in Acknowledgements section.

Extended author information available on the last page of the article

hypothesis that CNVs can be pathogenic and therefore often under negative selection. A recent study of CNVs in 100,000 individuals with European ancestry showed that > 98.5% CNV variants had a minor allele frequency < 0.01 (Li et al. 2020).

Rare pathogenic CNVs have previously been identified in cancer susceptibility genes, including known endometrial cancer syndrome genes (Truty et al. 2019). In a *MSH2*-associated Lynch syndrome cohort (n = 83), 11% of pathogenic variants identified in *MSH2* were CNVs (Romero et al. 2013). Similarly, single to multi-exon deletions make up 22% of pathogenic variants in *PMS2*, 21% of pathogenic variants in *MSH2* and *MLH1* and 4% of pathogenic variants in *MSH6* (Lagerstedt-Robinson et al. 2016). In a genome-wide analysis of 1209 endometrial cancer cases and 528 cancer-unaffected female controls, we previously reported that rare deletions of likely functional genomic regions (e.g. exons and CpG islands) were more frequent in cases compared to controls (Moir-Meyer et al. 2015). These results implicated rare germline deletions of functional and regulatory genomic regions as mechanisms for conferring risk of endometrial cancer.

To identify endometrial cancer CNV risk loci, we performed a gene-centric genome-wide association study (GWAS) using the OncoArray single nucleotide polymorphism (SNP) array on a large cohort (n = 21,933) of endometrial cancer cases and healthy controls with European ancestry. Additionally, we conducted analysis of global CNV burden in endometrial cancer cases compared to controls.

Methods

Study cohort and genotyping

The study cohort was comprised of female individuals from 28 studies, with cases sourced via the Endometrial Cancer Association Consortium (ECAC) and healthy female controls from the Breast Cancer Association Consortium (Supplementary Table S1). The characteristics of the cohorts have been previously described (O'Mara et al. 2018). DNA samples derived from whole blood were genotyped on the Infinium OncoArray-500K Beadchip (Illumina) across five genotyping facilities, all participants were of European descent. The OncoArray consists of 533,631 probes, half of which were selected from the HumanCore (Illumina) backbone and the other half placed in regions previously associated with cancer risk (Amos et al. 2017).

CNV calling

CNVs were called using CamCNV, a method designed to confidently call rare (MAF < 3%) CNVs with fewer probes and higher confidence (Dennis et al. 2021). Quality control was performed for samples and CNVs (Supplementary Table S2). Briefly, for each sample a derivative log ratio spread (DLRS) figure was calculated as the average variance in Log R Ratio (LRR) intensities of neighbouring probes by genome position over the whole genomes (Cooper et al. 2015). Samples with a DLRS Fig. 3.5 SD above the DLRS study mean (DLRS = 0.2) were removed. Principle component adjustment (PCA) of the LRR intensities at each probe was then performed to reduce batch effects in probe intensity and adjust for variation in hybridisation intensity (genomic waves) (Diskin et al. 2008). Following PCA, a second DLRS sample exclusion was applied, again removing samples with a DLRS 3.5 SD above post PCA-adjusted sample mean of DLRS = 0.1. Samples with excessive heterogeneity (4.89 SD from the study mean), or those with sex chromosome abnormalities were also excluded from study (Michailidou et al. 2017). Prior to CNV calling, probes with data that failed to be clustered by Illumina Gentrain algorithm (< 0.15), low intensity probes (mean intensity < 0.2) or any with high LRR variance (two SD above the mean variance of all probes) were removed. Additionally, CNVs predicted within immune-related loci (Immunoglobin heavy chain, T-cell receptor and major histocompatibility complex) or near centromeres and telomeres were also excluded. Only CNVs called using 3–200 probes were retained. Previous published thresholds of excess germline CNV count in human blood ranged between 30 and 200 CNVs (Aguirre et al. 2019; Macé et al. 2016). We adopted a lower threshold and excluded samples predicted to carry n ≥ 50 (Supplementary Table S2). The final analysis dataset included data for 4,115 endometrial cancer cases (371 removed) and 17,818 controls (1,073 removed).

CNV annotation

CNVs were annotated for overlap with protein coding genes and exons sourced using biomaRt and EnsDB (Hsapiens.v75) R packages, with the largest Ensembl transcript used to define gene boundaries (Durinck et al. 2009; Rainer 2017). All genomic features were restricted to chromosomes 1–23/X, and any elements mapping to alternative chromosomes (i.e., sequence scaffolds or mitochondrial chromosomes) were excluded from analysis. Genomic coordinates were based on the GRCh37/hg19 genome build. In situations where genomic data was in an alternative genome build, the UCSC LiftOver tool was used for conversion to GRCH37/hg19 (https://genome.ucsc.edu/cgi-bin/hgLiftOver). All CNVs were assessed for overlap (≥ 1 bp) with regions of interest in R using the GenomicRanges package (V1.4) (Lawrence et al. 2013).

CNV burden

CNV burden was estimated between endometrial cancer cases and controls for: total number of CNVs, the number of genic CNVs, the number of exonic CNVs and number of intergenic CNVs, respectively. Each burden analysis was repeated for CNV deletions, CNV duplications and all CNVs. Statistical significance of differences in CNV burden between cases and controls were determined by a two-sided Student's *t*-test, *p*-values < 0.05 were considered statistically significant.

Copy number variation (CNV)-GWAS

Associations between CNVs and endometrial cancer were assessed by performing a gene-specific test using gene boundaries to define regions of interest. Case and control CNV overlap frequency was determined for each gene region and association was tested by fitting a binomial logistic regression model. Given the varying modes of effects from copy number gain and copy number loss, deletions and duplications were tested independently. Additionally, models were estimated on putative loss of function. A CNV was included in the loss of function GWAS if it was either predicted as a deletion or a duplication that partially overlapped a gene region. A genome-wide significance threshold was calculated for each GWAS conducted: this was represented as 0.05/6014, 0.05/8377 and 0.05/8613 for deletion-only, duplication-only and loss of function respectively.

Additionally, to explicitly model the level of evidence for genes already associated with endometrial cancer, the Bayesian false discovery probability (BFDP) approach was applied (Wakefield 2008) with the prior probabilities assigned at 0.5, for the genes associated with Lynch syndrome, 0.2 for genes with previous associations and 0.05 for genes with little to no prior evidence (Supplementary Tables S5-S7). An upper bound of 8.0 was applied on the odds ratio for any association, all parameters were chosen to reflect the rare nature and large effect of the tested CNV. Lastly, associations at $p < 0.01$ were considered as candidate associations.

Overlap with previously identified risk SNPs

SNPs associated with disease risk were directly downloaded from the NHGRI-EBI GWAS Catalog (accessed Jan 2024) for the following traits; endometrial cancer (MONDO_0011962, n = 84), Type 2 Diabetes (MONDO_0005148, n = 3516) and Obesity (EFO_0001073, n = 297). SNP associated with these traits were expanded to include any variant in linkage disequilibrium (LD, $R^2 > 0.8$) in the 'EUR' population from 1000 genomes. Germline CNVs overlapping candidate endometrial cancer risk genes were first assessed for direct overlap with SNP, and the candidate gene list was compared to GWAS mapped gene(s).

Pathway analysis

Over-representation analysis was performed in R v3.14 using the gProfiler2 package by applying a hypergeometric test to assess enrichment, all results presented are Bonferroni corrected (Kolberg et al. 2020). To allow for variation among candidate endometrial cancer risk genes ($p < 0.01$) derived from different GWAS, top hits from each GWAS were assessed independently. Additionally, FUMA-GWAS was used to test if candidate genes were enriched for genes reported in the GWAS (Watanabe et al. 2017).

Expression in endometrial tissue and dosage sensitivity

Expression of candidate genes was assessed in normal and tumour tissue using publicly available data. The R packages hpar and ExperimentHub were used to retrieve RNA levels (Transcripts per million (TPM)) directly from the Human Protein Atlas repository (L and Martin 2022; Morgan and Shepherd 2022). Genes were grouped into expression categories using thresholds defined by Expression Atlas (Papatheodorou et al. 2018). Dosage sensitivities of candidate genes were assessed using mRNA expression data and putative copy number of genes from The Cancer Genome Atlas- Uterine Corpus Endometrial Carcinoma (TCGA-UCEC) dataset using the cBioPortalData package from R (Bonneville et al. 2017; Ramos et al. 2020). Candidate risk genes were deemed dosage sensitive if there was a positive, significant ($P < 0.0001$) relationship between copy number and expression.

CNV validation

Accessible whole-blood DNA samples from the study cohort were used to validate 17 putative CNV regions. CNV validation was carried out using NanoString nCounter (NanoString Technologies, Inc) following the manufacturer's protocol. Custom Nanostring probes for CNV regions are listed in Supplementary Table S3. Where possible, three independent probe pairs were designed for each CNV unless the region was too small to accommodate, in which case two probes were used. nSolver 4.0 analysis software was used to perform quality control on raw counts and normalised to a set of invariant control probe pairs. CNVs were partitioned by carrier status and count ratios were calculated to call CNV status.

Results

Identification of CNVs in the study cohort

A total of 63,349 rare deletions and 48,555 rare duplications were identified across the 21,933 study participants, of which 46,234 were unique (25,047 deletions and 21,187 duplications). On average, duplications were 2.4 times larger than deletions (mean length 99 kilobases (kb) for duplications vs 41 kb for deletions). In total, 10,637 unique protein coding genes were predicted to be encompased by 24,390 unique CNVs, with 40.7% of deletions and 52.7% of duplications predicted to overlap at least one gene region (Supplementary Table S4). On average, we identified 5.10 CNVs per sample (range = 0–47) and 2.34 genic CNVs per sample (range = 0–47), with 96.3% of samples estimated to carry at least one CNV. The highest minor allele frequency for CNVs called with CamCNV was 2.2%. The majority of CNVs (79% of deletions and 81% of duplications) identified were only identified in a single sample (allele frequency = 0.0045%) highlighting the uniqueness of these events.

Explicitly modelling prior knowledge lifted *MSH6* to significance however none of the 41 genes with prior probability 0.2 were significant in either frequentist or Bayesian analysis. Bayesian analysis showed significant evidence for 2 additional genes, *VWA1* and *ATAD3C* at a BFDP of 0.0074 however, these both had an adjusted P value of 0.079. Given the convergence of the Bayesian and frequentist analysis, subsequent analysis proceeded with the genes identified in the frequentist analysis; further details are available in Supplementary Tables 5–7.

Comparison of global CNV burden between endometrial *cancer* cases and controls

The impact of an individual's CNV burden on endometrial cancer risk was estimated for all CNVs, deletions-only and duplications-only. On average, the total number of CNVs in endometrial cancer cases was 1.22-fold greater than controls ($p = 4.4 \times 10^{-63}$) and was consistent for CNVs predicted as deletions (fold change [FC] = 1.16, $p = 1.2 \times 10^{-25}$) and duplications (Table 1, FC = 1.31, $p = 1.5 \times 10^{-50}$). We further investigated the genomic location of CNVs and estimated the burden of CNVs overlapping genes and exons or in intergenic regions (Table 1). Compared to the burden analysis of total CNVs, the estimated burden was greater for CNVs overlapping genes (FC = 1.30, $p = 2.1 \times 10^{-50}$) and exons (FC = 1.31, $p = 7.1 \times 10^{-48}$). In contrast, intergenic CNVs (FC = 1.16, $p = 1.9 \times 10^{-32}$) displayed reduced burden compared to total CNVs (Table 1).

Rare CNV association analysis

To identify specific CNVs associated with endometrial cancer risk, we conducted GWASs for three different association models: a deletion-only, a duplication-only and a loss of function models (all genic deletions and any partial gene duplications) (Supplementary Tables S5-7). We performed

Table 1 Global burden association analysis of rare CNVs

Genomic feature	Mean frequency		Mean difference	95% CI	p-value[a]	Fold change
	Cases (n = 4115)	Controls (n = 17,818)				
CNVs						
All	5.99	4.90	1.10	0.97–1.22	4.3×10^{-63}	1.22
Deletions	3.26	2.80	0.45	0.37–0.54	1.1×10^{-25}	1.16
Duplications	2.74	2.09	0.64	0.56–0.73	1.4×10^{-50}	1.31
Genic CNVs						
All	2.89	2.22	0.67	0.59–0.76	2.1×10^{-50}	1.3
Deletions	1.40	1.12	0.28	0.22–0.33	2.2×10^{-20}	1.25
Duplications	1.49	1.09	0.4	0.34–0.46	2.1×10^{-38}	1.36
Exonic CNVs						
All	2.51	1.92	0.59	0.51–0.67	7.1×10^{-48}	1.31
Deletions	1.19	0.94	0.25	0.2–0.31	7.0×10^{-21}	1.27
Duplications	1.32	0.98	0.34	0.28–0.39	4.1×10^{-34}	1.34
Intergenic CNVs						
All	3.10	2.68	0.42	0.35–0.49	1.9×10^{-32}	1.16
Deletions	1.86	1.68	0.18	0.13–0.23	7.0×10^{-12}	1.11
Duplications	1.25	1.00	0.25	0.2–0.29	2.6×10^{-28}	1.25

[a]Student's two-sample *t test*

gene-centric tests under the assumption that non-overlapping CNVs impacting the same gene locus may have similar effects. The deletion-only model identified a total of 59 gene loci associated ($p < 0.01$) with endometrial cancer, including two loci (*SLCO1B3* and *SALL3*) that met the Bonferroni genome-wide threshold of significance (Fig. 1; Supplementary Table S5). The analysis of duplication variants identified a total of 58 risk-associated loci ($p < 0.01$), including three loci (*SLC6A3, ANTXRL* and *KIF25*) that met genome-wide significance (Fig. 1; Supplementary Table S6). The

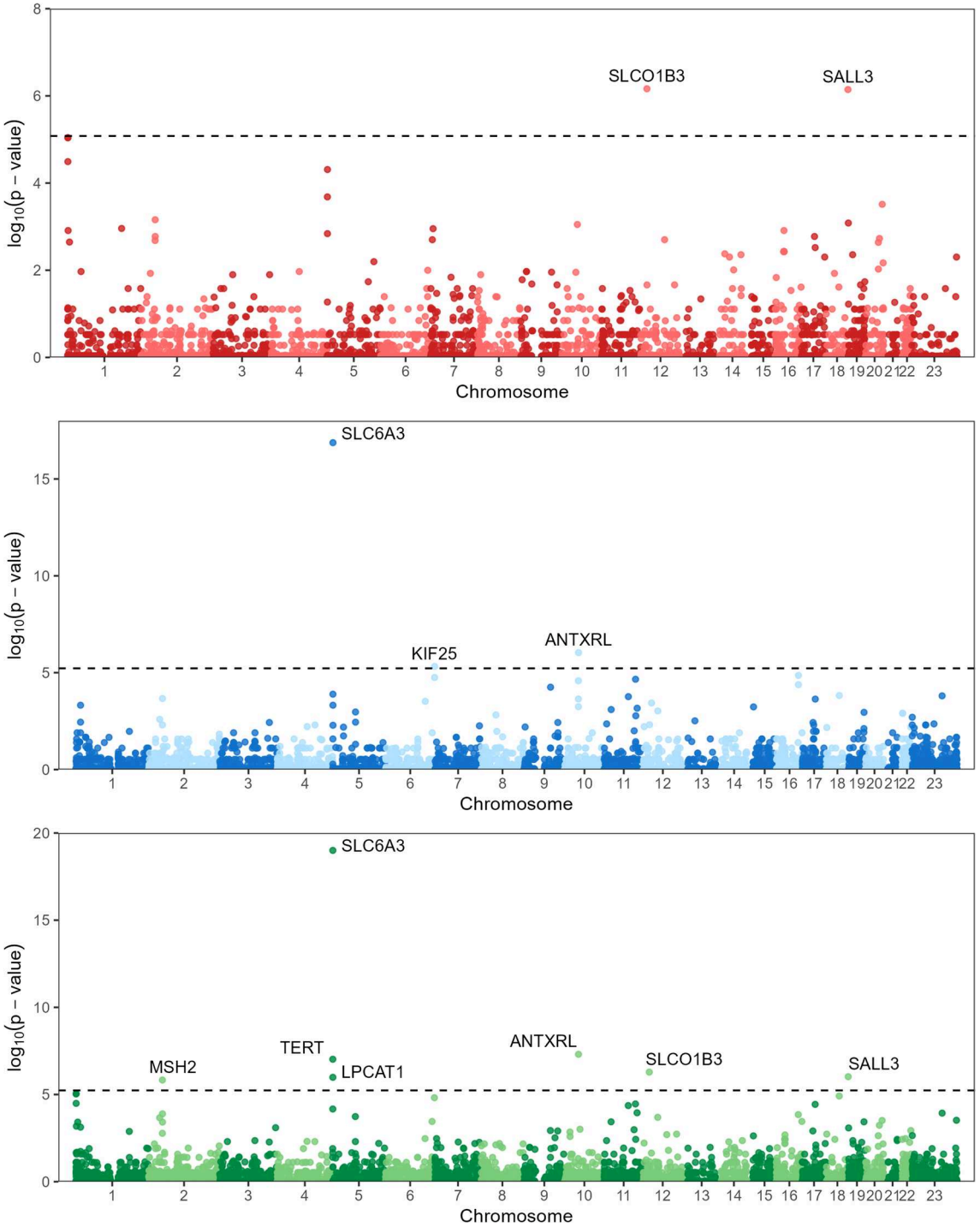

Fig. 1 Manhattan plots for CNV-GWAS of 4,115 endometrial cancer cases and 17,818 controls. Genome-wide *p*-values for deletion-only (top), duplication-only (middle) and loss of function CNVs (bottom). Dashed line indicates Bonferroni derived genome-wide significance thresholds at 8.31×10^{-6} for deletion-only, 5.97×10^{-6} for duplication-only and 5.81×10^{-6} for loss of function

analysis of loss-of-function variants identified a total of 116 endometrial cancer risk loci ($p < 0.01$), including seven loci (*SLC6A3, ANTXRL, TERT, SLCO1B3, SALL3, LPCAT1* and *MSH2*) that met genome-wide significance (Fig. 1; Supplementary Table S7). Candidate genes ($p < 0.01$) identified by the loss of function model, which includes all deletion variants, captured 93% (55/59) and 64% (37/58) of the candidate genes identified in the deletion-only and duplication-only models, respectively (Supplementary Fig. 1). Additionally, 28 candidate genes were exclusively identified by the loss-of-function model. Only four genes (*LPCAT1, TERT, MSH2* and *SLC6A3*) were identified as candidate risk loci across all three genome-wide association analyses (Supplementary Fig. 1). For each of the genes, all duplications partially overlapped the respective gene boundaries suggesting a shared loss-of-function mechanism with deletions. In total 141 candidate genes (1,525 unique CNVs, $p < 0.01$) were identified across the three association models, including 5 genes (190 unique CNVs) that met genome-wide significance.

Associations of candidate CNV risk loci at established risk associated SNPs

We next sought to assess if any of the 1,525 risk-associated candidate CNVs had direct overlap with previously identified GWAS risk SNPs (n = 84) for endometrial cancer risk (Type 2 diabetes [n = 3,516] and obesity [n = 297]). Seven cases and three controls had CNVs that colocalised with two endometrial cancer risk SNPs (rs11263763 and rs11651052) located in intron 1 of *HNF1B* (Fig. 2). Furthermore, CNVs overlapping *HNF1B* were more than six times as frequent in endometrial cancer cases compared to controls (OR = 7.59, 95% CI = 2.29–28.99, $p = 0.001$, Supplementary Table S8). For the traits associated with endometrial cancer risk, 33 Type 2 diabetes-associated and no obesity-associated SNPs were overlapped by at least one candidate endometrial cancer CNV, respectively. Of the 141 candidate gene regions assessed, 50 had at least one CNV overlapping a previously identified risk-SNP. This was driven by a large, multigenic deletion that mapped to the proximal 16p11.2 recurrent breakpoints (BP) 4 and 5 (Supplementary Fig. 2A) that overlaps two Type 2 diabetes risk SNPs (rs8054556 and rs11642340) and 25 risk-associated candidate genes. An additional six lead SNPs had at least one variant in LD ($R^2 > 0.8$) that overlapped a risk-associated candidate CNVs. This included three lead SNPs associated with endometrial cancer (rs11263761, rs2278868 and rs882380) and three associated with Type 2 diabetes (rs11651755, rs4430796 and rs8010382). No SNPs associated with obesity (EFO_0001073) from the GWAS Catalog (MacArthur et al. 2017) were found to map to the CNV risk loci.

Validation of putative rare CNVs

We attempted to validate 17 CNVs (localised to 12 genes), selected from a range of allele frequencies (0.005%-1.49%), in 11 samples using NanoString technology. In total, 12 risk-associated candidate genes were assessed with eight (80%, 8/10) deletions and one (50%, 1/2) duplication validated (Table 2). These data support the reported predictive accuracy of the CamCNV tool (Dennis et al. 2021). This included, validation of three deletions overlapping the known endometrial cancer risk genes (*MSH2* and *PMS2*) in three cases. These three validated CNVs (chr2:47,637,511–47,673,515, chr2:47,639,553–47,639,699, chr7:6,029,431–6029586) overlapped CNVs predicted in a further 26 samples (20 cases, 6 controls). In total, there were 73 CNVs (46 deletions and 27 duplications) overlapping *MLH1, MSH2, MSH6* and *PMS2* in 86 samples (1.28% of cases and 0.19% of controls).

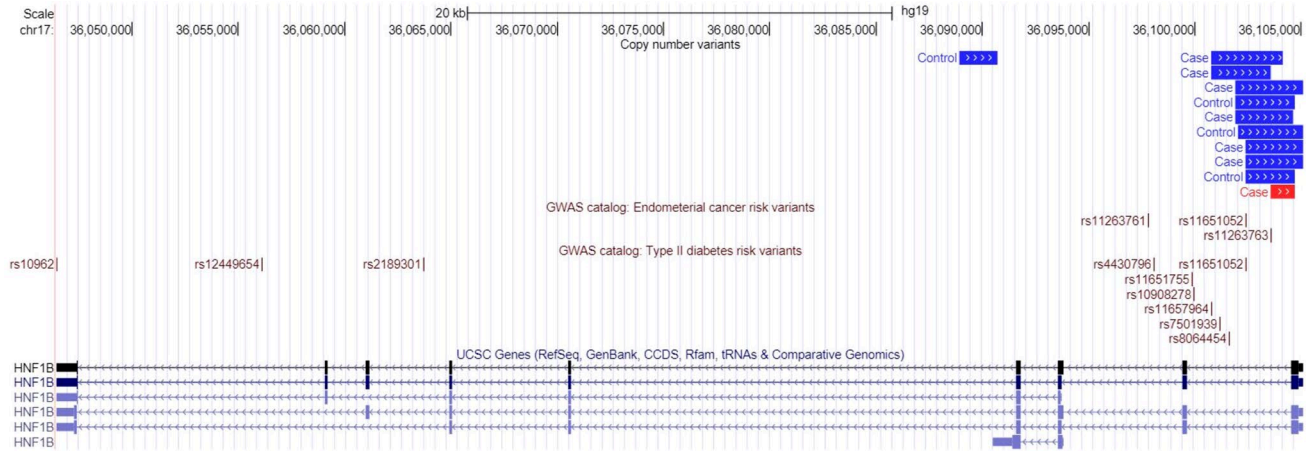

Fig. 2 Overlap of putative endometrial cancer risk copy number variants with previously identified endometrial cancer risk and type II diabetes risk variants. Copy number deletions (red) and duplications (blue) in the region of *HNF1B*

Table 2 Validation results for predicted CNVs

Gene/Loci	MAF[a]	Probes	OR (CI)[b]	*p-value*	Nanostring Validation
Deletions					
16p11.2	0.05%	47	6.05 (1.83–23.05)	3.74E–03	100% (1/1)
CTNNA3	2.58%	17	1.07 (0.87–1.32)	5.13E–01	100% (1/1)
MSH2	0.05%	2–18	6.07 (1.93–19.13)	2.08E–03	100% (2/2)
MUTYH	0.02%	7	17.34 (1.93–155.14)	1.07E–02	100% (1/1)
NPL	0.65%	3	1.83 (1.27–2.62)	1.10E–03	100% (2/2)
PMS2	0.07%	2	4.95 (1.78–13.68)	1.99E–03	100% (1/1)
FTO	0.10%	25	2.67 (1.11–6.44)	2.91E–02	100% (1/1)
SKAP1	0.13%	13	3.06 (1.46–6.42)	3.02E–03	100% (1/1)
SALL3	0.09%	15–19	16.29 (5.40–49.12)	7.16E–07	0% (0/2)
XRCC1	0.01%	2–3	8.66 (0.78–95.57)	7.70E–02	0% (0/5)
Duplications					
KIF25	2.01%	37	1.64 (1.33–2.03)	4.77E–06	100% (2/2)
SLC6A3	0.49%	9–15	8.57 (5.23–14.02)	1.34E-17	0% (0/3)

MAF minor allele frequency, *OR* odds ratio, *CI* confidence intervals (95%)

[a]Frequencies based on array data

[b]Odds ratios and p-value were calculated using logistic regression

A 600 kb deletion at 16p11.2 was validated in one sample (Table 2) using two NanoString probes targeting two different sequences located at chr16:29,653,084–29653175 and chr16:29,875,711–29,875,781. A third probe (16p11_2_389916_32171.1:87) located at chr16:30,125,121–30125192 within the predicted deletion region had insufficient counts (< 100 average counts, Supplementary Table S9). Additionally, two risk-associated deletions overlapping *NPL* ($OR_{DEL} = 1.8$, $p = 0.001$; Supplementary Table S5) and *SKAP1* ($OR_{DEL} = 3.1$, $p = 0.003$; Supplementary Table S5) were confirmed in two samples and one sample, respectively.

Pathway analysis of candidate endometrial cancer risk genes

Due to high degree of overlap between loss of function and deletion-only models (Supplementary Fig. 1), pathway analysis was independently performed on candidate endometrial cancer risk genes for duplication-only and loss-of-function CNV-GWASs (Fig. 3, Supplementary Table S10). The most significantly enriched pathway for loss-of-function CNV-GWAS is 16p11.2 proximal deletion syndrome (MIM: 611,913; $p = 6.3 \times 10^{-39}$), driven by the recurrent 600 kb long deletion (chr16:29,595,483-30,159,693) identified in six endometrial cancer cases and four controls (0.15% vs 0.02% respectively). This recurrent deletion encompasses 25 genes entirely with 24/25 genes solely impacted by this deletion. The one exception, *MAP3K*, had a single small deletion (28 kb) in one other case sample.

Moreover, when GWAS-SNP gene sets were tested for enrichment, many of the traits over-represented were driven by those overlapped by this CNV (Fig. 4; Supplementary Table S11). Interestingly, these traits included the enrichment of genes previously linked to body fat distribution (arm fat ratio) ($p = 1.2 \times 10^{-9}$).

The gene expression data Human Protein Atlas and The Cancer Genome Atlas (TCGA) were used to assess the expression in the endometrium of the genes within the 16p11.2 deletion (n = 25). Additionally, TCGA-UCEC data was used to correlate the expression of each gene with the number of DNA copies (dosage sensitivity, Supplementary Table S12, Supplementary Fig. 3). In normal endometrial tissue, one gene had no detectable expression (*C16orf92*), eight genes had low expression (0.5 < Transcripts per million [TPM] < 10; *ZG16, ASPHD1, TBX6, DOC2A, C16orf54, SPN, KCTD13, GDPD3*) and the remaining 16 exhibited high levels of expression (10 < TPM < 1000). Of the eight 'low' expression genes, only the expression levels of *TBX6* and *KCTD13* positively correlated with gene dosage. In contrast, of the more highly expressed genes in normal tissue, all except *TMEM219* and *PRRT2* had a gene dosage effect in endometrial tumour tissue (p < 0.0001). Overall, expression levels correlated positively with gene dosage ($p < 0.0001$) for 16/25 genes in endometrial tumour tissue, supporting the possibility that CNV-related impact on function results in gene expression changes and a potentially abnormal phenotype (Supplementary Table S12; Supplementary Figs. 2b).

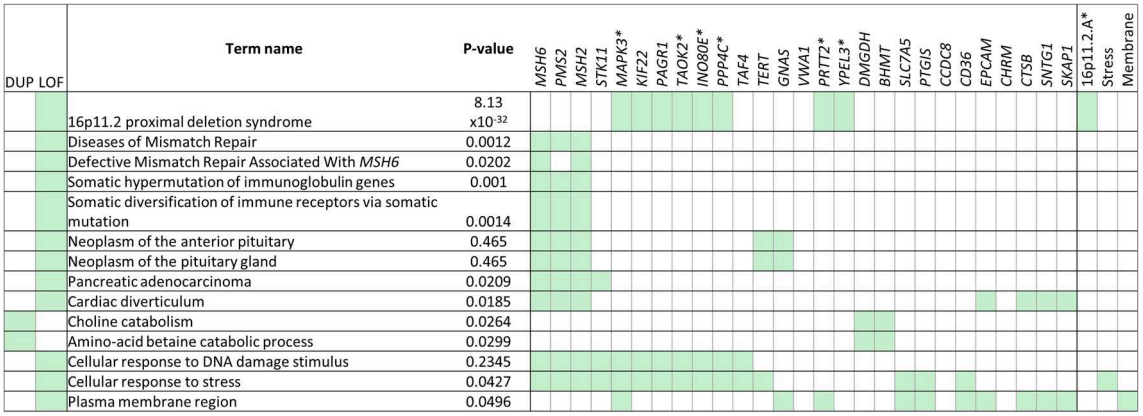

Fig. 3 Significantly over-represented pathways for candidate genes derived from duplication-only (DUP) and loss of function (LOF) CNV_GWAS. Significantly enriched pathways are ordered by adjusted p-value (most-to-least significant) of 58 and 116 candidate genes derived from duplication-only and loss of function GWAS. Reactome (REAC) (Fabregat et al., 2018), KEGG (Kanehisa et al., 2019), WikiPathways (WP) (Slenter et al., 2018), Gene Ontology (GO) (Ashburner et al., 2011), Human Phenotype Ontology (HP) (Köhler et al., 2019) were selected as annotation databases. Heatmap on left depicts which CNV-GWAS candidate genes were overrepresented. Gene sets on right side of figure encompass multiple genes: **16p11.2A** = *SPN, QRPT, C16orf54, ZG16, MAZ, MVP, CDIPT, SEZ6L2, ASPHD1, KCTD13, TMEM219, HIRIP3, DOC2A, C16orf92, ALDOA, TBX6, GDPD.* **Stress** = *CYP1B1, FGF12, PPARA, BCLAF1, POLQ, FANCM, ERCC2, GML.* **Membrane** = *SLC6A3, SLCO1B3, DLG2, TMEM231, SLC19A1, SLC4A7.* Genes denoted with * denote additional gene loci identified via recurrent 16p deletion identified in LOF CNV-GWAS but were also represented in other enriched pathways

Discussion

While a proportion of endometrial cancer cases that are not currently explained by known genetic risk factors will be explained by epistatic and gene-environment interaction, it is likely that some risk loci have yet to be identified. The objective of this study was to identify rare, germline CNVs that may be associated with endometrial cancer predisposition. We have utilised a large SNP array dataset from the Endometrial Cancer Association and Breast Cancer Association Consortium to conduct a CNV-based GWAS of endometrial cancer. A small proportion (~3%) of the cases cohort are likely to be attributed to Lynch Syndrome, however these data were not available within the study cohort (Ryan et al. 2019).

The number of rare CNVs we predicted per individual (5.1 CNVs per individual) is consistent with other case–control studies in breast (5.4 CNVs per individual) and ovarian (5.3 CNVs per individual) cancer cohorts using the same CNV calling methods (Dennis et al. 2022; DeVries et al. 2022). (DE) In our study, endometrial cancer cases had a 1.2-fold greater number of CNVs compared to controls. This is consistent with our previous analyses of a cohort of endometrial cancer cases and controls, that reported an increased burden of rare deletions involving genes and other likely functional regions (Moir-Meyer et al. 2015). However, the increased CNV burden is not observed between ovarian cancer cases and controls (DeVries et al. 2022). The cause of the discrepancy in CNV burden between studies is unclear.

Analysing rare variants is often more challenging than common variants due to the larger sample sizes needed to reach statistical significance (W. Chen et al. 2022). However, the chosen CNV calling method, CamCNV was specifically designed to detect rare variants from genotyping array data while reducing false positives (Dennis et al. 2021). The estimated false discovery rate (FDR) for CNVs called using CamCNV reduces with increasing probe coverage, with the FDR for deletions called by five probes or more estimated at 5.8% and dropping to 1.2% with 10 or more probes. Approximately 53% of deletions were called with ≥ five probes, and 27% of deletions with ≥ 10 probes, increasing our confidence in these findings. Consistent with the estimated FDR for CamCNV, 83% of candidate deletions were validated by Nanostring in deletions called with 2–47 probes. Interestingly, 67% of CNVs called with 2 probes were validated, including two deletions in the Lynch syndrome genes *MSH2* and *PMS2*. CamCNV is less reliable for duplication with a FDR for ≥ 3 probes calls estimated at 62.4% (Dennis et al. 2021). However, approximately 48% of duplications were called using ten or more probes where the FDR is estimated at 8.5%. In this study, we validated three singletons CNVs using NanoString, further increasing our confidence in CamCNVs ability to detect rare, true events.

One way to overcome difficulties associated with rare variant analysis is to perform region-based aggregation tests of multiple variants (Lee et al. 2014). In contrast to SNPs, the impact on a gene by different overlapping CNVs are assumed allowing the aggregation of these CNVs. This

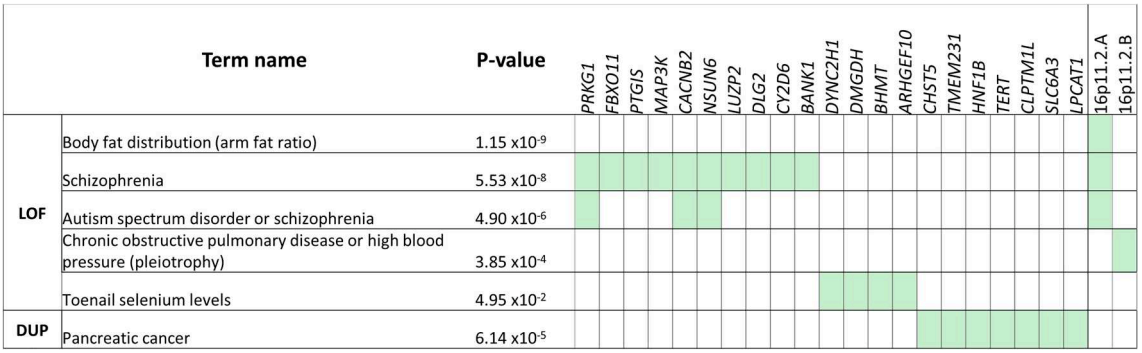

Fig. 4 Gene set enrichment analysis for candidate risk genes derived from duplication-only (DUP) and loss of function (LOF) CNV-GWAS. FUMA gene set enrichment analysis results for candidate genes derived from DUP and LOF CNV-GWAS (n = 58 and n = 116, respectively). Adjusted p-values presented. Gene sets on right side encompass two sets of genes, all of which are at 16p11.2 and driven by recurrent deletion identified. **16p11.2.A** = *SEZ6L2, ASPHD1, KCTD13, TMEM219, TAOK2, HIRIP3, INO80E, DOC2A, ALDOA, PPP4C, TBX6, YPEL3, GDPD3*. **16p11.2.B** = *TMEM219, TAOK2, HIRIP3, INO80E, DOC2A, ALDOA, PPP4C*

approach allowed us to identify genes overlapping with rare CNVs that were associated with endometrial cancer risk, including genes previously implicated in risk by SNP-based association studies. Rare CNVs are over 800 times more likely to be deleterious when compared with single nucleotide variants of the same frequency (Abel et al. 2020). A strength of this study was the loss of function CNV-GWAS in which we tested CNVs based on their likely impact of gene regions. A total of 28 gene regions were found to be significantly associated with endometrial cancer in the loss of function CNV-GWAS. *LPCAT1, TERT, MSH2* and *SLC6A3* were consistently associated with endometrial cancer risk across the three models (deletions-only, duplications-only and loss of function), suggesting a shared loss of function mechanism across CNV type. It is unclear how the loss of function of LPCAT1, TERT or SLC6A3 might contribute to endometrial cancer risk. LPCAT1 is involved in lipid metabolism (Nakanishi et al. 2006) a cellular process which when disrupted may be associated with increased endometrial cancer risk (Rosato et al. 2011). TERT has multiple functions including maintenance of telomere ends, and its activity can have oncogenic effects, such as promoting cell growth and proliferation of cancer cells (Yuan et al. 2019). SLC6A3 functions as a dopamine transporter, as can be found overexpressed in cancers, including renal cell carcinoma and gastric cancer (Hansson et al. 2017).

Our loss of function GWAS recapitulated risk associations identified in endometrial cancer SNP-studies, including variants involving *SKAP1* (O'Mara et al. 2018; Painter et al. 2018). A corresponding transcriptome wide association study (TWAS) demonstrated that decreased expression of *SKAP1* in blood was associated with an increased risk of endometrial cancer (Kho et al. 2021a, b). In this study we report a risk association between loss-of-function variants involving *SKAP1* (OR: 2.4, $p = 0.008$) and endometrial cancer risk, which is consistent with these findings. A novel finding from this study is the association between deletions involving *NPL* and endometrial cancer risk (OR: 1.8, $p = 0.001$). *NPL* regulates intracellular levels of sialic acid, with functional studies demonstrating genetic disruption of *NPL* leads to sialic acid accumulation (Wen et al. 2018). Increased sialic acid levels, or hypersialyation is commonly seen in tumour tissues and leads to accelerated cancer progression (Büll et al. 2014; Dobie & Skropeta 2021; Sun et al. 2020). Moreover, high levels of sialyation in endometrial cells has been shown to promote endometriosis outbreaks via TGF- β1 (Choi et al. 2018). Given the shared biological aetiology between endometrial cancer and non-cancerous gynaecological diseases such as endometriosis (Kho et al. 2021a, b; Painter et al. 2018), the association identified between deletions involving *NPL* and endometrial cancer risk warrants further investigation.

Obesity traits are well established risk factors for endometrial cancer (Aune et al. 2015; Painter et al. 2016), at least partly due to the accumlation of unopposed oestrogen (Lukanova et al. 2004). In this study, pathway enrichment analyses of candidate endometrial cancer risk genes revealed a strong over-representation of genes involved in 16p11.2 proximal deletion syndrome (MIM: 611,913), that is characterised by clinical heterogeneity and incomplete penetrance (Fetit et al. 2020). Proximal 16p11.2 BP4-BP5 deletions are highly pleiotropic and have been associated with many neurocognitive phenotypes, neurological tumours, morbid obesity and epilepsy (Auwerx et al. 2024; Bijlsma et al. 2009; Egolf et al. 2019; Fetit et al. 2020; Jacquemont et al. 2011; Shinawi et al. 2010; Ventura et al. 2019). This is consistent with genetic correlation between obesity traits and endometrial cancer risk (O'Mara et al. 2018). Repetitive regions at 16p11.2 result in recurrent structural changes, the most common of which being a proximal 16p11.2 BP4-BP5 deletion

at chr16: 29.6—30.2 Mb (Zufferey et al. 2012). We observed a risk-associated deletion among ten women at this locus, that is completely retained within this clinically defined region. Microdeletions at 16p11.2 result in a predisposition to obesity, with reciprocal deletions and duplications being respectively associated with obesity and being underweight, highlighting a gene dosage mechanism (Bochukova et al. 2010; Jacquemont et al. 2011; Macé et al. 2017; Walters et al. 2010). Expression levels for some but not all genes within the proximal 16p11.2 BP4-BP5 have previously been shown to correlate with copy number in pluripotent stem cells, lymphoblastoid cell lines and adipose tissues (Jacquemont et al. 2011; Roth et al. 2020; Walters et al. 2010). To our knowledge, this is the first time the relationship between gene copy and expression of genes involved in this deletion have been assessed in endometrial tissue and our results suggest potential dosage effects for the majority of genes assessed. Interestingly, the transcription factor *TBX6* is expressed at low levels in normal endometrial tissue but a correlation between *TBX6* gene dosage and expression was identified in endometrial tumour tissue. *TBX6* has been implicated as a candidate gene for another associated clinical manifestation of microdeletions at 16p11.2 which leads to a complete absence, or underdevelopment, of the female reproductive system (with Mayer-Rokitansky-Küster-Hauser syndrome [MRKH; MIM: 277000]). Studies have reported a significant association of 16p11.2 deletions among individuals with MRKH, potentially indicating that genes near this locus are involved in uterine development (Chen et al. 2021; Gatti et al. 2018). Results from this study support loss-of-function at this region is associated with endometrial cancer risk, with possible risk mechanisms being linked to obesity and/or uterine development.

Despite this being the largest endometrial cancer CNV-dataset analysed to date, the rarity of the CNVs identified results in limited power for detecting significant associations. We therefore used a nominal threshold of $p < 0.01$ to prioritise gene regions as candidate risk genes. Explicitly modelling prior associations with a generous prior did not materially alter our results providing some assurance that the genome wide adjustment used in our standard analysis is best practise, at least with our current knowledge of the genomic landscape of endometrial cancer. With this current study we aimed to identify a broad array of candidates, and thus all results reported on require further validation in independent datasets. We acknowledge that this is a limitation of the study, however in silico assessment and prioritisation was employed as a way to compliment the empirical approach. Pathway analysis of candidate genes revealed an enrichment of obesity and cancer pathways and identified multiple genes/loci that warrant further investigation.

In summary, we have conducted the largest CNV-GWAS for endometrial cancer predisposition. We have shown a global burden of rare CNVs and support the association between increased genomic load of rare CNVs and endometrial cancer risk. Our prioritisation workflow led to the identification of 141 candidate endometrial cancer susceptibility genes, many of which have plausible biological mechanisms to suggest an involvement in endometrial cancer susceptibility. Clinical features previously associated with proximal 16p11.2 BP4-BP5 deletions, including predisposition to obesity and congenital reproductive tract development, make this a particularly intriguing risk association that warrants further study.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00439-024-02707-9.

Acknowledgements The OncoArray endometrial cancer analysis were supported by NHMRC project grants [ID#1031333 & ID#1109286] to ABS, DFE, AMD, DJT and IT. This study and the authors LCW, GARW, CES, JP, VLL, ANS, MW and PS were supported by a grant from the New Zealand Health Research Council (#19/460). ABS (APP1061779), PMW, and TO'M (APP1111246) are supported by the NHMRC Fellowship scheme. AMD was supported by the Joseph Mitchell Trust. IT is supported by Cancer Research UK and the Oxford Comprehensive Biomedical Research Centre. OncoArray genotyping of ECAC cases was performed with the generous assistance of the Ovarian Cancer Association Consortium (OCAC). We particularly thank the efforts of Cathy Phelan. The OCAC OncoArray genotyping project was funded through grants from the US National Institutes of Health (CA1X01HG007491-01 (Christopher I Amos), U19-CA148112 (Thomas A Sellers), R01-CA149429 (Catherine M Phelan) and R01-CA058598 (Marc T Goodman)); Canadian Institutes of Health Research (MOP-86727 (Linda E Kelemen)); and the Ovarian Cancer Research Fund (Andrew Berchuck). CIDR genotyping for the Oncoarray was conducted under contract 268201200008I. **ANECS** recruitment was supported by project grants from the NHMRC [ID#339435], The Cancer Council Queensland [ID#4196615] and Cancer Council Tasmania [ID#403031 and ID#457636]. **SEARCH** recruitment was funded by a programme grant from Cancer Research UK [C490/A10124]. **NSECG** was supported by the EU FP7 CHIBCHA grant, Wellcome Trust Centre for Human Genetics Core Grant 090532/Z/09Z, and CORGI was funded by Cancer Research UK. The Bavarian Endometrial Cancer Study (**BECS**) was partly funded by the ELAN fund of the University of Erlangen. The Hannover-Jena Endometrial Cancer Study (HJECS) was partly supported by the Wilhelm Sander Foundation. The Leuven Endometrium Study (**LES**) was supported by the Verelst Foundation for endometrial cancer. The Mayo Endometrial Cancer Study (**MECS**) was supported by grants from the National Cancer Institute of United States Public Health Service [R01 CA122443, P30 CA15083, P50 CA136393, and GAME-ON the NCI Cancer Post-GWAS Initiative U19 CA148112], the Fred C and Katherine B Andersen Foundation, the Mayo Foundation, and the Ovarian Cancer Research Fund with support of the Smith family, in memory of Kathryn Sladek Smith. The Newcastle Endometrial Cancer Study (**NECS**) acknowledges contributions from the University of Newcastle, The NBN Children's Cancer Research Group, Ms Jennie Thomas and the Hunter Medical Research Institute. **RENDOCAS** was supported through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet [numbers: 20110222, 20110483, 20110141 and DF 07015], The Swedish Labor Market Insurance [number 100069] and The Swedish Cancer Society [number 11 0439]. The Cancer Hormone Replacement Epidemiology in Sweden Study (**CAHRES**, formerly called The Singapore and Swedish Breast/Endometrial Cancer Study; SASBAC)

was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institutes of Health and the Susan G. Komen Breast Cancer Foundation. The authors thank the many individuals who participated in this study and the numerous institutions and their staff who have supported recruitment. **ANECS** thanks members of the Molecular Cancer Epidemiology and Cancer Genetic laboratories at QIMR Berghofer Medical Research Institute for technical assistance, and the ANECS research team for assistance with the collection of risk factor information and blood samples. ANECS also gratefully acknowledges the cooperation of the following institutions: NSW: John Hunter Hospital, Liverpool Hospital, Mater Misericordiae Hospital (Sydney), Mater Misericordiae Hospital (Newcastle), Newcastle Private Hospital, North Shore Private Hospital, Royal Hospital for Women, Royal Prince Alfred Hospital, Royal North Shore Hospital, Royal Prince Alfred Hospital, St George Hospital; Westmead Hospital, Westmead Private Hospital; Qld: Brisbane Private Hospital, Greenslopes Hospital, Mater Misericordiae Hospitals, Royal Brisbane and Women's Hospital, Wesley Hospital, Queensland Cancer Registry; SA: Adelaide Pathology Partners, Burnside Hospital, Calvary Hospital, Flinders Medical Centre, Queen Elizabeth Hospital, Royal Adelaide Hospital, South Australian Cancer Registry; Tas: Launceston Hospital, North West Regional Hospitals, Royal Hobart Hospital; Vic: Freemasons Hospital, Melbourne Pathology Services, Mercy Hospital for Women, Royal Women's Hospital, Victorian Cancer Registry; WA: King Edward Memorial Hospital, St John of God Hospitals Subiaco & Murdoch, Western Australian Cancer Registry. **SEARCH** thanks the SEARCH research team for recruitment, and also acknowledges the assistance of the Eastern Cancer Registration and Information Centre for subject recruitment. **BECS** thanks Reiner Strick, Silke Landrith and Sonja Oeser for their logistic support during the study. **CAHRES** (formerly known as SASBAC) thanks Li Yuqing from the Genome Institute of Singapore for contributions to this study, and also acknowledges previous input to SASBAC resource creation by Anna Christensson, Boel Bissmarck, Kirsimari Aaltonen, Karl von Smitten, Nina Puolakka, Christer Halldén, Lim Siew Lan and Irene Chen, Lena U. Rosenberg, Mattias Hammarström, and Eija Flygare. **HJECS** thanks Wen Zheng, Hermann Hertel, and Tjoung-Won Park-Simon at Hannover Medical School for their contribution to sample recruitment. **LES** gratefully acknowledges Helena Soenen, Gilian Peuteman and Dominiek Smeets for their technical assistance. **MECS** thanks Tom Sellers, Catherine Phelan, Andrew Berchuck, and Kimberly Kalli, Amanda von Bismarck, Luisa Freyer and Lisa Rogmann. **NECS** thanks staff at the University of Newcastle and the Hunter Medical Research Institute. **NSECG** thank Ella Barclay and Lynn Martin for their contribution, and acknowledge the invaluable help of the National Cancer Research Network with the collection of study participants. **RENDOCAS** thanks Berith Wejderot, Sigrid Sahlen, Tao Liu, Margareta Ström, Maria Karlsson, and Birgitta Byström for their contribution to the study. BCAC Funding and Acknowledgments: This work was supported by Cancer Research UK grant: PPRPGM-Nov20\100002 and by core funding from the NIHR Cambridge Biomedical Research Centre (NIHR203312) [*]. *The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Additional funding for BCAC is provided by the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health, the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and the PERSPECTIVE I&I project, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation. The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report. Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer Research UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC L3 Investigator Fellow. The Australian Breast Cancer Tissue Bank (ABCTB) was supported by the National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BBCS is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). The BCEES was funded by the National Health and Medical Research Council, Australia and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foundation (JS). The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, theRobert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany. The GESBC was supported by the Deutsche Krebshilfe e. V. [70492] and the German Cancer Research Center (DKFZ). The HABCS study was supported by German Research Foundation (DFG Do761/15-1), the Claudia von Schilling Foundation for Breast Cancer Research, by the Lower Saxonian Cancer Society, and by the Rudolf Bartling Foundation. The KARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. LMBC is supported by the 'Stichting tegen Kanker'. DL is supported by the FWO. The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. The MCBCS was supported by the NIH grants R35CA253187, R01CA192393, R01CA116167, R01CA176785 a NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [P50CA116201], and the Breast Cancer Research Foundation. The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the Australian Cancer Database. The MISS study was supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society CAN 2018/675, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation FBKS 2021-19, Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. MSKCC is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics

Initiative. The SMC is funded by the Swedish Cancer Foundation and the Swedish Research Council (VR 2017-00644) grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER). The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. ABCFS thank Maggie Angelakos, Judi Maskiell, Gillian Dite. ABCS thanks the Blood bank Sanquin, The Netherlands. ABCTB Investigators: Christine Clarke, Deborah Marsh, Rodney Scott, Robert Baxter, Desmond Yip, Jane Carpenter, Alison Davis, Nirmala Pathmanathan, Peter Simpson, J. Dinny Graham, Mythily Sachchithananthan. Samples are made available to researchers on a non-exclusive basis. BBCS thanks Eileen Williams, Elaine Ryder-Mills, Kara Sargus. BCEES thanks Allyson Thomson, Christobel Saunders, Jennifer Girschik, Jane Heyworth and Terry Boyle. The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [RH, Hiltrud Brauch, Wing-Yee Lo], Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany [Yon-Dschun Ko, Christian Baisch], Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany [Ute Hamann], Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany [Thomas Brüning, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth]. HABCS thanks Peter Schürmann, Peter Hillemanns, Natalia Bogdanova, Michael Bremer, Johann Karstens, Hans Christiansen and the Breast Cancer Network in Lower Saxony for continuous support. KARMA and SASBAC thank the Swedish Medical Research Counsel. LMBC thanks Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden and Kathleen Corthouts. MARIE thanks Petra Seibold, Nadia Obi, Sabine Behrens, Ursula Eilber and Muhabbet Celik. The MCCS was made possible by the contribution of many people, including the original investigators, the teams that recruited the participants and continue working on follow-up, and the many thousands of Melbourne residents who continue to participate in the study. The MISS study group acknowledges the former Principal Investigator, Professor Håkan Olsson. We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast cancer. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre.

The OncoArray endometrial cancer analysis were supported by NHMRC project grants [ID#1031333 & ID#1109286] to ABS, DFE, AMD, DJT and IT. This study and the authors LCW, GARW, TO'M, ABS, CES, JP, VLL, ANS, MW and PS were supported by a grant from the New Zealand Health Research Council (#19/460). ABS (APP1061779), PMW, and TO'M (APP1111246) are supported by the NHMRC Fellowship scheme. AMD was supported by the Joseph Mitchell Trust. IT is supported by Cancer Research UK and the Oxford Comprehensive Biomedical Research Centre. OncoArray genotyping of ECAC cases was performed with the generous assistance of the Ovarian Cancer Association Consortium (OCAC). We particularly thank the efforts of Cathy Phelan. The OCAC OncoArray genotyping project was funded through grants from the US National Institutes of Health (CA1X01HG007491-01 (Christopher I Amos), U19-CA148112 (Thomas A Sellers), R01-CA149429 (Catherine M Phelan) and R01-CA058598 (Marc T Goodman)); Canadian Institutes of Health Research (MOP-86727 (Linda E Kelemen)); and the Ovarian Cancer Research Fund (Andrew Berchuck). CIDR genotyping for the Oncoarray was conducted under contract 268201200008I. ANECS recruitment was supported by project grants from the NHMRC [ID#339435], The Cancer Council Queensland [ID#4196615] and Cancer Council Tasmania [ID#403031 and ID#457636]. SEARCH recruitment was funded by a programme grant from Cancer Research UK [C490/A10124]. NSECG was supported by the EU FP7 CHIBCHA grant, Wellcome Trust Centre for Human Genetics Core Grant 090532/Z/09Z, and CORGI was funded by Cancer Research UK. The Bavarian Endometrial Cancer Study (BECS) was partly funded by the ELAN fund of the University of Erlangen. The Hannover-Jena Endometrial Cancer Study was partly supported by the Rudolf Bartling Foundation. The Leuven Endometrium Study (LES) was supported by the Verelst Foundation for endometrial cancer. The Mayo Endometrial Cancer Study (MECS) was supported by grants from the National Cancer Institute of United States Public Health Service [R01 CA122443, P30 CA15083, P50 CA136393, and GAME-ON the NCI Cancer Post-GWAS Initiative U19 CA148112], the Fred C and Katherine B Andersen Foundation, the Mayo Foundation, and the Ovarian Cancer Research Fund with support of the Smith family, in memory of Kathryn Sladek Smith. The Newcastle Endometrial Cancer Study (NECS) acknowledges contributions from the University of Newcastle, The NBN Children's Cancer Research Group, Ms Jennie Thomas and the Hunter Medical Research Institute. RENDOCAS was supported through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet [numbers: 20110222, 20110483, 20110141 and DF 07015], The Swedish Labor Market Insurance [number 100069] and The Swedish Cancer Society [number 11 0439]. The Cancer Hormone Replacement Epidemiology in Sweden Study (CAHRES, formerly called The Singapore and Swedish Breast/Endometrial Cancer Study; SASBAC) was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institutes of Health and the Susan G. Komen Breast Cancer Foundation. The authors thank the many individuals who participated in this study and the numerous institutions and their staff who have supported recruitment. ANECS thanks members of the Molecular Cancer Epidemiology and Cancer Genetic laboratories at QIMR Berghofer Medical Research Institute for technical assistance, and the ANECS research team for assistance with the collection of risk factor information and blood samples. ANECS also gratefully acknowledges the cooperation of the following institutions: NSW: John Hunter Hospital, Liverpool Hospital, Mater Misericordiae Hospital (Sydney), Mater Misericordiae Hospital (Newcastle), Newcastle Private Hospital, North Shore Private Hospital, Royal Hospital for Women, Royal Prince Alfred Hospital, Royal North Shore Hospital, Royal Prince Alfred Hospital, St George Hospital; Westmead Hospital, Westmead Private Hospital; Qld: Brisbane Private Hospital, Greenslopes Hospital, Mater Misericordiae Hospitals, Royal Brisbane and Women's Hospital, Wesley Hospital, Queensland Cancer Registry; SA: Adelaide Pathology Partners, Burnside Hospital, Calvary Hospital, Flinders Medical Centre, Queen Elizabeth Hospital, Royal Adelaide Hospital, South Australian Cancer Registry; Tas: Launceston Hospital, North West Regional Hospitals, Royal Hobart Hospital; Vic: Freemasons Hospital, Melbourne Pathology Services, Mercy Hospital for Women, Royal Women's Hospital, Victorian Cancer Registry; WA: King Edward Memorial Hospital, St John of God Hospitals Subiaco & Murdoch, Western Australian Cancer Registry. SEARCH thanks the SEARCH research team for recruitment, and also acknowledges the assistance of the Eastern Cancer Registration and Information Centre for subject recruitment. BECS thanks Reiner Strick, Silke Landrith and Sonja Oeser for their logistic support during the study. CAHRES (formerly known as SASBAC) thanks Li Yuqing from the Genome Institute of Singapore for contributions to this study, and also acknowledges previous input to SASBAC resource creation by Anna Christensson, Boel Bissmarck, Kirsimari Aaltonen, Karl von Smitten, Nina Puolakka, Christer Halldén, Lim Siew Lan and Irene Chen, Lena U. Rosenberg,

Mattias Hammarström, and Eija Flygare. HJECS thanks Wen Zheng, Hermann Hertel, and Tjoung-Won Park-Simon at Hannover Medical School for their contribution to sample recruitment. LES gratefully acknowledges Helena Soenen, Gilian Peuteman and Dominiek Smeets for their technical assistance. MECS thanks Tom Sellers, Catherine Phelan, Andrew Berchuck, and Kimberly Kalli, Amanda von Bismarck, Luisa Freyer and Lisa Rogmann. NECS thanks staff at the University of Newcastle and the Hunter Medical Research Institute. NSECG thank Ella Barclay and Lynn Martin for their contribution, and acknowledge the invaluable help of the National Cancer Research Network with the collection of study participants. RENDOCAS thanks Berith Wejderot, Sigrid Sahlen, Tao Liu, Margareta Ström, Maria Karlsson, and Birgitta Byström for their contribution to the study. BCAC Funding and Acknowledgments This work was supported by Cancer Research UK grant: PPRPGM-Nov20\100002 and by core funding from the NIHR Cambridge Biomedical Research Centre (NIHR203312) [*]. *The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Additional funding for BCAC is provided by the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health, the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and the PERSPECTIVE I&I project, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation. The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report. Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer Research UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC L3 Investigator Fellow. The Australian Breast Cancer Tissue Bank (ABCTB) was supported by the National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BBCS is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). The BCEES was funded by the National Health and Medical Research Council, Australia and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foundation (JS). The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany. The GESBC was supported by the Deutsche Krebshilfe e. V. [70492] and the German Cancer Research Center (DKFZ). The HABCS study was supported by German Research Foundation (DFG Do761/15-1), the Claudia von Schilling Foundation for Breast Cancer Research, by the Lower Saxonian Cancer Society, and by the Rudolf Bartling Foundation. The KARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. LMBC is supported by the 'Stichting tegen Kanker'. DL is supported by the FWO. The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. The MCBCS was supported by the NIH grants R35CA253187, R01CA192393, R01CA116167, R01CA176785 a NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [P50CA116201], and the Breast Cancer Research Foundation. The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the Australian Cancer Database. The MISS study was supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society CAN 2018/675, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation FBKS 2021-19, Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. MSKCC is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics Initiative. The SMC is funded by the Swedish Cancer Foundation and the Swedish Research Council (VR 2017-00644) grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER). The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. ABCFS thank Maggie Angelakos, Judi Maskiell, Gillian Dite. ABCS thanks the Blood bank Sanquin, The Netherlands. ABCTB Investigators: Christine Clarke, Deborah Marsh, Rodney Scott, Robert Baxter, Desmond Yip, Jane Carpenter, Alison Davis, Nirmala Pathmanathan, Peter Simpson, J. Dinny Graham, Mythily Sachchithananthan. Samples are made available to researchers on a non-exclusive basis. BBCS thanks Eileen Williams, Elaine Ryder-Mills, Kara Sargus. BCEES thanks Allyson Thomson, Christobel Saunders, Jennifer Girschik, Jane Heyworth and Terry Boyle. The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [RH, Hiltrud Brauch, Wing-Yee Lo], Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany [Yon-Dschun Ko, Christian Baisch], Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany [Ute Hamann], Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany [Thomas Brüning, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth]. HABCS thanks Peter Schürmann, Peter Hillemanns, Natalia Bogdanova, Michael Bremer, Johann Karstens, Hans Christiansen and the Breast Cancer Network in Lower Saxony for continuous support. KARMA and SASBAC thank the Swedish Medical Research Counsel. LMBC thanks Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden and Kathleen Corthouts. MARIE thanks Petra

Seibold, Nadia Obi, Sabine Behrens, Ursula Eilber and Muhabbet Celik. The MCCS was made possible by the contribution of many people, including the original investigators, the teams that recruited the participants and continue working on follow-up, and the many thousands of Melbourne residents who continue to participate in the study. The MISS study group acknowledges the former Principal Investigator, Professor Håkan Olsson. We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast cancer. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre.

Author contributions L.C.W, S.C, G.A.R.W, J.P, V.L.L, A.B.S, A.N.S, P.S and W.S designed the study; G.A.R.W, C.E.S, J.P and L.C.W drafted the manuscript. G.A.R.W, C.E.S, and J.P conducted statistical analysis. J.D conducted copy number calling. T.A.O'M. co-ordinated the iCOGS and OncoArray case genotyping, and associated data management; J.D, and K.M co-ordinated quality control and data cleaning for the iCOGS and OncoArray control datasets. A.B.S and T.A.O'M co-ordinated the ANECS stage 1 genotyping; A.M.D co-ordinated the SEARCH stage 1 genotyping; I.T funded and implemented the NSECG GWAS; I.T, co-ordinated the National Study of Endometrial Cancer Genetics (NSECG), and collation of CORGI control GWAS data; A.B.S. co-ordinated the Australian National Endometrial Cancer Study (ANECS); R.J.S., co-ordinated collation of GWAS data for the Hunter Community Study; P.D.P.P, D.F.E, and M.S. co-ordinated Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH); The following authors provided samples and/or phenotypic data: F.A, D.A, W.DW, P.A.F, E.L.G, D.M.G, E.L.G, D.L, E.T, M.K.B, F.J.C, K.C, T.D, O.F, M.G-C, R.H, H.J, R.K, N.O, M.C.S, J.S, and Q.W. All authors provided critical review of the manuscript. L.C.W, S.C, G.A.R.W, J.P, V.L.L, A.B.S, A.N.S, P.S and W.S designed the study; G.A.R.W, C.E.S, J.P and L.C.W drafted the manuscript. G.A.R.W, C.E.S, and J.P conducted statistical analysis. J.D conducted copy number calling. T.A.O'M. co-ordinated the iCOGS and OncoArray case genotyping, and associated data management; J.D, and K.M co-ordinated quality control and data cleaning for the iCOGS and OncoArray control datasets. A.B.S and T.A.O'M co-ordinated the ANECS stage 1 genotyping; A.M.D co-ordinated the SEARCH stage 1 genotyping; I.T funded and implemented the NSECG GWAS; I.T, co-ordinated the National Study of Endometrial Cancer Genetics (NSECG), and collation of CORGI control GWAS data; A.B.S. co-ordinated the Australian National Endometrial Cancer Study (ANECS); R.J.S., co-ordinated collation of GWAS data for the Hunter Community Study; P.D.P.P, D.F.E, and M.S. co-ordinated Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH); The following authors provided samples and/or phenotypic data: F.A, D.A, W.DW, P.A.F, E.L.G, D.M.G, E.L.G, D.L, E.T, M.K.B, F.J.C, K.C, T.D, O.F, M.G-C, R.H, H.J, R.K, N.O, M.C.S, J.S, and Q.W. All authors provided critical review of the manuscript.

Funding statement This study was supported by funding from New Zealand Health Research Council (#19/460).

Data availability OncoArray germline genotype data for the ECAC studies have been deposited through the database of Genotypes and Phenotypes (dbGaP; accession number phs000893.v1.p1). Genotype data for non-cancer controls were provided by the Breast Cancer Association Consortium (BCAC) by application to the BCAC Data Access Coordination Committee (http://bcac.ccge.medschl.cam.ac.uk/).

Declarations

Competing interests The authors declare no competing interests.

References

Abel HJ, Larson DE, Regier AA, Chiang C, Das I, Kanchi KL, Layer RM, Neale BM, Salerno WJ, Reeves C, Buyske S, Abecasis GR, Appelbaum E, Baker J, Banks E, Bernier RA, Bloom T, Boehnke M, Boerwinkle E, Hall IM (2020) Mapping and characterization of structural variation in 17,795 human genomes. Nature. https://doi.org/10.1038/s41586-020-2371-0

Aguirre M, Rivas MA, Priest J (2019) Phenome-wide burden of copy-number variation in the UK Biobank. Am J Human Genet 105(2):373–383. https://doi.org/10.1016/J.AJHG.2019.07.001

Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, Casey G, Hunter DJ, Sellers TA, Gruber SB, Dunning AM, Michailidou K, Fachal L, Doheny K, Spurdle AB, Li Y, Xiao X, Romm J, Pugh E, Easton DF (2017) The oncoarray consortium: a network for understanding the genetic architecture of common cancers. Cancer Epidemiol Biomark Prev. https://doi.org/10.1158/1055-9965.EPI-16-0106

Aune D, Navarro Rosenblatt DA, Chan DSM, Vingeliene S, Abar L, Vieira AR, Greenwood DC, Bandera EV, Norat T (2015) Anthropometric factors and endometrial cancer risk: A systematic review and dose-response meta-analysis of prospective studies. Ann Oncol. https://doi.org/10.1093/annonc/mdv142

Auwerx C, Moix S, Kutalik Z, Reymond A (2024) Disentangling mechanisms behind the pleiotropic effects of proximal 16p11.2 BP4–5 CNVs. MedRxiv. https://doi.org/10.1101/2024.03.20.24304613

Bijlsma EK, Gijsbers ACJ, Schuurs-Hoeijmakers JHM, van Haeringen A, Fransen van de Putte DE, Anderlid BM, Lundin J, Lapunzina P, Pérez Jurado LA, Delle Chiaie B, Loeys B, Menten B, Oostra A, Verhelst H, Amor DJ, Bruno DL, van Essen AJ, Hordijk R, Sikkema-Raddatz B, Ruivenkamp CAL (2009) Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. Eur J Med Genet 52(2–3):77–87. https://doi.org/10.1016/J.EJMG.2009.03.006

Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, Saeed S, Hamilton-Shield J, Clayton-Smith J, O'Rahilly S, Hurles ME, Farooqi IS (2010) Large, rare chromosomal deletions associated with severe early-onset obesity. Nature. https://doi.org/10.1038/nature08689

Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen H-Z, Reeser JW, Yu L, Roychowdhury S (2017) Landscape of microsatellite instability across 39 cancer types. JCO Precis Oncol. https://doi.org/10.1200/po.17.00073

Büll C, Stoel MA, den Brok MH, Adema GJ (2014) Sialic acids sweeten a tumor's life. Can Res 74(12):3199–3204

Chen MM, O'Mara TA, Thompson DJ, Painter JN, Attia J, Black A, Brinton L, Chanock S, Chen C, Cheng TH, Cook LS, Crous-Bou M, Doherty J, Friedenreich CM, Garcia-Closas M, Gaudet MM, Gorman M, Haiman C, Hankinson SE, De Vivo I (2016) GWAS meta-analysis of 16 852 women identifies new susceptibility locus for endometrial cancer. Human Mol Genet 25(12):ddw092. https://doi.org/10.1093/hmg/ddw092

Chen N, Zhao S, Jolly A, Wang L, Pan H, Yuan J, Chen S, Koch A, Ma C, Tian W, Jia Z, Kang J, Zhao L, Qin C, Fan X, Rall K, Coban-Akdemir Z, Chen Z, Jhangiani S, Zhu L (2021) Perturbations of genes essential for Müllerian duct and Wölffian duct development in Mayer-Rokitansky-Küster-Hauser syndrome. Am J Human Genet 108(2):337–345. https://doi.org/10.1016/j.ajhg.2020.12.014

Chen W, Coombes BJ, Larson NB (2022) Recent advances and challenges of rare variant association analysis in the biobank sequencing era. Front Genet 13:1014947. https://doi.org/10.3389/fgene.2022.1014947

Choi HJ, Chung TW, Choi HJ, Han JH, Choi JH, Kim CH, Ha KT (2018) Increased α2-6 sialylation of endometrial cells contributes to the development of endometriosis. Exp Mol Med. https://doi.org/10.1038/s12276-018-0167-1

Cooper NJ, Shtir CJ, Smyth DJ, Guo H, Swafford AD, Zanda M, Hurles ME, Walker NM, Plagnol V, Cooper JD, Howson JMM, Burren OS, Onengut-Gumuscu S, Rich SS, Todd JA (2015) Detection and correction of artefacts in estimation of rare copy number variants and analysis of rare deletions in type 1 diabetes. Hum Mol Genet 24(6):1774–1790. https://doi.org/10.1093/HMG/DDU581

Dennis J, Walker L, Tyrer J, Michailidou K, Easton DF (2021) Detecting rare copy number variants from Illumina genotyping arrays with the CamCNV pipeline: segmentation of z-scores improves detection and reliability. Genet Epidemiol. https://doi.org/10.1002/gepi.22367

Dennis J, Tyrer JP, Walker LC, Michailidou K, Dorling L, Bolla MK, Wang Q, Ahearn TU, Andrulis IL, Anton-Culver H, Antonenkova NN, Arndt V, Aronson KJ, Freeman LEB, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Bogdanova NV, Easton DF (2022) Rare germline copy number variants (CNVs) and breast cancer risk. Commun Biol 5(1):1–15. https://doi.org/10.1038/s42003-021-02990-6

DeVries AA, Dennis J, Tyrer JP, Peng PC, Coetzee SG, Reyes AL, Plummer JT, Davis BD, Chen SS, Dezem FS, Aben KKH, Anton-Culver H, Antonenkova NN, Beckmann MW, Beeghly-Fadiel A, Berchuck A, Bogdanova NV, Bogdanova-Markov N, Brenton JD, Jones MR (2022) Copy number variants are ovarian cancer risk alleles at known and novel risk loci. J Natl Cancer Inst 114(11):1533–1544. https://doi.org/10.1093/JNCI/DJAC160

Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, Bucan M, Maris JM, Wang K (2008) Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. https://doi.org/10.1093/nar/gkn556

Dobie C, Skropeta D (2021) Insights into the role of sialylation in cancer progression and metastasis. Br J Cancer. https://doi.org/10.1038/s41416-020-01126-7

Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. Nat Protoc. https://doi.org/10.1038/nprot.2009.97

Egolf LE, Vaksman Z, Lopez G, Rokita JL, Modi A, Basta PV, Hakonarson H, Olshan AF, Diskin SJ (2019) Germline 16p11.2 microdeletion predisposes to neuroblastoma. Am J Human Genet 105(3):658–668. https://doi.org/10.1016/J.AJHG.2019.07.020

Fetit R, Price DJ, Lawrie SM, Johnstone M (2020) Understanding the clinical manifestations of 16p11.2 deletion syndrome: a series of developmental case reports in children. Psychiatric Genet. https://doi.org/10.1097/YPG.0000000000000259

Gatti M, Tolva G, Bergamaschi S, Giavoli C, Esposito S, Marchisio P, Milani D (2018) Mayer–Rokitansky–Küster–Hauser syndrome and 16p11.2 recurrent microdeletion: a case report and review of the literature. J Pediatric Adolesc Gynecol 31(5):533–535

Hansson J, Lindgren D, Nilsson H, Johansson E, Johansson M, Gustavsson L, Axelson H (2017) Overexpression of functional SLC6A3 in clear cell renal cell carcinoma. Clin Cancer Res 23(8):2105–2115. https://doi.org/10.1158/1078-0432.CCR-16-0496

Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, Martinet D, Shen Y, Valsesia A, Beckmann ND, Thorleifsson G, Belfiore M, Bouquillon S, Campion D, De Leeuw N, De Vries BBA, Esko T, Fernandez BA, Fernández-Aranda F, Froguel P (2011) Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature. https://doi.org/10.1038/nature10406

Kho PF, Mortlock S, Yang J, Digna R, Nyegaard M, Low S-K, Zondervan KT, Missmer SA (2021a) Genetic analyses of gynecological disease identify genetic relationships between uterine fibroids and endometrial cancer, and a novel endometrial cancer genetic risk region at the WNT4 1p36. 12 locus. Hum Genet 140:1353–1365

Kho PF, Wang X, Cuéllar-Partida G, Dörk T, Goode EL, Lambrechts D, Scott RJ, Spurdle AB, O'Mara TA, Glubb DM (2021b) Multi-tissue transcriptome-wide association study identifies eight candidate genes and tissue-specific gene expression underlying endometrial cancer susceptibility. Commun Biol. https://doi.org/10.1038/s42003-021-02745-3

Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H (2020) gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Research. https://doi.org/10.12688/f1000research.24956.2

Lagerstedt-Robinson K, Rohlin A, Aravidis C, Melin B, Nordling M, Stenmark-Askmalm M, Lindblom A, Nilbert M (2016) Mismatch repair gene mutation spectrum in the Swedish Lynch syndrome population. Oncol Rep. https://doi.org/10.3892/or.2016.5060

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). Software forcomputing and annotating genomic ranges. PLoS Comput Biol 9(8):e1003118.https://doi.org/10.1371/journal.pcbi.1003118

LG, Martin M (2022) hpar: Human Protein Atlas in R. R Package Version 1.40.0

Lee S, Abecasis GR, Boehnke M, Lin X (2014) Rare-variant association analysis: study designs and statistical tests. Am J Human Genet. https://doi.org/10.1016/j.ajhg.2014.06.009

Li YR, Glessner JT, Coe BP, Li J, Mohebnasab M, Chang X, Connolly J, Kao C, Wei Z, Bradfield J, Kim C, Hou C, Khan M, Mentch F, Qiu H, Bakay M, Cardinale C, Lemma M, Abrams D, Hakonarson H (2020) Rare copy number variants in over 100,000 European ancestry subjects reveal multiple disease associations. Nat Commun 11(1):1–9. https://doi.org/10.1038/s41467-019-13624-1

Lortet-Tieulent J, Ferlay J, Bray F, Jemal A (2018) International patterns and trends in endometrial cancer incidence, 1978–2013. J Natl Cancer Inst 110(4):354–361. https://doi.org/10.1093/jnci/djx214

Lukanova A, Lundin E, Micheli A, Arslan A, Ferrari P, Rinaldi S, Krogh V, Lenner P, Shore RE, Biessy C, Muti P, Riboli E, Koenig KL, Levitz M, Stattin P, Berrino F, Hallmans G, Kaaks R, Toniolo P, Zeleniuch-Jacquotte A (2004) Circulating levels of sex steroid hormones and risk of endometrial cancer in postmenopausal women. Int J Cancer 108(3):425–432. https://doi.org/10.1002/IJC.11529

MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, MayPendlington Z, Welter D, Burdett T, Hindorff L, Flicek P, Cunningham F, Parkinson H (2017) The new NHGRI-EBI Catalog of published genome-wide

association studies (GWAS Catalog). Nucleic Acids Res 45(D1):D896–D901. https://doi.org/10.1093/NAR/GKW1133

Macé A, Tuke MA, Beckmann JS, Lin L, Jacquemont S, Weedon MN, Reymond A, Kutalik Z (2016) New quality measure for SNP array based CNV detection. Bioinformatics 32(21):3298–3305. https://doi.org/10.1093/BIOINFORMATICS/BTW477

Macé A, Tuke MA, Deelen P, Kristiansson K, Mattsson H, Nõukas M, Sapkota Y, Schick U, Porcu E, Rüeger S, McDaid AF, Porteous D, Winkler TW, Salvi E, Shrine N, Liu X, Ang WQ, Zhang W, Feitosa MF, Kutalik Z (2017) CNV-association meta-analysis in 191,161 European adults reveals new loci associated with anthropometric traits. Nat Commun. https://doi.org/10.1038/s41467-017-00556-x

Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, Lemaçon A, Soucy P, Glubb D, Rostamianfar A, Bolla MK, Wang Q, Tyrer J, Dicks E, Lee A, Wang Z, Allen J, Keeman R, Eilber U, Easton DF (2017) Association analysis identifies 65 new breast cancer risk loci. Nature. https://doi.org/10.1038/nature24284

Moir-Meyer GL, Pearson JF, Lose F, Scott RJ, McEvoy M, Attia J, Holliday EG, Pharoah PD, Dunning AM, Thompson DJ, Easton DF, Spurdle AB, Walker LC, The Australian National Endometrial Cancer Study Group, The Hunter Community Study, Studies of Epidemiology (2015) Rare germline copy number deletions of likely functional importance are implicated in endometrial cancer predisposition. Human Genet. https://doi.org/10.1007/s00439-014-1507-4

Morgan M, Shepherd L (2022) ExperimentHub: Client to access ExperimentHub resources. R Package Version 2.6.0

Nakanishi H, Shindou H, Hishikawa D, Harayama T, Ogasawara R, Suwabe A, Taguchi R, Shimizu T (2006) Cloning and characterization of mouse lung-type acyl-CoA:lysophosphatidylcholine acyltransferase 1 (LPCAT1). Expression in alveolar type II cells and possible involvement in surfactant production. J Biol Chem 281(29):20140–20147. https://doi.org/10.1074/JBC.M600225200

O'Mara TA, Glubb DM, Amant F, Annibali D, Ashton K, Attia J, Auer PL, Beckmann MW, Black A, Bolla MK, Brauch H, Brenner H, Brinton L, Buchanan DD, Burwinkel B, Chang-Claude J, Chanock SJ, Chen C, Chen MM, Thompson DJ (2018) Identification of nine new susceptibility loci for endometrial cancer. Nat Commu. https://doi.org/10.1038/s41467-018-05427-7

Painter JN, O'Mara TA, Marquart L, Webb PM, Attia J, Medland SE, Cheng T, Dennis J, Holliday EG, McEvoy M, Scott RJ, Ahmed S, Healey CS, Shah M, Gorman M, Martin L, Hodgson SV, Beckmann MW, Ekici AB, Spurdle AB (2016) Genetic risk score mendelian randomization shows that obesity measured as body mass index, but not waist: hip ratio, is causal for endometrial cancer. Cancer Epidemiol Biomark Prev 25(11):1503–1510. https://doi.org/10.1158/1055-9965.EPI-16-0147

Painter JN, O'Mara TA, Morris AP, Cheng THT, Gorman M, Martin L, Hodson S, Jones A, Martin NG, Gordon S, Henders AK, Attia J, McEvoy M, Holliday EG, Scott RJ, Webb PM, Fasching PA, Beckmann MW, Ekici AB, Spurdle AB (2018) Genetic overlap between endometriosis and endometrial cancer: evidence from cross-disease genetic correlation and GWAS meta-analyses. Cancer Med 7(5):1978–1987. https://doi.org/10.1002/cam4.1445

Papatheodorou I, Fonseca NA, Keays M, Tang YA, Barrera E, Bazant W, Burke M, Füllgrabe A, Fuentes AM-P, George N (2018) Expression Atlas: gene and protein expression across multiple studies and organisms. Nucleic Acids Res 46(D1):D246–D251

Rainer J (2017) EnsDb.Hsapiens.v75: Ensembl based annotation package

Ramos M, Geistlinger L, Oh S, Schiffer L, Azhar R, Kodali H, de Bruijn I, Gao J, Carey VJ, Morgan M, Waldron L (2020) Multiomic integration of public oncology databases in bioconductor. JCO Clin Cancer Inform. https://doi.org/10.1200/cci.19.00119

Rodríguez-Palacios DÁ, Colorado-Yohar SM, Velten M, Vaamonde-Martín RJ, Ballesta M, Chirlaque MD (2022) Incidence and trend of type I and II endometrial cancer in women from Two Population-Based European Cancer Registries (1998–2012). Int J Environ Res Public Health. https://doi.org/10.3390/ijerph19073789

Romero A, Garre P, Valentin O, Sanz J, Pérez-Segura P, Llovet P, Díaz-Rubio E, de la Hoya M, Caldés T (2013) Frequency and variability of genomic rearrangements on MSH2 in Spanish Lynch Syndrome families. PLoS ONE. https://doi.org/10.1371/journal.pone.0072195

Rosato V, Zucchetto A, Bosetti C, Dal Maso L, Montella M, Pelucchi C, Negri E, Franceschi S, La Vecchia C (2011) Metabolic syndrome and endometrial cancer risk. Ann Oncol 22(4):884–889. https://doi.org/10.1093/ANNONC/MDQ464

Roth JG, Muench KL, Asokan A, Mallett VM, Gai H, Verma Y, Weber S, Charlton C, Fowler JL, Loh KM, Dolmetsch RE, Palmer TD (2020) 16p11.2 microdeletion imparts transcriptional alterations in human iPSC-derived models of early neural development. Elife 9:1–34. https://doi.org/10.7554/ELIFE.58178

Ryan NAJ, Glaire MA, Blake D, Cabrera-Dandy M, Evans DG, Crosbie EJ (2019) The proportion of endometrial cancers associated with Lynch syndrome: a systematic review of the literature and meta-analysis. Genet Med 21(10):2167–2180. https://doi.org/10.1038/S41436-019-0536-8

Shinawi M, Liu P, Kang SHL, Shen J, Belmont JW, Scott DA, Probst FJ, Craigen WJ, Graham BH, Pursley A, Clark G, Lee J, Proud M, Stocco A, Rodriguez DL, Kozel BA, Sparagana S, Roeder ER, McGrew SG, Lupski JR (2010) Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. J Med Genet 47(5):332–341. https://doi.org/10.1136/JMG.2009.073015

Sun H, Zhou Y, Jiang H, Xu Y (2020) Elucidation of functional roles of sialic acids in cancer migration. Front Oncol. https://doi.org/10.3389/fonc.2020.00401

Truty R, Paul J, Kennemer M, Lincoln SE, Olivares E, Nussbaum RL, Aradhya S (2019) Prevalence and properties of intragenic copy-number variation in Mendelian disease genes. Genet Med 21(1):114–123. https://doi.org/10.1038/S41436-018-0033-5

Ventura M, Gomes L, Rosmaninho-Salgado J, Barros L, Paiva I, Melo M, Oliveira D, Carrilho F (2019) Bifocal germinoma in a patient with 16p112 microdeletion syndrome. Endocrinol Diabetes Metab Case Rep. https://doi.org/10.1530/EDM-18-0149

Wakefield J (2008) Reporting and interpretation in genome-wide association studies. Int J Epidemiol 37(3):641–653. https://doi.org/10.1093/IJE/DYM257

Walters RG, Jacquemont S, Valsesia A, De Smith AJ, Martinet D, Andersson J, Falchi M, Chen F, Andrieux J, Lobbens S, Delobel B, Stutzmann F, El-Sayed Moustafa JS, Chèvre JC, Lecoeur C, Vatin V, Bouquillon S, Buxton JL, Boute O, Beckmann JS (2010) A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature. https://doi.org/10.1038/nature08727

Wang X, Glubb DM, O'Mara TA (2022) 10 Years of GWAS discovery in endometrial cancer: aetiology, function and translation. EBioMedicine. https://doi.org/10.1016/J.EBIOM.2022.103895

Watanabe K, Taskesen E, Van Bochoven A, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. Nat Commun. https://doi.org/10.1038/s41467-017-01261-5

Wen XY, Tarailo-Graovac M, Brand-Arzamendi K, Willems A, Rakic B, Huijben K, Da Silva A, Pan X, El-Rass S, Ng R, Selby K, Philip AM, Yun J, Ye XC, Ross CJ, Lehman AM, Zijlstra F, Abu Bakar N, Drögemöller B, Lefeber DJ (2018) Sialic acid catabolism by N-acetylneuraminate pyruvate lyase is essential for muscle function. JCI Insight. https://doi.org/10.1172/jci.insight.122373

Yuan X, Larsson C, Xu D (2019) Mechanisms underlying the activation of TERT transcription and telomerase activity in human cancer: old actors and new players. Oncogene 38(34):6172–6183. https://doi.org/10.1038/s41388-019-0872-9

Zhang F, Gu W, Hurles ME, Lupski JR (2009) Copy number variation in human health, disease, and evolution. Annu Rev Genom Hum Genet 10:451. https://doi.org/10.1146/ANNUREV.GENOM.9.081307.164217

Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L, Macé A, Ferrari C, Kutalik Z, Andrieux J, Aylward E, Barker M, Bernier R, Bouquillon S, Conus P, Delobel B, Faucett WA, Goin-Kochel RP, Grant E, Roberts TPL (2012) A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. J Med Genet. https://doi.org/10.1136/jmedgenet-2012-101203

Authors and Affiliations

Cassie E. Stylianou[1] · George A. R. Wiggins[1] · Vanessa L. Lau[1] · Joe Dennis[2] · Andrew N. Shelling[3] · Michelle Wilson[4] · Peter Sykes[5] · Frederic Amant[6,7] · Daniela Annibali[7] · Wout De Wispelaere[7] · Douglas F. Easton[2,8] · Peter A. Fasching[9] · Dylan M. Glubb[10] · Ellen L. Goode[11] · Diether Lambrechts[12,13] · Paul D. P. Pharoah[14] · Rodney J. Scott[15,16,17] · Emma Tham[18,19] · Ian Tomlinson[20] · Manjeet K. Bolla[2] · Fergus J. Couch[21] · Kamila Czene[22] · Thilo Dörk[23] · Alison M. Dunning[8] · Olivia Fletcher[24] · Montserrat García-Closas[25] · Reiner Hoppe[26,27] · ABCTB Investigators[28] · Helena Jernström[29] · Rudolf Kaaks[30] · Kyriaki Michailidou[2,31] · Nadia Obi[32,33] · Melissa C. Southey[34,35,36] · Jennifer Stone[37,38] · Qin Wang[2] · Amanda B. Spurdle[39] · Tracy A. O'Mara[10] · John Pearson[40] · Logan C. Walker[1]

✉ George A. R. Wiggins
george.wiggins@otago.ac.nz

[1] Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand

[2] Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK

[3] Department of Obstetrics and Gynaecology, University of Auckland, Auckland, New Zealand

[4] Te Pūriri o Te Ora Regional Cancer and Blood Service, Auckland Hospital, Auckland, New Zealand

[5] Department of Obstetrics and Gynaecology, University of Otago, Christchurch, New Zealand

[6] Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University Hospitals KU Leuven, University of Leuven, Leuven, Belgium

[7] Gynecological Oncology Laboratory, Department of Oncology, KU Leuven and Leuven Cancer Institute (LKI), Leuven, Belgium

[8] Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK

[9] Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

[10] Cancer Research Program, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

[11] Division of Epidemiology, Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA

[12] Laboratory for Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium

[13] VIB Center for Cancer Biology, VIB, Leuven, Belgium

[14] Department of Computational Biomedicine, Cedars-Sinai Medical Center, West Hollywood, CA, USA

[15] Division of Molecular Medicine, Pathology North, John Hunter Hospital, Newcastle, NSW, Australia

[16] Faculty of Health, Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan, NSW, Australia

[17] Hunter Medical Research Institute, John Hunter Hospital, Newcastle, NSW, Australia

[18] Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

[19] Clinical Genetics and Genomics, Karolinska University Hospital, Stockholm, Sweden

[20] Department of Oncology, University of Oxford, Oxford, UK

[21] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

[22] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[23] Gynaecology Research Unit, Hannover Medical School, Hannover, Germany

[24] The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK

[25] Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK

[26] Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany

[27] University of Tübingen, Tübingen, Germany

[28] Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, NSW, Australia

[29] Oncology, Department of Clinical Sciences in Lund, Lund University, Lund, Sweden

[30] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

[31] Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus

[32] Institute for Occupational and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[33] Institute for Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[34] Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia

[35] Department of Clinical Pathology, The University of Melbourne, Melbourne, VIC, Australia

[36] Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia

[37] Genetic Epidemiology Group, School of Population and Global Health, University of Western Australia, Perth, WA, Australia

[38] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia

[39] Public Health Program, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

[40] Department of Medicine, University of Otago, Christchurch, New Zealand