MicroRNA Processing and Binding Site Polymorphisms Are Not Replicated in the Ovarian Cancer Association Consortium

Published in final edited form as:
Cancer Epidemiol Biomarkers Prev. 2011 August ; 20(8): 1793–1797. doi: 10.1158/1055-9965.EPI-11-0397.

MicroRNA Processing and Binding Site Polymorphisms are not Replicated in the Ovarian Cancer Association Consortium

Jennifer Permuth-Wey[1], **Zhihua Chen**[2], **Ya-Yu Tsai**[1], **Hui-Yi Lin**[3], **Y. Ann Chen**[3], **Jill Barnholtz-Sloan**[4], **Michael J. Birrer**[5], **Stephen J. Chanock**[6], **Daniel W. Cramer**[7], **Julie M. Cunningham**[8], **David Fenstermacher**[2], **Brooke L. Fridley**[9], **Montserrat Garcia-Closas**[10], **Simon A. Gayther**[11], **Aleksandra Gentry-Maharaj**[12], **Jesus Gonzalez-Bosquet**[13], **Edwin Iversen**[14], **Heather Jim**[15], **John McLaughlin**[16], **Usha Menon**[12], **Steven A. Narod**[17], **Catherine M. Phelan**[1], **Susan J. Ramus**[11], **Harvey Risch**[18], **Honglin Song**[19], **Rebecca Sutphen**[20], **Kathryn L. Terry**[7], **Jonathan Tyrer**[19], **Robert A. Vierkant**[9], **Nicolas Wentzensen**[6], **Johnathan M. Lancaster**[13], **Jin Q. Cheng**[21], **Andrew Berchuck**[22], **Paul D.P. Pharoah**[19], **Joellen M. Schildkraut**[23], **Ellen L. Goode**[9], and **Thomas A. Sellers**[1] **on behalf of the Ovarian Cancer Association Consortium (OCAC)**

[1] Cancer Epidemiology Program, Division of Population Sciences, Moffitt Cancer Center, Tampa, FL, USA

[2] Department of Biomedical Informatics, Moffitt Cancer Center, Tampa, FL, USA

[3] Department of Biostatistics, Moffitt Cancer Center, Tampa, FL, USA

[4] Case Comprehensive Cancer Center, Case School of Medicine, Cleveland, OH, USA

[5] Massachusetts General Hospital, Boston, MA, USA

[6] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA

[7] Obstetrics and Gynecology Epidemiology Center, Brigham Women's Hospital, Boston, MA, USA

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

[9] Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, USA

[10] Sections of Epidemiology and Genetics at the Institute of Cancer Research and Breakthrough Breast Cancer Research Centre, London UK

[11] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA

[12] Department of Gynaecological Oncology, University College London, EGA Institute for Women's Health, London, UK

[13] Department of Women's Oncology, Moffitt Cancer Center, Tampa, FL, USA

[14] Department of Statistical Science, Duke University, Durham, NC, USA

[15] Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, USA

[16] Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada

Address correspondence to: Thomas A. Sellers, PhD, Director, Moffitt Research Institute, Executive Vice President, Moffitt Cancer Center, Department of Epidemiology, 12902 Magnolia Drive, MRC-CANCONT, Tampa, FL 33612-9416, Phone: 813-745-1315 Fax: 813-449-8126, Thomas.Sellers@Moffitt.org.

[17] Center for Research in Women's Health, Toronto, ON, Canada

[18] Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, USA

[19] Department of Oncology, University of Cambridge, Cambridge, UK

[20] Pediatrics Epidemiology Center, College of Medicine, University of South Florida, Tampa, FL, USA

[21] Department of Interdisciplinary Oncology, Moffitt Cancer Center, Tampa, FL, USA

[22] Department of Obstetrics and Gynecology, Duke Comprehensive Cancer Center, Durham, NC, USA

[23] Department of Community and Family Medicine, Duke University Medical Center, Durham, NC, USA

Abstract

Background—Single nucleotide polymorphisms (SNPs) in microRNA-related genes have been associated with epithelial ovarian cancer (EOC) risk in two reports, yet associated alleles may be inconsistent across studies.

Methods—We conducted a pooled analysis of previously-identified SNPs by combining genotype data from 3,973 invasive EOC cases and 3,276 controls from the Ovarian Cancer Association Consortium. We also conducted imputation to obtain dense coverage of genes and comparable genotype data for all studies. In total, 226 SNPs within 15 kilobases of 4 miRNA biogenesis genes (*DDX20, DROSHA, GEMIN4,* and *XPO5)* and 23 SNPs located within putative miRNA binding sites of 6 genes (*CAV1, COL18A1, E2F2, IL1R1, KRAS,* and *UGT2A3*) were genotyped or imputed and analyzed in the entire dataset.

Results—After adjustment for European ancestry, no overall association was observed between any of the analyzed SNPs and EOC risk.

Conclusions—Common variants in these evaluated genes do not appear to be strongly associated with EOC risk.

Impact—This analysis suggests earlier associations between EOC risk and SNPs in these genes may have been chance findings, possibly confounded by population admixture. To more adequately evaluate the relationship between genetic variants and cancer risk, large sample sizes are needed, adjustment for population stratification should be performed, and use of imputed SNP data should be considered.

Keywords

miRNA processing; binding sites; inherited susceptibility; ovarian cancer; genetic variants

Introduction

MicroRNAs (miRNAs) are short, non-coding RNAs that regulate translation (1). SNPs in precursor and mature miRNAs, their processing machinery, or in miRNA binding sites of target genes have been implicated in cancer risk (2). Liang et al. (3) analyzed 238 SNPs from 8 miRNA processing genes and 138 genes containing potential miRNA binding sites in 339 EOC cases and 349 controls self-reported to be Caucasian, and identified associations between EOC risk and 13 SNPs from 4 processing genes (*DDX20, DROSHA/RNASEN, GEMIN4, XPO5)* and 7 binding site genes (*ATG4A, CAV1, COL18A1, E2F2, IL1R1, KRAS,* and *UGT2A3)*. We (4) genotyped 318 SNPs in 18 miRNA processing genes in 2,172 EOC

cases and 3,052 controls of European ancestry, and identified 6 SNPs from 4 genes (*DROSHA, FMR1, LIN28, LIN28B*) as significantly associated with EOC risk. Here we conducted a pooled analysis of variants reported as risk-associated by Liang et al (3) in 3,973 cases and 3,276 controls from the international Ovarian Cancer Association Consortium (OCAC) (5). We imputed SNPs to expand coverage of genes and regions, totaling 249 SNPs from 10 of the 11 highlighted genes (3).

Material and Methods

Participating OCAC studies were from North America (US-CAN), the United Kingdom (UK), and Poland (POL). Study characteristics have been reported (4) and are summarized in Table 1. Briefly, cases had pathologically-confirmed primary invasive EOC. Controls had at least one ovary intact when interviewed. All studies collected data on disease status, self-reported ethnicity, and histologic subtype. Subjects with <80% European ancestry were excluded (4), and the first two principal components (PCs) representing European ancestry were estimated for all SNPs with call rates >99% using Golden Helix SVS PCA function, algorithmically equivalent to EigenSTRAT. The protocol was approved by the institutional review board at each site, and all participants provided written informed consent. Pooled data included 3,973 cases (51% serous) and 3,276 controls.

SNP genotyping and quality control have been described (4, 6). SNP imputation was carried out within studies (US-CAN, UK, POL) with MACH version 1.0.16 using CEU phased data from HapMap release 22 (genome build 36). We imputed data for 186 SNPs that span 15 kb upstream and downstream of each miRNA processing gene or reside in a putative miRNA binding site in the 3′ UTR of target genes as predicted by SNPInfo (7) and/or PolymiRTS (8); the remaining 63 SNPs were directly genotyped.

Study-specific odds ratios (OR) and 95% confidence intervals (CI) were estimated using unconditional logistic regression. Log-additive genetic models were fit for each SNP, modeling the number of copies of the minor allele. For imputed SNPs, we used expected counts of minor alleles obtained from MACH. Study-specific estimates were adjusted for age at diagnosis/interview (US-CAN, POL), component study sites (US-CAN), and the first two PCs (US-CAN, UK, POL). Allele frequencies across studies were similar, suggesting low genetic heterogeneity between populations and appropriateness for combining data. Pooled estimates were adjusted for a) study (US-CAN, UK, POL) and b) study and the first two PCs. We used PLINK for statistical analysis (10).

Results

Two hundred twenty-six SNPs were evaluated within or near miRNA processing genes *DDX20* (n=17), *DROSHA* (n=179), *GEMIN4* (n=11), and *XPO5* (n=19). Table 2 displays association results for the 6 processing SNPs (or their tagSNPs) identified by Liang et al. (3); none were risk-associated. Of all other miRNA processing SNPs evaluated, only 3 *DROSHA* SNPs were associated with risk ($P<0.05$) when accounting for study site only, but none retained statistical significance after further adjustment for ancestry (See Supplemental Table 1).

There were 23 SNPs predicted to disrupt miRNA binding within 6 of the 7 candidate genes (3). We did not evaluate SNPs within *ATG4A* because neither genotype nor imputed data were available for SNPs within the 3′ UTR. Table 2 shows results from the 6 binding site SNPs (or their tagSNPs) identified by Liang et al. (3). To minimize redundancy due to tagSNPs, results from 21 of the 23 binding site SNPs evaluated are displayed in Supplemental Table 1. Only one previously-identified binding site SNP, *CAV1* rs9920 (3),

and two imputed *CAV1* SNPs (rs1049314 and rs8713) were associated with risk in the pooled, study site-adjusted analysis (Table 2; Supplemental Table 1). However, none of these *CAV1* SNPs were risk-associated after further adjustment for ancestry.

Study-specific estimates were generally similar across studies, and results did not change appreciably when considering a dominant genetic model or serous-only histology (data not shown).

Discussion

We did not detect consistent associations between the majority of previously-identified polymorphisms (3) and EOC risk. Although we did identify associations between EOC risk and 3 SNPs flanking the 3′UTR of *DROSHA* and 3 SNPs in miRNA binding sites of *CAV1*, none retained statistical significance after controlling for European ancestry. Consistent with recent large-scale (11) but not smaller studies (3, 12), we did not identify associations between EOC risk and SNPs in miRNA binding sites of *KRAS*.

Several explanations exist for not replicating the findings presented by Liang et al. (3). First, our analysis suggests their results may be confounded by population admixture, underscoring the importance of estimating population stratification rather than relying on self-reported ancestry in genetic association studies. Due to their relatively small sample size (3), chance is an alternate explanation for their findings. Our pooled sample had at least 90% statistical power to detect a SNP with a minor allele frequency of 0.09 and a log-additive OR of 1.2. This analysis highlights the importance of having large studies and/or combining genotype data from multiple studies to increase statistical power to detect true associations, and demonstrates the utility of population stratification and imputation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all of the individuals who participated in this research along with all of the researchers, clinicians, and staff who have contributed to the participating studies. The genotyping, bioinformatic and biostatistical data analysis for MAY, NCO, and TOR was supported by R01-CA-114343 and R01-CA114343-S1. The MAY study is supported by R01-CA-122443 and P50-CA-136393 and funding from the Mayo Foundation. The NCO study is supported by R01-CA-76016. The TBO study is supported by R01-CA-106414, the American Cancer Society (CRTG-00-196-01-CCE), and the Advanced Cancer Detection Center Grant, Department of Defense (DAMD-17-98-1-8659). The TOR study is supported by grants from the Canadian Cancer Society and the National Institutes of Health (R01-CA-63682 and R01-CA-63678). The Mayo Clinic Genotyping Shared Resource is supported by the National Cancer Institute (P30-CA-15083). The NEC study is supported by grants CA-54419 and P50 CA105009.The POL study was supported by the Intramural Research Program of the NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics, and the Center for Cancer Research. The SEA study is funded by a program grant from Cancer Research UK. The UKO study is supported by funding from Cancer Research UK, the Eve Appeal, and the OAK Foundation; some of this work was undertaken at UCLH/UCL who received some funding from the Department of Health's NIHR Biomedical Research Centre funding scheme. UK genotyping and data analysis was supported by a project grant from Cancer Research UK. UK studies also make use of data generated by the Wellcome Trust Case-Control consortium. A list of investigators who contributed to the generation of data is available at www.wtccc.org.uk.

References

1. Li SD, Zhang JR, Wang YQ, Wan XP. The role of microRNAs in ovarian cancer initiation and progression. J Cell Mol Med. 2010
2. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer. 2010; 10:389–402. [PubMed: 20495573]

3. Liang D, Meyer L, Chang DW, Lin J, Pu X, Ye Y, et al. Genetic variants in MicroRNA biosynthesis pathways and binding sites modify ovarian cancer risk, survival, and treatment response. Cancer Res. 2010; 70:9765–76. [PubMed: 21118967]
4. Permuth-Wey J, Kim D, Tsai YY, Lin HY, Chen YA, Barnholtz-Sloan J, et al. LIN28B polymorphisms influence susceptibility to epithelial ovarian cancer. Cancer Res. 2011
5. Berchuck A, Schildkraut JM, Pearce CL, Chenevix-Trench G, Pharoah PD. Role of genetic polymorphisms in ovarian cancer susceptibility: development of an international ovarian cancer association consortium. Adv Exp Med Biol. 2008; 622:53–67. [PubMed: 18546618]
6. Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nat Genet. 2009; 41:996–1000. [PubMed: 19648919]
7. Xu, Z.; Taylor, JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies; Nucleic Acids Res. 2009. p. W600-5.Available at: http://manticore.niehs.nih.gov/
8. Bao, L.; Zhou, M.; Wu, L.; Lu, L.; Goldowitz, D.; Williams, RW., et al. PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits; Nucleic Acids Res. 2007. p. D51-4Available at http://compbio.uthsc.edu/miRSNP
9. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]
10. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, MA.; Bender, D., et al. PLINK: a tool set for whole-genome association and population-based linkage analyses; Am J Hum Genet. 2007. p. 559-75.Available at: http://pngu.mgh.harvard.edu/~purcell/plink/
11. Pharoah PD, Palmieri RT, Ramus SJ, Gayther SA, Andrulis IL, Anton-Culver HA, et al. The role of KRAS rs61764370 in invasive epithelial ovarian cancer: implications for clinical testing. Clin Cancer Res. 2011
12. Ratner E, Lu L, Boeke M, Barnett R, Nallur S, Chin LJ, et al. A KRAS-variant in ovarian cancer acts as a genetic marker of cancer risk. Cancer Res. 2010; 70:6509–15. [PubMed: 20647319]

Table 1

Characteristics of Participating Studies of Epithelial Ovarian Cancer

Study Name (Abbreviation)	Study Population	Genotyping Platform	Study Type	Number of subjects[1]	
				cases	controls
North America (US-CAN)					
Mayo Clinic Ovarian Cancer Study (MAY)	Upper Midwest, USA	Illumina 610K	Clinic based	359	520
North Carolina Ovarian Cancer Study (NCO)	North Carolina, USA	Illumina 610K	Population based	494	654
Tampa Bay Ovarian Cancer Study (TBO)	Tampa, USA	Illumina 610K	Population based	227	169
Familial Ovarian Tumor Study (TOR)	Ontario, Canada	Illumina 610K	Population based	734	524
New England Case-Control Study of Ovarian Cancer (NEC)	New England, USA	Illumina 317K, 370K	Population based	133[2]	142
US/CAN Subtotal				**1947**	**2009**
United Kingdom (UK)					
SEARCH (SEA)	England	Illumina 610K	Population based	1118	-
United Kingdom Ovarian Cancer Population Study (UKO)	England	Illumina 610K	Population based	506	-
Cancer Research UK Familial Ovarian Cancer Register (FOCR)	England	Illumina 610K	Familial Cancer Register	44	-
Royal Marsden Hospital Study (RMH)	England	Illumina 610K	Hospital based	146	-
UK 58 Birth Cohort (58 BC)	England, Wales, Scotland	Illumina 550K	Cohort	-	712
UK Subtotal				**1814**	**712**
Poland (POL)					
Polish Ovarian Cancer Study (POL)	Warsaw and Lodz, Poland	Illumina 660w	Population based	**212**	**555**
OVERALL TOTAL				**3973**	**3276**

[1] Totals represent the number of non-Hispanic white Europeans passing genotyping quality control criteria and meeting study site-specific inclusion/exclusion criteria.

[2] Cases from NEC that were evaluated as part of this investigation represent postmenopausal advanced papillary serous carcinomas; 26 of these cases were ascertained as part of a hospital-based pre-operative study

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2012 February 1.

Table 2

Association between selected miRNA processing and miRNA binding site SNPs and epithelial ovarian cancer risk in a pooled analysis.

Gene (locus)	SNP (maj/min allele[a])	Location (putative miRs) [b]	OR (95% CI) reported by Liang et al (Ref. 3)	MAF [c]	Pooled OR (95% CI), adjusted for study [d]	*P*	Pooled OR (95% CI), adjusted for study and ancestry [e]	*P*
miRNA processing								
DDX20 (1p21,1-p13.2)	rs197414 (C/A) [f]	Missense	0.69 (0.48–0.99)	0.13	1.02 (0.92,1.12)	0.70	1.04 (0.94,1.15)	0.49
DROSHA (5p13.3)	rs9292427 (C/T)[g]	Intron	0.71 (0.51–0.99)	0.46	1.01 (0.95,1.08)	0.72	1.01 (0.94,1.08)	0.79
GEMIN4 (17p13)	rs2740349 (A/C) [h]	exon 1, ns	0.70 (0.51–0.96)	0.18	0.99 (0.92,1.09)	0.97	1.02 (0.93,1.11)	0.71
	rs2740351 (T/C) [i]	flanks 5′UTR	0.71 (0.57–0.87)	0.45	0.98 (0.91,1.04)	0.46	1.00 (0.94,1.07)	0.98
	rs7813 (T/G) [i]	exon 1, ns	0.71 (0.57–0.88)	0.46	0.97 (0.91,1.04)	0.38	1.00 (0.93,1.07)	0.91
XPO5 (6p21.1)	rs2257082 (C/A)	exon 1, ss	0.73 (0.54–0.99)	0.27	0.99 (0.92,1.07)	0.87	1.00 (0.93,1.08)	0.95
miRNA binding sites								
CAV1 (7q31.1)	rs9920 (G/A)	3′UTR (miR 630)	1.50 (1.04–2.17)	0.10	1.13 (1.10,1.26)	**0.03**	1.06 (0.95,1.19)	0.29
COL18A1 (21q22.3)	rs7499 (G/A)	3′UTR (miR-594)	1.47 (1.07–2.02)	0.42 [c]	0.98 (0.92,1.05)	0.57	0.98 (0.92,1.05)	0.50
E2F2 (1p36)	rs2075993 (A/C) [j]	3′UTR (miR-663,486-3p)	1.24 (1.00–1.54)	0.48	1.01 (0.95,1.08)	0.67	1.01 (0.94,1.08)	0.87
ILIR1 (2q12)	rs3917328 (C/T)	3′UTR (miR-335, 31)	1.65 (1.03–2.64)	0.05 [c]	1.06 (0.91,1.23)	0.49	1.00 (0.86,1.17)	0.99
KRAS (12p12.1)	rs13096 (A/G) [k]	3′UTR (miR-1244)	1.26 (1.01–1.57)	0.45	1.00 (0.94,1.07)	0.94	0.99 (0.93,1.06)	0.85
UGT2A3 (4q13.2)	rs17147016 (T/A) [h]	3′UTR (miR-224, 1279)	1.47 (1.08–2.01)	0.19 [c]	1.02 (0.93,1.11)	0.70	1.01 (0.93,1.10)	0.88

Abbreviations: US-CAN=United States-Canada; UK=United Kingdom; POL=Poland; maj=major; min=minor; miR=miRNA; UTR= untranslated region; ns=non- synonymous SNP; ss=synonymous SNP; OR (CI) =odds ratio (confidence interval); MAF=minor allele frequency among all controls; all *P*-values are two-sided.

[a] The major allele represents the most frequently-occurring allele and serves as the reference allele during modeling.

[b] SNP location derived from Illumina annotation files, HapMap2 data (http://hapmap.ncbi.nlm.nih.gov/), and dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/). SNPinfo http://snpinfo.niehs.nih.gov/ and the PolymiRTS database (http://compbio.uthsc.edu/miRSNP) were used to predict miRNAs whose binding activity may be altered due to the SNP location.

[c] Genotype data was imputed for all participants using MACH version 1.0.16 using phased data from HapMap release 22 (genome build 36) derived from individuals with European ancestry (CEU).

[d] Pooled OR and 95% CI estimated using a log-additive model adjusted for study (US-CAN, UK, POL)

[e] Pooled OR and 95% CI estimated using a log-additive model adjusted for study and the first two principal components representing European ancestry

[f] *DDX20* rs19714 is in linkage disequilibrium (LD) (r^2 =0.90) with rs197383 identified by Liang et al.

[g] *DROSHA* rs9292427 is in LD (r^2 =0.98) with rs4867329 identified by Liang et al.

[h] SNP deviates from Hardy Weinberg Equilibrium among all controls with P_{HWE} values of 0.020 for rs607613, 0.040 for rs615435, 0.013 for rs2740349, 0.004 for rs3732133, and 0.034 for rs17147016, respectively.

[i] *GEMIN4* SNP pair in LD (r^2 =1)

[j] *E2F2* SNP pair in LD (r^2 =0.97)

[k] *KRAS* rs13096 is in LD (r^2 =1) with rs10771184 identified by Liang et al.