Germline copy number variation and ovarian cancer survival

Brooke L. Fridley[1][†], Prabhakar Chalise[2][†], Ya-Yu Tsai[3], Zhifu Sun[1], Robert A. Vierkant[1], Melissa C. Larson[1], Julie M. Cunningham[4], Edwin S. Iversen[5], David Fenstermacher[6], Jill Barnholtz-Sloan[7], Yan Asmann[1], Harvey A. Risch[8], Joellen M. Schildkraut[9], Catherine M. Phelan[3], Rebecca Sutphen[10], Thomas A. Sellers[3] and Ellen L. Goode[1]*

[1] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA
[2] Biostatistics Department, University of Kansas Medical Center, Kansas City, KS, USA
[3] Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA
[4] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA
[5] Department of Statistical Science, Duke University, Durham, NC, USA
[6] Department of Biomedical Informatics, Moffitt Cancer Center, Tampa, FL, USA
[7] Case Comprehensive Cancer Center, Case School of Medicine, Cleveland, OH, USA
[8] Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, USA
[9] Department of Community and Family Medicine, Duke University Medical Center, Durham, NC, USA
[10] Pediatrics Epidemiology Center, College of Medicine, University of South Florida, Tampa, FL, USA

Copy number variants (CNVs) have been implicated in many complex diseases. We examined whether inherited CNVs were associated with overall survival among women with invasive epithelial ovarian cancer. Germline DNA from 1,056 cases (494 deceased, average of 3.7 years follow-up) was interrogated with the Illumina 610 quad genome-wide array containing, after quality control exclusions, 581,903 single nucleotide polymorphisms (SNPs) and 17,917 CNV probes. Comprehensive analysis capitalized upon the strengths of three complementary approaches to CNV classification. First, to identify small CNVs, single markers were evaluated and, where associated with survival, consecutive markers were combined. Two chromosomal regions were associated with survival using this approach (14q31.3 rs2274736 $p = 1.59 \times 10^{-6}$, $p = 0.001$; 22q13.31 rs2285164 $p = 4.01 \times 10^{-5}$, $p = 0.009$), but were not significant after multiple testing correction. Second, to identify large CNVs, genome-wide segmentation was conducted to characterize chromosomal gains and losses, and association with survival was evaluated by segment. Four regions were associated with survival (1q21.3 loss $p = 0.005$, 5p14.1 loss $p = 0.004$, 9p23 loss $p = 0.002$, and 15q22.31 gain $p = 0.002$); however, again, after correcting for multiple testing, no regions were statistically significant, and none were in common with the single marker approach. Finally, to evaluate associations with general amounts of copy number changes across the genome, we estimated CNV burden based on genome-wide numbers of gains and losses; no associations with survival were observed ($p > 0.40$). Although CNVs that were not well-covered by the Illumina 610 quad array merit investigation, these data suggest no association between inherited CNVs and survival after ovarian cancer.

Keywords: association testing, copy number variation, genotyping array, ovarian cancer, overall survival

INTRODUCTION

Epithelial ovarian cancer is the fifth leading cause of cancer death among women in the United States, accounting for 5% of cancer deaths (Jemal et al., 2011). Though rare, most patients are diagnosed with advanced disease due to the non-specific nature of early symptoms and lack of effective screening strategies. For the approximate 75% of women diagnosed with stage III or IV disease, the likelihood of long term disease-free survival ranges from 15 to 20% (Hoskins et al., 1992; McGuire et al., 2002; Barnholtz-Sloan et al., 2003). As women may inherently vary in their ability to eradicate disease or tolerate treatment, genetic association studies have sought to identify inherited variants related to overall survival. In 2010, the first ovarian cancer genome-wide association study (GWAS) to examine survival did not identify any replicated survival-associated single nucleotide polymorphisms (SNPs; Bolton et al., 2010).

Like SNPs, copy number variants (CNVs) occur commonly in the genome and have been implicated in risk of complex diseases including schizophrenia (Need et al., 2009), neuroblastoma (Diskin et al., 2009), and prostate cancer (International Schizophrenia Consortium, 2008; Liu et al., 2009). The size and frequency of the detected CNV (deletions) from these studies varied, with size of deletions ranging from intermediate (4 kb) to large (2 Mb). CNVs are a priori more likely to have larger phenotypic effects

than SNPs (Cooper et al., 2008), and they have been shown to have adequate coverage on current SNP arrays, at least for large and intermediate size CNVs (CNVs > 5 kb; McCarroll, 2008). Therefore, to better understand inherited factors in ovarian cancer, we used the Illumina 610 quad array to characterize CNVs and evaluate associations with survival among over 1,000 ovarian cancer cases.

MATERIALS AND METHODS

STUDY POPULATION
Participants were from three previously described studies of invasive epithelial ovarian cancer which enrolled cases from 2000 to 2008 from the upper Midwest [the Mayo Clinic Ovarian Cancer Study (MAY)], North Carolina [the North Carolina Ovarian Cancer Study (NCO)], and the Tampa Bay, FL, USA area (the Tampa Bay Ovarian Cancer Study, TBO; Permuth-Wey et al., 2011). Research protocols were approved by institutional review boards at each site, and all participants provided written informed consent. Cases were followed for vital status through 2009 using active contact, medical record review, and linkage to the National Death Index.

GENOTYPING AND QUALITY CONTROL
DNA extracted from blood were genotyped at the Mayo Clinic Genotyping Shared Resource Facility (Rochester, MN, USA) using the Illumina Infinium 610 quad array as described previously (Permuth-Wey et al., 2011). Genotyping was attempted for 4,169 samples including ineligible participants (population controls, case without follow-up data, laboratory controls). Samples with call rate <95%, ambiguous gender, unresolved identical genotypes, self-reported non-Caucasian race, or less than 80% European ancestry as predicted by structure (Pritchard et al., 2000) analysis were excluded. This resulted in a sample size of 3,715, including 1,056 for the current analysis. Markers on the X chromosome or with call rate <80% were excluded, leaving 599,820 markers (581,903 SNPs, 17,917 CNV probes) for CNV analysis.

NORMALIZATION OF INTENSITY DATA
Normalization of allelic intensities was completed similar to the approach used by Barnes et al. (2008). Let A_{ij} and B_{ij} represent the intensities of the two alleles for marker j measured on subject i. First, systematic differences between the A and B intensities were corrected using data for subjects with heterozygous genotype calls, with correction factor for marker j equal to $\psi_j = \left(\frac{1}{K}\right)\sum_{i=1}^{K}\left(\frac{A_{ij}}{B_{ij}}\right)$, with K representing the number of subjects with heterozygous genotype calls for marker j. This correction factor was then used to compute the total intensity for marker j measured on subject i, $I_{ij} = \log_2(A_{ij} + \Psi_j B_{ij})$.

Next, a two-step normalization procedure was completed to remove plate and other experimental artifacts using quantile normalization to produce similar intensity distributions for each subject, followed by a median normalization for each marker by plate. Using normalized intensity values, $Z_{ij} = I_{ij} - \bar{I}_i$ was computed for each marker j measured on subject i, where $\bar{I}_i$ is the mean intensity for subject i computed over all the markers (mean intensity for the subject).

PARAMETERIZATION OF CNVs
In order to comprehensively address the genetic architecture of CNV association with ovarian cancer survival, CNVs were characterized using three complementary approaches, as simulation showed different approaches to have maximal statistical power under varied genetic models (Breheny et al., 2012).

First, to identify associations with small, common CNVs, normalized intensities at individual markers (Z) were evaluated and, where associated with survival (association methods described below), results across neighboring genetic loci were combined (Ionita-Laza et al., 2008). To combine results at consecutive markers, we performed fused lasso regression of $-\log_{10}$ (p-values; Tibshirani et al., 2005), using the R package *cghFLasso*.[1] In regions of interest [smooth $-\log_{10}$ (p-value) >2], chromosomal segmentation was completed using Partek® Discovery Suite™ (version 6.3) to more precisely define the associated CNV boundaries (parameter settings: minimum of five markers, p-value threshold = 0.001, signal to noise ratio = 0.3, region below = −0.3, region above = 0.15).

Second, to identify associations with large, rare CNVs, genome-wide segmentation was conducted to define chromosomal gains and losses, and association with survival was evaluated by segment. The number of copies present at each chromosomal segment for each individual was estimated based on normalized intensities (Z; i.e., CNVs were "called") using circular binary segmentation (CBS) methods of the R package *DNAcopy* with a three marker minimum[2] (Venkatraman and Olshen, 2007). The R package *cghMCR* was used to determine common CNV regions and combine adjacent regions[3] (Aguirre et al., 2004; lower threshold = 3rd percentile, upper threshold = 97th percentile, required recurrence = five). At every segment for which copy number was variable, association testing was conducted (association methods described below).

Finally, because markers or regions may not individually predict ovarian cancer survival, we also evaluated the general amount of copy number changes across the genome (i.e., CNV burden; Kathiresan et al., 2009). Genome-wide CNV burden for each participant was defined based on CBS of normalized intensities (Z) in three ways: the total number of gains or losses, the total number of gains, and the total number of losses. Association testing was then conducted for these three values (association methods described below). However, a limitation of this burden analysis approach is that the sizes of the CNVs are not taken into account during the analysis, as estimation of CNV breakpoints (and sizes) based GWAS data is difficult.

ASSOCIATION TESTING
Association with overall survival was completed using Cox proportional hazards regression, accounting for left truncation, with estimation of hazard ratios (HRs) and 95% confidence intervals (CIs; Therneau and Grambsch, 2000). Time at risk was defined as date of diagnosis to death with censoring at last follow-up. As described above, the CNV variable of interest was either the

---

[1]http://cran.r-project.org/web/packages/cghFLasso/
[2]http://www.bioconductor.org/packages/release/bioc/html/DNAcopy.html
[3]http://www.bioconductor.org/packages/release/bioc/html/cghMCR.html

normalized intensity at each marker (Z), the number of copies of certain chromosomal sections, or one of three measures of CNV burden. To control for population stratification, the first principal component from eigen-analysis of non-Hispanic white participants (Price et al., 2006; Permuth-Wey et al., 2011) was included as a covariate along with study site and age at diagnosis.

RESULTS

With an average 3.7 years of follow-up, 494 recurrences or deaths were observed among 1,056 successfully genotyped invasive epithelial ovarian cancer cases with follow-up for disease outcome. As shown in **Table 1**, 62% of cases were of serous histology, 71% were diagnosed at advanced stage, and the majority of subjects were enrolled within 4 months of diagnosis.

Analysis of individual normalized marker intensities followed by combination of results across multiple consecutive markers is the most powerful approach for the detection of associations with small, common CNVs. Two regions showed suggestive association with ovarian cancer survival at multiple markers [smoothed $-\log_{10}(p) > 2$]. On 14q31.3 (379 kb, 63 markers) a smoothed $p = 0.001$ was observed suggesting a modest regional association with survival. As shown in **Figure 1A**, normalized intensities at rs2274736, a non-synonymous SNP in *PTPN21*, alone appeared to be driving the regional association. In fact, this SNP was the most significant single marker in genome-wide analysis
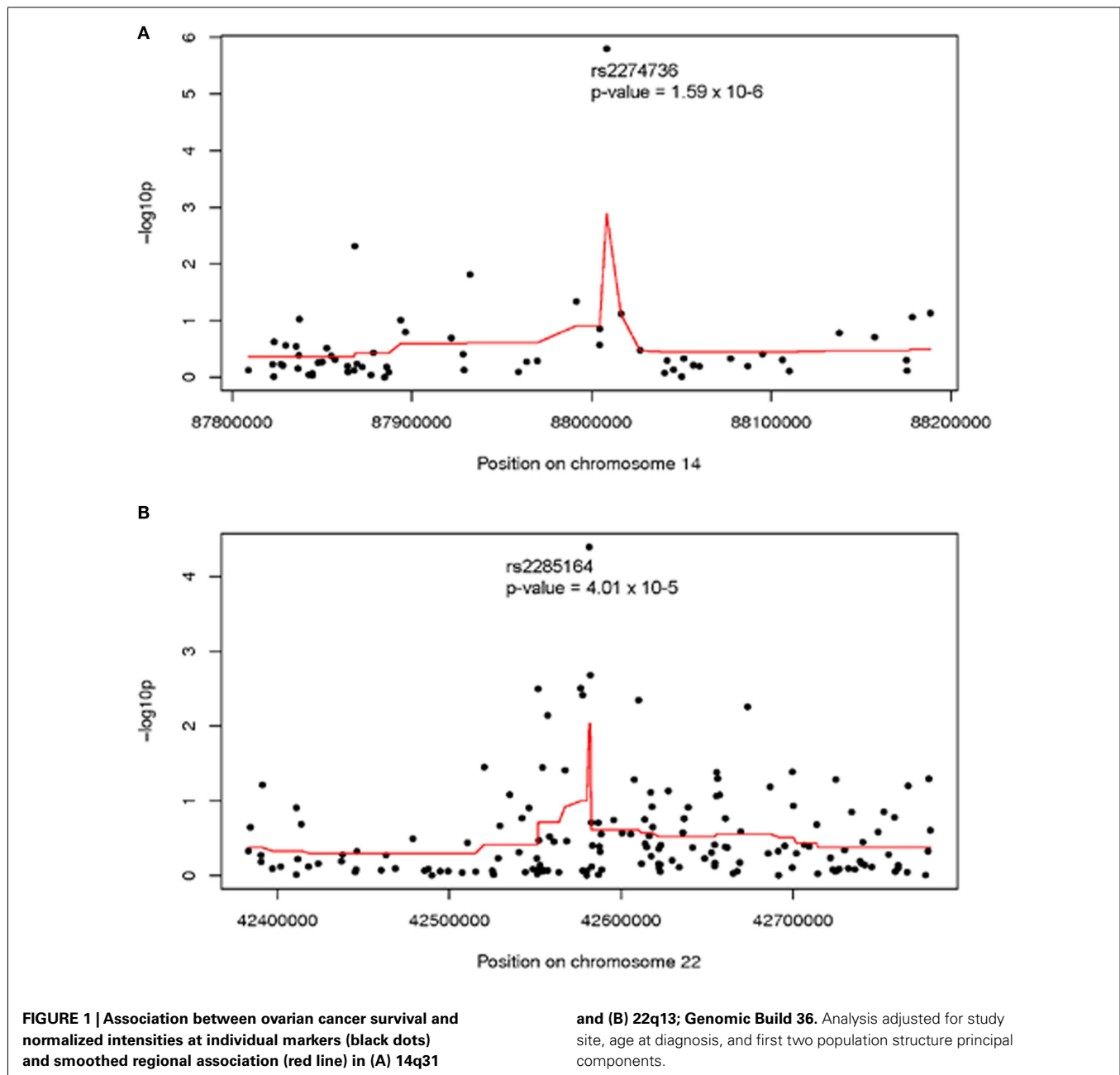
($p = 1.6 \times 10^{-6}$); note, however that it did not reach traditional genome-wide significance. Genomic segmentation in 14q31.3 was then done to identify specific gains or loss among study participants; however, only two samples were detected with gains and 18 samples with losses. Due to the lack of called CNVs in this region, further analysis was not carried out in 14q31.3.

A second region with suggestive association with cancer survival at multiple markers was on 22q13.31 centered at rs2285164 (smoothed $p = 0.009$, 397 kb, 160 markers; **Figure 1B**). Genomic segmentation of 22q13 was then computed to identify specific gains or loss among study participants (Inc, 2008), where 35 showed gain and 154 showed loss. Association testing of gain ($N = 35$), normal ($N = 867$), or loss ($N = 154$) with survival did not reveal association ($p = 0.29$ for two degrees of freedom test; treating CNV as a categorical variable; $p = 0.67$ for one degree of freedom trend test treating CVN as a continuous variable). Thus, even though a signal was observed for association from the single marker analysis on 22q13, CNV calling, and subsequent analysis of this region showed no association between CNV and overall survival (**Figure 2**).

An inverse approach analyzing pre-defined regions of CNV change is most powerful for the detection of associations with large, rare CNVs. Genome-wide CBS identified 564 regions with variable copy number among the study population, including 78 regions with gain and 486 regions with loss (available upon

Table 1 | Clinical characteristics of invasive epithelial ovarian cancer cases.

	MAY (*N*=352)	NCO (*N*=492)	TBO (*N*=212)	Total (*N*=1,056)
	N (%)	*N* (%)	*N* (%)	*N* (%)
Deceased	171 (49)	229 (47)	94 (44)	494 (47)
STAGE (FIGO)				
I	67 (19)	125 (25)	36 (17)	228 (22)
II	26 (7)	36 (7)	16 (8)	78 (7)
III	199 (57)	314 (64)	139 (66)	652 (62)
IV	56 (16)	15 (3)	21 (10)	92 (9)
Missing	4 (1)	2 (0)	0 (0)	6 (1)
GRADE				
I	13 (4)	65 (13)	15 (7)	93 (9)
II	45 (13)	137 (28)	38 (18)	220 (21)
III	288 (82)	278 (57)	159 (75)	725 (69)
Missing	6 (2)	12 (2)	0 (0)	18 (2)
HISTOLOGY				
Serous	221 (63)	293 (60)	138 (65)	652 (62)
Endometrioid	71 (20)	81 (16)	27 (13)	179 (17)
Clear Cell	24 (7)	58 (12)	11 (5)	93 (9)
Mucinous	11 (3)	20 (4)	12 (6)	43 (4)
Mixed/Other	25 (7)	40 (8)	19 (9)	84 (8)
Missing	0 (0)	0 (0)	5 (2)	5 (0)
	Mean (range)	Mean (range)	Mean (range)	Mean (range)
Age at diagnosis	61.4 (28–91)	57.1 (22–74)	61.5 (26–93)	59.4 (22–93)
Days from diagnosis to enrollment	35.5 (0–835)	102.7 (5–1,178)	99.3 (0–947)	79.6 (0–1,178)
Years from diagnosis to last follow-up	3.3 (0.02–9.5)	4.6 (0.5–10.0)	2.5 (0.02–8.3)	3.7 (0.02–10.0)

FIGURE 1 | Association between ovarian cancer survival and normalized intensities at individual markers (black dots) and smoothed regional association (red line) in (A) 14q31 **and (B) 22q13; Genomic Build 36.** Analysis adjusted for study site, age at diagnosis, and first two population structure principal components.

request). Association testing of these regions revealed 14 regions with p-values <0.05, including one region of gain and 13 regions of loss (**Table 2**). Results at the most statistically significant regions ($p = 0.002$) suggested that loss of a region on 9p23 was associated with poorer survival (HR 1.44, 95% CI 1.14–1.81) as was gain of a region on 15q22.31 (HR 1.34, 95% CI 1.11–1.61). However, no region was statistically significant after correction for multiple testing using a Bonferroni procedure (564 tests).

Because the overall amount of variation from normal copy number across the genome (CNV burden) may contribute to disease, CNV burden for each case was estimated as summarized in **Table 3**. There was no association between survival and number

of gains ($p = 0.42$), number of losses ($p = 0.94$), or total number of gains and losses ($p = 0.84$).

DISCUSSION

Single nucleotide polymorphisms GWASs have yielded great insights into the etiology several complex diseases including ovarian cancer[4] (Song et al., 2009; Bolton et al., 2010). However, germline variation related to disease outcome has been more elusive, with the possible exception of acute lymphocytic leukemia.[5] With much unexplained variation in ovarian cancer outcome,

[4]http://www.nature.com/ng/journal/v42/n10/full/ng.668.html
[5]http://www.nature.com/ng/journal/v43/n3/full/ng.763.html

investigation of additional inherited factors remains warranted. Here, in the first germline ovarian cancer CNV analysis, we have harnessed CNV data gleaned from dense genome-wide SNP genotyping in order to address the hypothesis that copy number variation is associated with survival. With over 1,000 cases (almost 500 deaths) representing three study populations, results were largely null suggesting no strong associations between small CNVs, large CNVs, or general CNV burden and ovarian cancer outcome.

There are numerous strengths to this study. First, the Illumina Infinium 610 quad array included markers in known

FIGURE 2 | Kaplan–Meier plot of the 22q13 CNV association with ovarian cancer survival. The different lines represent the survival curve for subjects with a "loss," "gain," or "normal" for the genomic segment.

regions according to the Toronto Database of Genomic Variation with approximately 38 markers per CNV region, markers in the "unSNPable" genome, markers in novel CNV regions, and intensity-only probes resulting in an average (median) coverage of 93% (100%) for white non-Hispanic populations.[6] Second, our analysis leveraged a large sample size providing power to detect a HR of 1.75 for a CNV with frequency of 20% at a significance level of 10^{-5} (or HR = 2.00 at $p < 10^{-7}$). Third, based on the available SNP data, assessment, and correction for possible population stratification to eliminate spurious associations due to confounding was performed. Fourth, extensive normalization of the intensity data from the SNP array was completed in order to produce reliable CNV calls. Finally, three complementary approaches for assessment and testing association between CNVs and ovarian cancer survival were utilized.

However, we cannot rule out more modest associations (HR < 2.0) between CNVs and ovarian cancer survival, and we cannot rule out associations with very small CNVs (<10 kb) which were not covered in Infinium 610 quad array. Despite this study being the largest of its kind to determine germline CNVs associated survival, the sample size was still relatively small and restricted to subjects of European ancestry. Finally, it is possible that CNVs exhibit survival effects only in the context of certain chemotherapeutic regimens or only among certain subtypes; these analyses are important and may reveal associations masked in the current analysis.

In conclusion, this study did not detect any CNVs associated with ovarian cancer survival using the data from the Infinium 610 quad array in three populations. Future research may reveal whether germline CNVs play a role in ovarian cancer etiology or in outcome within certain clinical contexts.

---
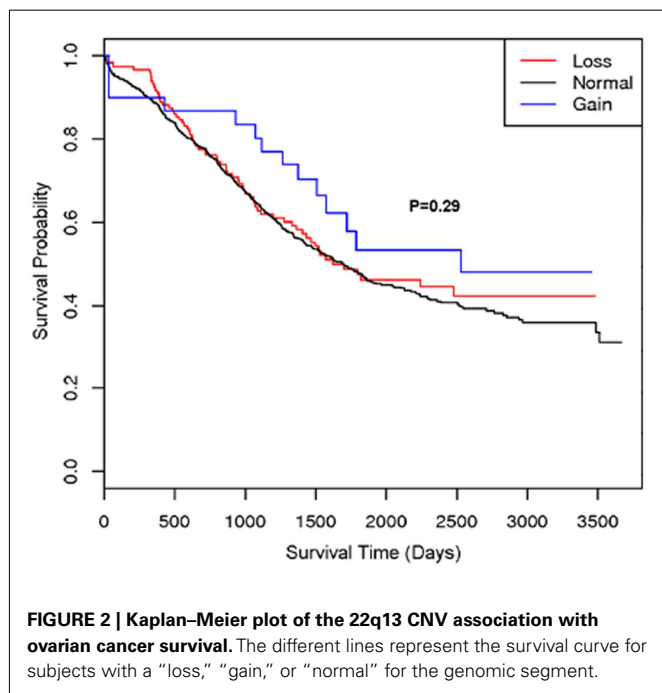
[6]www.illumina.com

Table 2 | Association between CNV region and ovarian cancer survival ($p < 0.05$).

Chr	Location (bp)	*N* Gain	*N* Normal	*N* Loss	HR (95% CI)	*p*-Value
1p31.1	72,538,673–72,549,855	0	992	64	1.52 (1.08–2.15)	0.017
1q21.3	151,026,302–151,033,105	0	947	109	1.46 (1.12–1.89)	0.005
4p15.31	19,130,834–19,131,054	0	994	62	1.48 (1.03–2.11)	0.032
5p14.1	27,462,485–27,462,654	0	975	81	1.54 (1.15–2.07)	0.004
6p21.33	30,017,499–30,017,538	0	1001	55	1.44 (1.01–2.06)	0.043
6p21.32	32,756,221–32,758,787	0	996	60	1.44 (1.01–2.05)	0.045
6q24.1	141,045,394–141,083,022	0	969	87	1.45 (1.07–1.94)	0.015
7q11.21	65,415,344–65,417,964	0	948	108	1.35 (1.03–1.76)	0.031
7q31.1	109,237,359–109,238,466	0	940	116	1.34 (1.04–1.74)	0.026
8q13.3	70,795,078–70,795,945	0	893	163	0.75 (0.57–0.97)	0.031
9p23	11,398,647–11,398,865	0	906	150	1.44 (1.14–1.81)	0.002
12q12	38,668,966–38,671,005	0	969	87	1.43 (1.07–1.91)	0.015
12q23.1	95,745,302–95,745,411	0	963	93	0.62 (0.43–0.90)	0.013
15q22.31	63,701,811–63,710,800	339	717	0	1.34 (1.11–1.61)	0.002

HR, hazard ratio; CI, confidence interval adjusted for study site, age at diagnosis, and first two population structure principal components.

The start and end point based on the minimum common region determined from R package cghMCR (Aguirre et al., 2004), genomic build 36; normal number of copies is the reference group.

Table 3 | Mean genome-wide CNV burden by vital status and association with overall survival.

	Alive (*N*=562)	Deceased (*N*=494)	HR (95% CI)	*p*-Value
	Mean (SD)	Mean (SD)		
N Losses	46.96 (22.47)	47.98 (20.76)	0.9998 (0.9958–1.0039)	0.94
N Gains	11.85 (6.75)	12.51 (6.85)	1.0051 (0.9926–1.0179)	0.42
Total *N* Gains and Loses	58.80 (23.35)	60.49 (21.44)	1.0000 (0.9997–1.0002)	0.84

p-Value, hazard ratio (HR), and confidence intervals (CI) from hazards regression adjusted for study site, age at diagnosis, and first two population structure principal components.

REFERENCES

Aguirre, A. J., Brennan, C., Bailey, G., Sinha, R., Feng, B., Leo, C., Zhang, Y., Zhang, J., Gans, J. D., Bardeesy, N., Cauwels, C., Cordon-Cardo, C., Redston, M. S., Depinho, R. A., and Chin, L. (2004). High-resolution characterization of the pancreatic adenocarcinoma genome. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9067–9072.

Barnes, C., Plagnol, V., Fitzgerald, T., Redon, R., Marchini, J., Clayton, D., and Hurles, M. E. (2008). A robust statistical method for case-control association testing with copy number variation. *Nat. Genet.* 40, 1245–1252.

Barnholtz-Sloan, J. S., Schwartz, A. G., Qureshi, F., Jacques, S., Malone, J., and Munkarah, A. R. (2003). Ovarian cancer: changes in patterns at diagnosis and relative survival over the last three decades. *Am. J. Obstet. Gynecol.* 189, 1120–1127.

Bolton, K. L., Tyrer, J., Song, H., Ramus, S. J., Notaridou, M., Jones, C., Sher, T., Gentry-Maharaj, A., Wozniak, E., Tsai, Y. Y., Weidhaas, J., Paik, D., Van Den Berg, D. J., Stram, D. O., Pearce, C. L., Wu, A. H., Brewster, W., Anton-Culver, H., Ziogas, A., Narod, S. A., Levine, D. A., Kaye, S. B., Brown, R., Paul, J., Flanagan, J., Sieh, W., McGuire, V., Whittemore, A. S., Campbell, I., Gore, M. E., Lissowska, J., Yang, H. P., Medrek, K., Gronwald, J., Lubinski, J., Jakubowska, A., Le, N. D., Cook, L. S., Kelemen, L. E., Brook-Wilson, A., Massuger, L. F., Kiemeney, L. A., Aben, K. K., Van Altena, A. M., Houlston, R., Tomlinson, I., Palmieri, R. T., Moorman, P. G., Schildkraut, J., Iversen, E. S., Phelan, C., Vierkant, R. A., Cunningham, J. M., Goode, E. L., Fridley, B. L., Kruger-Kjaer, S., Blaeker, J., Hogdall, E., Hogdall, C., Gross, J.,

Karlan, B. Y., Ness, R. B., Edwards, R. P., Odunsi, K., Moyisch, K. B., Baker, J. A., Modugno, F., Heikkinenen, T., Butzow, R., Nevanlinna, H., Leminen, A., Bogdanova, N., Antonenkova, N., Doerk, T., Hillemanns, P., Durst, M., Runnebaum, I., Thompson, P. J., Carney, M. E., Goodman, M. T., Lurie, G., Wang-Gohrke, S., Hein, R., Chang-Claude, J., Rossing, M. A., Cushing-Haugen, K. L., Doherty, J., Chen, C., Rafnar, T., Besenbacher, S., Sulem, P., Stefansson, K., Birrer, M. J., Terry, K. L., Hernandez, D., Cramer, D. W., Vergote, I., Amant, F., Lambrechts, D., Despierre, E., Fasching, P. A., Beckmann, M. W., Thiel, F. C., Ekici, A. B., Chen, X., Australian Ovarian Cancer Study Group, Australian Cancer Study (Ovarian Cancer), Ovarian Cancer Association Consortium, Johnatty, S. E., Webb, P. M., Beesley, J., Chanock, S., Garcia-Closas, M., Sellers, T., Easton, D. F., Berchuck, A., Chenevix-Trench, G., Pharoah, P. D., and Gayther, S. A. (2010). Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nat. Genet.* 42, 880–884.

Breheny, P., Chalise, P., Batzler, A., Wang, L., and Fridley, B. L. (2012). Genetic association studies of copy-number variation: should assignment of copy number states precede testing? *PloS ONE* 7, e34262. doi:10.1371/journal.pone.0034262

Cooper, G. M., Zerr, T., Kidd, J. M., Eichler, E. E., and Nickerson, D. A. (2008). Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nat. Genet.* 40, 1199–1203.

Diskin, S. J., Hou, C., Glessner, J. T., Attiyeh, E. F., Laudenslager, M., Bosse, K., Cole, K., Mosse, Y. P., Wood, A., Lynch, J. E., Pecor, K., Diamond, M., Winter, C., Wang, K.,

Kim, C., Geiger, E. A., McGrady, P. W., Blakemore, A. I., London, W. B., Shaikh, T. H., Bradfield, J., Grant, S. F., Li, H., Devoto, M., Rappaport, E. R., Hakonarson, H., and Maris, J. M. (2009). Copy number variation at 1q21.1 associated with neuroblastoma. *Nature* 459, 987–991.

Hoskins, P. J., O'reilly, S. E., Swenerton, K. D., Spinelli, J. J., Fairey, R. N., and Benedet, J. L. (1992). Ten-year outcome of patients with advanced epithelial ovarian carcinoma treated with cisplatin-based multimodality therapy. *J. Clin. Oncol.* 10, 1561–1568.

Inc, P. (2008). "*Partek® Discovery Suite™ [or Partek® Genomics Suite™ or Partek® Screening Solution™ or Partek® QSAR Solution™]*," Revision 6.3 edn. St. Louis, MO: Partek Inc.

International Schizophrenia Consortium (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455, 237–241.

Ionita-Laza, I., Perry, G. H., Raby, B. A., Klanderman, B., Lee, C., Laird, N. M., Weiss, S. T., and Lange, C. (2008). On the analysis of copy-number variations in genome-wide association studies: a translation of the family-based association test. *Genet. Epidemiol.* 32, 273–284.

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. *CA Cancer J. Clin.* 61, 69–90.

Kathiresan, S., Voight, B. F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P. M., Anand, S., Engert, J. C., Samani, N. J., Schunkert, H., Erdmann, J., Reilly, M. P., Rader, D. J., Morgan, T., Spertus, J. A.,

Stoll, M., Girelli, D., McKeown, P. P., Patterson, C. C., Siscovick, D. S., O'donnell, C. J., Elosua, R., Peltonen, L., Salomaa, V., Schwartz, S. M., Melander, O., Altshuler, D., Merlini, P. A., Berzuini, C., Bernardinelli, L., Peyvandi, F., Tubaro, M., Celli, P., Ferrario, M., Fetiveau, R., Marziliano, N., Casari, G., Galli, M., Ribichini, F., Rossi, M., Bernardi, F., Zonzin, P., Piazza, A., Yee, J., Friedlander, Y., Marrugat, J., Lucas, G., Subirana, I., Sala, J., Ramos, R., Meigs, J. B., Williams, G., Nathan, D. M., MacRae, C. A., Havulinna, A. S., Berglund, G., Hirschhorn, J. N., Asselta, R., Duga, S., Spreafico, M., Daly, M. J., Nemesh, J., Korn, J. M., McCarroll, S. A., Surti, A., Guiducci, C., Gianniny, L., Mirel, D., Parkin, M., Burtt, N., Gabriel, S. B., Thompson, J. R., Braund, P. S., Wright, B. J., Balmforth, A. J., Ball, S. G., Hall, A. S., Linsel-Nitschke, P., Lieb, W., Ziegler, A., Konig, I., Hengstenberg, C., Fischer, M., Stark, K., Grosshennig, A., Preuss, M., Wichmann, H. E., Schreiber, S., Ouwehand, W., Deloukas, P., Scholz, M., Cambien, F., Li, M., Chen, Z., Wilensky, R., Matthai, W., Qasim, A., Hakonarson, H. H., Devaney, J., Burnett, M. S., Pichard, A. D., Kent, K. M., Satler, L., Lindsay, J. M., Waksman, R., Knouff, C. W., Waterworth, D. M., Walker, M. C., Mooser, V., Epstein, S. E., Rader, D. J., Scheffold, T., Berger, K., Stoll, M., Huge, A., Girelli, D., Martinelli, N., Olivieri, O., Corrocher, R., Morgan, T., Spertus, J. A., McKeown, P., Patterson, C. C., Schunkert, H., Erdmann, E., Linsel-Nitschke, P., Lieb, W., Ziegler, A., König, I. R., Hengstenberg, C., Fischer, M., Stark, K., Grosshennig, A., Preuss, M., Wichmann, H. E., Schreiber, S., Hólm, H., Thorleifsson, G., Thorsteinsdottir, U., Stefansson, K.,

Engert, J. C., Do, R., Xie, C., Anand, S., Kathiresan, S., Ardissino, D., Mannucci, P. M., Siscovick, D., O'Donnell, C. J., Samani, N. J., Melander, O., Elosua, R., Peltonen, L., Salomaa, V., Schwartz, S. M., and Altshuler, D. (2009). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* 41, 334–341.

Liu, W., Sun, J., Li, G., Zhu, Y., Zhang, S., Kim, S. T., Wiklund, F., Wiley, K., Isaacs, S. D., Stattin, P., Xu, J., Duggan, D., Carpten, J. D., Isaacs, W. B., Gronberg, H., Zheng, S. L., and Chang, B. L. (2009). Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. *Cancer Res.* 69, 2176–2179.

McCarroll, S. A. (2008). Extending genome-wide association studies to copy-number variation. *Hum. Mol. Genet.* 17, R135–R142.

McGuire, V., Jesser, C. A., and Whittemore, A. S. (2002). Survival among U.S. women with invasive epithelial ovarian cancer. *Gynecol. Oncol.* 84, 399–403.

Need, A. C., Ge, D., Weale, M. E., Maia, J., Feng, S., Heinzen, E. L., Shianna, K. V., Yoon, W., Kasperaviciute, D., Gennarelli, M., Strittmatter, W. J., Bonvicini, C., Rossi, G., Jayathilake, K., Cola, P. A., McEvoy, J. P., Keefe, R. S., Fisher, E. M., St Jean, P. L., Giegling, I., Hartmann, A. M., Moller, H. J., Ruppert, A., Fraser, G., Crombie, C., Middleton,

L. T., St Clair, D., Roses, A. D., Muglia, P., Francks, C., Rujescu, D., Meltzer, H. Y., and Goldstein, D. B. (2009). A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet.* 5, e1000373. doi:10.1371/journal.pgen. 1000373

Permuth-Wey, J., Chen, Y. A., Tsai, Y. Y., Chen, Z., Qu, X., Lancaster, J. M., Stockwell, H., Dagne, G., Iversen, E., Risch, H., Barnholtz-Sloan, J., Cunningham, J. M., Vierkant, R. A., Fridley, B. L., Sutphen, R., McLaughlin, J., Narod, S. A., Goode, E. L., Schildkraut, J. M., Fenstermacher, D., Phelan, C. M., and Sellers, T. A. (2011). Inherited variants in mitochondrial biogenesis genes may influence epithelial ovarian cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 20, 1131–1145.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Song, H., Ramus, S. J., Tyrer, J., Bolton, K. L., Gentry-Maharaj, A., Wozniak, E., Anton-Culver, H., Chang-Claude, J., Cramer, D. W., Dicioccio, R., Dork, T., Goode, E. L., Goodman, M. T., Schildkraut, J. M., Sellers, T., Baglietto, L., Beckmann, M. W., Beesley, J.,

Blaakaer, J., Carney, M. E., Chanock, S., Chen, Z., Cunningham, J. M., Dicks, E., Doherty, J. A., Durst, M., Ekici, A. B., Fenstermacher, D., Fridley, B. L., Giles, G., Gore, M. E., De Vivo, I., Hillemanns, P., Hogdall, C., Hogdall, E., Iversen, E. S., Jacobs, I. J., Jakubowska, A., Li, D., Lissowska, J., Lubinski, J., Lurie, G., McGuire, V., McLaughlin, J., Medrek, K., Moorman, P. G., Moysich, K., Narod, S., Phelan, C., Pye, C., Risch, H., Runnebaum, I. B., Severi, G., Southey, M., Stram, D. O., Thiel, F. C., Terry, K. L., Tsai, Y. Y., Tworoger, S. S., Van Den Berg, D. J., Vierkant, R. A., Wang-Gohrke, S., Webb, P. M., Wilkens, L. R., Wu, A. H., Yang, H., Brewster, W., Ziogas, A., Houlston, R., Tomlinson, I., Whittemore, A. S., Rossing, M. A., Ponder, B. A., Pearce, C. L., Ness, R. B., Menon, U., Kjaer, S. K., Gronwald, J., Garcia-Closas, M., Fasching, P. A., Easton, D. F., Chenevix-Trench, G., Berchuck, A., Pharoah, P. D., and Gayther, S. A. (2009). A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat. Genet.* 41, 996–1000.

Therneau, T. M., and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model.* New York, NY: Springer.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* 67, 91–108.

Venkatraman, E. S., and Olshen, A. B. (2007). A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 23, 657–663.