



OPEN

MassArray analysis of genomic susceptibility variants in ovarian cancer

Sonali Verma^{1,3}✉, Indu Sharma², Varun Sharma², Amrita Bhat³, Ruchi Shah⁴, Gh. Rasool Bhat³, Bhanu Sharma³, Divya Bakshi³, Ashna Nagpal³, Ajay Wakhloo⁵, Audesh Bhat⁶ & Rakesh Kumar^{1,3}✉

Ovarian cancer (OC), a multifaceted and genetically heterogeneous malignancy is one of the most common cancers among women. The aim of the study is to unravel the genetic factors associated with OC and the extent of genetic heterogeneity in the populations of Jammu and Kashmir (J&K). Using the high throughput Agena MassARRAY platform, present case control study was designed which comprises 200 histopathological confirmed OC patients and 400 age and ethnicity matched healthy controls to ascertain the association of previously reported eleven single nucleotide polymorphisms (SNPs) spread over ten genes (*DNMT3A*, *PIK3CA*, *FGFR2*, *GSTP1*, *ERCC5*, *AKT1*, *CASC16*, *CYP19A1*, *BCL2* and *ERCC1*) within the OC population of Jammu and Kashmir, India. The association of each variant was estimated using logistic regression analyses. Out of the 11 SNPs the odds ratio observed for three SNPs; rs2699887 was (1.72 at 95% CI: 1.19–2.48, $p=0.004$), rs1695 was (1.87 at 95% CI: 1.28–2.71, $p=0.001$), and rs2298881 was (0.66 at 95% CI: 0.46–0.96, $p=0.03$) were found significantly associated with the OC after correction with confounding factors i.e. age & BMI. Furthermore, the estimation of interactive analyses was performed and odds ratio observed was 2.44 (1.72–3.47), p value <0.001 suggests that there was a strong existence of interplay between the selected genetic variants in OC, which demonstrate that interactive analysis highlights the role of gene–gene interaction that provides an insight among multiple little effects of various polymorphisms in OC.

Report published by GLOBOCAN in the year 2018 confirm about 18.1 million people were diagnosed with cancer and over 0.7 million are registered in cancer registry every year with estimates deaths of 9.6 million, worldwide^{1,2}. In India, the number of new cancer cases in year 2018 were 1.1 million and estimated deaths were approximately 0.75 million^{1,3,4}. OC ranks 3rd among females in India after Breast and Cervix cancers³. Sixty-one percent of women of J&K (India) are having reproductive health issues like abnormal vaginal discharge, symptoms of a urinary tract infection, pain or bleeding associated with intercourse, uterine endometriosis or even benign and malignant ovarian tumors⁵. It has been observed that there is substantial difference among different Indian populations with special emphasis on their topographical, language, social relationships and genetics⁶.

In the Kashmir valley of North India, OC was third foremost reason of cancer-related deaths in females from 2002 to 2012 (7.45%)⁷. The rise in OC cases in recent past in the population groups of J&K has mandated to identify the reasons including genetic factors that might be responsible for this increased incidence. The population of Jammu and Kashmir is ethnically and genetically diverse compared to the other population groups⁸, thus creating a heterogeneous gene pool. Despite being ethnically diverse, the population groups of J&K have high incidence rate of OC⁹.

There are genetic and environmental factors associated with the risk of OC¹⁰. Genetics is vitally associated with risk of OC in both sporadic and familial cases¹¹. Due to the low mutation rates of *BRCA1* and *BRCA2* genes, they only accounts for fewer cases¹², leaving a lot of OC cases genetically unidentified. The evaluation of genetic determinants and epidemiologic factors of OC might help in advancing better detection and screening methods¹³. Aggregation of specific genetic alterations, particularly Single Nucleotide Polymorphisms (SNPs) contribute

¹Indian Council of Medical Research-Centre for Advance Research, Shri Mata Vaishno Devi University, Katra, Jammu and Kashmir, India. ²Ancient DNA Laboratory, Birbal Shani Institute of Paleo Sciences, Lucknow, Uttar Pradesh, India. ³School of Biotechnology, Shri Mata Vaishno Devi University, Katra, India. ⁴Department of Biotechnology, Kashmir University, Srinagar, Jammu and Kashmir, India. ⁵Department of Obstetrics and Gynecology, Government Medical College, Jammu, Jammu and Kashmir, India. ⁶Centre for Molecular Biology, Central University of Jammu, Jammu, Jammu and Kashmir, India. ✉email: Sonali.verma@smvdu.ac.in; drrakeshthusoo@gmail.com

Characteristics	Cases (200)	Controls (400)	p value
Age (years) Mean \pm SD	59.2 \pm 10.1	56.7 \pm 14.4	0.02
BMI Mean \pm SD	22.6 \pm 4.52	25.4 \pm 4.89	9.74E-12
Menopausal status			
Pre-menopausal	124	276	0.33
Post-menopausal	74	124	
Stage			
I/II	78	–	–
III/IV	110	–	–
Age at menarche (years)			
> 12	107	215	0.02
< 12	93	185	
Histology of tumors			
Epithelial	123	–	–
Endometroid	15	–	
Germ cell	9	–	
Sex cord stromal cell	33	–	
Metastasis	20	–	
Oral contraceptive use			
Yes	80	–	–
No	120	–	
Breast nodules			
Yes	22	–	–
No	162	–	

Table 1. Clinical details of cases and controls.

to OC predisposition¹⁴. SNPs investigated in the present study include genes like *DNMT3A*, *PIK3CA*, *FGFR2*, *GSTP1*, *ERCC5*, *AKT1*, *CASC16*, *CYP19A1*, *BCL2*, and *ERCC1*. SNPs of these genes have been found in association with cancers of ovary, breast, stomach and lung among different populations, globally^{15–30}. Recently, reported association of *XRCC1* (rs25487), *HoGG1* (rs1052133), *DNAH11* (rs2285947) and *LRFN2* (rs2494938) gene variants with OC provided an insight that genetic variants provide increased risk in the present populations^{9,31}. Besides these two reports, no genetic data is available on OC from the J&K region. In order to extend screening, more genetic variants in the J&K population, in-house cancer SNP panel was designed to screen the OC patients that comprises of eleven SNPs of ten genes (*DNMT3A*¹⁷, *PIK3CA*^{21,22}, *FGFR2*^{21,22}, *GSTP1*^{21,22}, *ERCC5*^{18–20}, *AKT1*¹⁶, *CASC16*¹⁵, *CYP19A1*, *BCL2*^{15,23,24}, *ERCC1*^{25–30}); and population based association study was conducted to assess the genetic predisposition of cancer susceptibility variants with OC.

Results

The clinical characteristics of both cases and controls are given in Table 1. The mean age of cases was 59.2 (\pm 10.1) years, which and that of controls was (56.7 \pm 14.4) years respectively. The average BMI of the cases (22.6 \pm 4.52) was significantly lower than that of the controls (25.4 \pm 4.89) ($p = 9.74E - 12$).

Out of 11 SNPs genotyped, only six SNPs (rs2699887, rs1695, rs2298881, rs10046, rs2981582, and rs751402) were having genotyping quality of more than 95% following stringent quality check and hence were included in the analyses. The allelic frequency distribution of these six SNPs in cases and controls are given in Table 2. Out of these six SNPs, only rs2699887, rs1695, and rs2298881 were found significantly associated with OC in the studied population. The observed allelic Odds Ratio (OR) of rs2699887 was (1.5 at 95% CI: 1.1–2.0, $p = 0.003$), rs1695 was (1.4 at 95% CI: 1.0–1.8, $p = 0.01$) and rs2298881 was (0.6 at 95% CI: 0.4–0.9, $p = 0.03$), respectively. The observed allelic OR for the non-significantly associated SNPs rs10046 of *CYP19A1*, rs2981582 of *FGFR2*, and rs751402 of *ERCC5* was (1.2 at 95% CI: 0.9–1.5, $p = 0.12$), (1.1 at 95% CI: 0.8–1.4, $p = 0.33$), and (0.8 at 95% CI (0.6–1.1), $p = 0.21$), respectively.

After applying logistic regression in order to avoid any biasness caused by confounding factors like age and BMI, all three significantly associated variants followed the same direction of association. Adjusted OR of rs2699887 was (1.72 at 95% CI: 1.19–2.48, $p = 0.004$), rs1695 was (1.87 at 95% CI: 1.28–2.71, $p = 0.001$), and rs2298881 was (0.66 at 95% CI: 0.46–0.96, $p = 0.03$). The SNPs that were not included in the final analyses because of their low call rate (< 95%) are summarized in Supplementary data: table S3.

Interaction analysis between the three significantly associated SNPs revealed that two SNPs were risk alleles with 10/10 cross-validation consistency (CVC) and one SNP had a Testing Balance Accuracy (TBA) score of above 0.50 in all the SNPs Table 3.

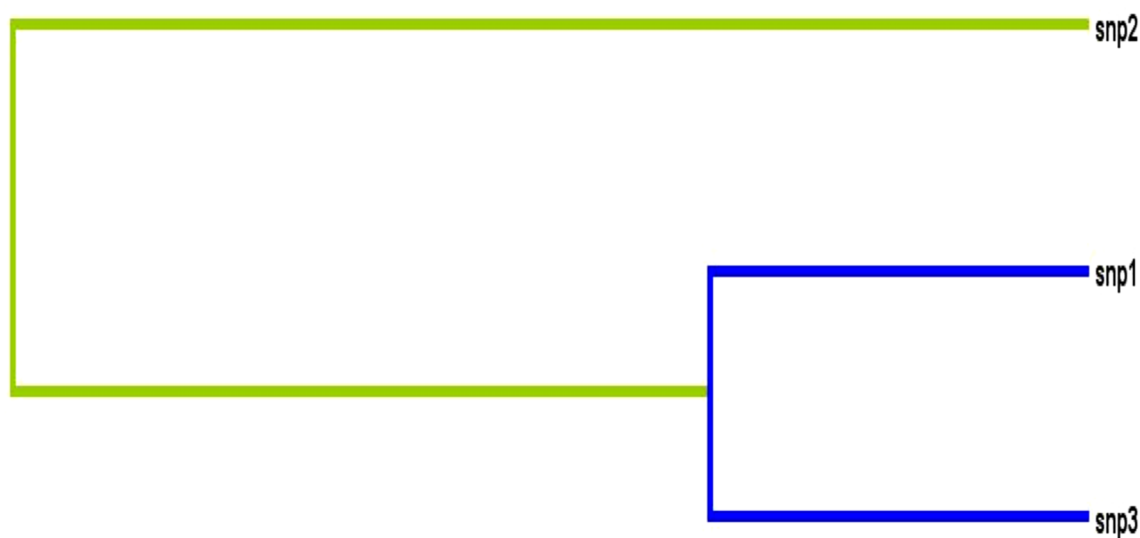
Entropy dendrogram revealed that SNP rs1695 has a synergistic effect on SNPs rs2699887 and SNP rs2298881. However, strong redundancy was observed between SNP rs2699887 and SNP rs2298881 Fig. 1. Interactive analyses revealed that there are interactions between genetic variants of *PIK3CA*, *GSTP1* and *ERCC* genes [OR-2.4483

S.No	Gene	SNPs	Frequency in cases	Frequency in control	HWE	OR 95% CI	p value	OR* 95% CI (dominant model)	p value * (dominant model)	PAR
1	CYP19A1	rs10046	A = 0.3447	A = 0.3	0.6271	1.2 (0.9–1.5)	0.12	1.36 (0.946–1.97)	0.95	–
			G = 0.6553	G = 0.7						
2	PIK3CA	rs2699887	T = 0.259	T = 0.185	0.0682	1.5 (1.1–2.0)	0.003	1.72 (1.19–2.48)	0.004	25.93
			C = 0.741	C = 0.815						
3	FGFR2	rs2981582	A = 0.3613	A = 0.3325	0.4957	1.1 (0.8–1.4)	0.33	1.30 (0.89–1.88)	0.165	–
			G = 0.6387	G = 0.6675						
4	GSTP1	rs1695	G = 0.2539	G = 0.1941	0.872	1.4 (1.0–1.8)	0.01	1.87 (1.28–2.71)	0.001	39.2
			A = 0.7461	A = 0.8059						
5	ERCC1	rs2298881	A = 0.2356	A = 0.3029	0.9036	0.7 (0.53–0.94)	0.01	0.66 (0.46–0.96)	0.03	–
			C = 0.7644	C = 0.6971						
6	ERCC5	rs751402	A = 0.2216	A = 0.2552	0.5037	0.8 (0.6–1.1)	0.21	0.71 (0.48–1.03)	0.07	–
			G = 0.7784	G = 0.7448						

Table 2. Distribution of risk allele frequency and association analyses of variants with genotyping call greater than 95%. *Adjusted with age and BMI.

SNP combination	Cross-validation statistics		p value
SNP1_rs2699887, SNP2_rs1695, SNP3_rs2298881	Balanced accuracy	0.605	< 0.0001
	Accuracy	0.64	
	Specificity	0.71	
	Odds ratio	2.4483 (1.7229–3.4791)	

Table 3. Interaction analysis OC cases and controls. *p < 0.05 was considered significant.



*SNP1-rs2699887, SNP2-rs1695, SNP3-rs2298881.

Color coding	Interpretation
Green —	Moderately Redundant
Blue —	Highly Redundant

Figure 1. SNP-SNP interaction analysis using MDR, color-coding of bars used to interpret interactions.

(1.7229–3.4791), $p > 0.0001$] via permutation test of 1000 iterations Table 3. Sample size included in the present study had 80–90% power assuming minor allele frequency 0.20 to detect the association with O.R (1.2–1.7).

Functional annotation of associated SNPs. As per the UCSC Genome browser database (<http://genome.ucsc.edu/>), SNPs rs2699887, rs1695 and rs2298881 fall in the regulatory region which was generated from Epigenome, Encyclopedia of DNA Elements (ENCODE) project where all the associated SNPs (rs2699887, rs1695 and rs2298881) are located nearby the enhancer region (H3K4Me1 mark) and are also observed near the active promoter region (H3K27Ac mark) as well as in seven cell lines (H1-hESC, GM12878, K562, HUVEC, HSMC, NHLF, and NHEK) which specified that SNPs (rs2699887, rs1695 and rs2298881) were possibly intricate with the gene expression, whereas SNPs (rs2699887, rs1695 and rs2298881) also knockdown into 125 types of cells in DNase I hypersensitive region (Supplementary data: figure S2, S3 & S4). The transcription factor data from ENCODE and haploreg of associated SNPs rs2699887, rs1695 and rs2298881 also indicates that these variants alter certain regulatory motifs (Supplementary data: figure S1). The role of rs2699887, rs1695 and rs2298881 SNPs; were also examined by using current biological servers (GTEx³², Haploreg³³). We assessed the gene expression of tissues using data from GTEx portal and found that the eQTL expression of ovarian tissue of variant rs1695 of *GSTP1* significant association with change in gene *NDUFV1*, showing change in expression of ovarian tissue with p value = 0.000018 and NES value is 0.45 and variant rs2298881 of *ERCC1* confirm expression change of ovarian tissue with p value = 7.9E – 11 and NES value = – 0.48 (reduced effect).

Therefore, we could speculate that functional SNPs located in *PIK3CA* and *GSTP1* gene may disrupt transcription factor response elements, and further affect the expression level of *PIK3CA* and *GSTP1* and ultimately affect the occurrence and development of ovarian cancer. On the basis of UCSC genome browser, haploreg and GTEx eQTL data for ovarian cancer. The study also found that rs2699887 and rs1695 were associated with the prognosis of ovarian cancer after adjustment for age and BMI.

These findings proved that functional annotation of risk associated variants rs2699887 and rs1695 were differentially regulated in ovarian tissues which suggested a possible mechanism for the effect on ovarian cancer risk. However, risk estimates suggested that *ERCC1* rs2298881 is a protection-associated genetic variation in ovarian cancer, whereas no association was found for rs10046, rs2981582 and rs751402.

Discussion

PI3K/AKT pathway under which the *PIK3CA* gene belongs; plays an important role in cell cycle progression, programmed cell death and drug resistance. Consequently, various SNPs in the *PI3K/AKT* pathway genes were found associated with different malignancies. The rs2699887 (C > T) variant in the upstream region of intron-1 has been associated with various cancers including OC^{34–36}. This variant has also been associated with lung cancer toxicity in patients who underwent platinum-based chemotherapy³⁷. Allele ‘T’ of variant rs2699887 was found associated with increased susceptibility towards OC. Interestingly, this variant has been reported to cause transcription factor binding effect which leads to change in normal splicing patterns³⁴. The role of this variant in OC and other cancer susceptibility may be due to diverse role of *PIK3CA* in the initiation and progression of OC. The mutation in *PIK3CA* gene have been frequently identified in endometrial ovarian carcinomas and not found in serous epithelial ovarian carcinomas³⁸. Over-expression of *AKT* gene from *PI3K/AKT* pathway may leads to progression and metastasis of OC^{34,39}. The association of rs1695 (A > G) variant with OC is the first studied population in India. It was previously associated with OC in Brazilian, Australian and French populations but not in Caucasian and Chinese populations^{22,40,41}. The rs1695 (A > G) variant, missense mutation in exon 5 of *GSTP1* that changes amino acid 105 from isoleucine to valine (I105V). The I105V change has prognostic effect in OC patients with paclitaxel plus carboplatin combination chemotherapy (TC therapy)²¹. Additionally, in some populations the mutant and wild alleles defend cells against programmed cell death through JNK pathway and wild type allele provided risk in epithelial OC risk⁴². In addition to the significant association of ‘AA’ genotype of rs1695 with OC, we also identified that rs2699887 and rs1695 variants in combination increases risk of OC in post-menopausal women than pre-menopausal women. Thus, indicating that *GSTP1* and *PIK3CA* mutants when in combination are highly responsible for development of OC in studied population group.

Published data regarding rs2298881 in the *ERCC1* gene and rs751402 in the *ERCC5* gene with OC although scanty, retrospective studies have associated rs2298881 and rs751402 with increased risk for breast, lung, colorectal, gastric and several other types of cancers^{25,27,43,44}. Contrary to this, our data supports a protective nature of rs2298881 variant in the studied population. Variant rs751402 of *ERCC5* on the other hand was not found associated with OC. These results were consistent with our previous report⁴⁵, whereas another polymorphisms of DNA repair pathway rs25487 (*XRCC1*) was significantly not associated with OC.

We excavated the following database (Epigenome, ENCODE-UCSC genome browser, haploreg, GTEx mobile portal) and observed that rs2699887, and rs1695 and rs2298881 could affect the gene expression and are likely to modify regulatory motifs and disrupt protein binding activities (Supplementary data: figure S1, S2, S3 and S4). All three SNPs (rs2699887, rs1695 and rs2298881) in ENCODE data fall in the hypersensitivity region of DNase, in enhancer (H3K4Me1) and in promoter (H3K27Ac) region in seven cell lines from Epigenome and ENCODE, suggesting that there was a probable mechanism which effected the regulation of gene expression resulting in OC risk.

In the light of significant association of three variants either directly or indirectly with OC in the studied population; we further evaluated these variants in GTEx mobile portal. The rs1695 variant showed eQTL expression with *NDUFV1* gene in ovarian tissue indicating the change in expression in *NDUFV1* gene, where *NDUFV1* gene plays a role in apoptosis due to reactive oxygen species (ROS) which may have strong implication on OC⁴⁶. The rs2298881 variant showed eQTL expression in ovarian tissues indicating the change in expression.

In summary, our study revealed that SNP rs2699887 of *PIK3CA*, rs1695 of *GSTP1* and rs2298881 of *ERCC1* gene might affect the expression directly or indirectly in OC and ultimately modify the OC risk in J&K. The findings support for additional functional studies to identify the biological mechanism behind the progression of OC. To the best of our knowledge this is the prelude study that has replicated the association between various polymorphisms and the risk of OC in J&K women; however, the functional validation of these variants is mandatory in order to better understand the association between these variables and its possible role in OC.

Conclusion

The observation from the study highlights the role of independent as well as interactive effect of SNPs in OC. The interactive analysis also provides insight that variants of *PIK3CA*, *GSTP1* and *ERCC1* have relatively high risk of OC. Hence, more replication studies with larger sample size are required along with their functional validations to unravel the biological significance of SNPs in OC.

Materials and methodology

Ethics statement. This case-control genetic association study was approved by the Institutional Ethical Review Board (IERB) of Shri Mata Vaishno Devi University under notification number (SMVDU/IERB/16/48). All experimental protocols were conducted according to the guidelines and regulations set by the Institutional Ethical Review Board (IERB).

Sampling of subjects. A total of 600 participants (200 histopathological confirmed OC patients and 400 healthy females) were included in the study with written informed consent. The participant details were duly filled pro forma and was signed by the participants. 2–3 ml of venous blood was collected in EDTA vials. Samples were collected from various hospitals and clinics of J&K region. The clinical parameters are listed in Table 1.

SNP selection. In the study, the variants were selected on the basis of following criteria:

- (1) The potential genetic variation data implicated in carcinogenesis and associated traits were retrieved from the NCBI's Single Nucleotide Polymorphism database (dbSNP-NCBI), including the variants previously associated with OC and other cancers.
- (2) Only variations having annotation in exonic promoter region, 5' un-translated regions (5' UTR) or 3' UTR, exonic and intronic SNPs (if the condition or criteria for SNP selection were met) were screened. The details of selected SNPs are given (Supplementary data: Table S1).

Genotyping. The genomic DNA was isolated by using the Qiagen DNA isolation kit (Catalogue No. 51206). Genotyping was performed using Agena MassArray platform, a robust and highly sensitive tool for genotyping of SNPs⁴⁷ available in the Central Analyzer Mass Array facility at SMVDU. The SNP panel was customized by using Agena Bioscience Assay Design Suite (version 2.0). The sequence of primers have been provided in (Supplementary data: Table S2). The whole methodology for pool PCR, Shrimp Alkaline Phosphatase (SAP) and iPLEX PCR was adopted from studies^{47,48}. The genotyping results were validated by replicating 10% of random samples and the concordance rate was 98.3%. In the reaction of 384 well plates, one negative and one positive control were added in every reaction to check the quality of the reaction.

Genotyping quality control. For genotyping quality assurance, the following criteria were applied: (i) SNPs having call rate >95% were included in the statistical analysis⁴⁹ and (ii) Hardy-Weinberg Equilibrium (HWE) among cases and controls was used for assessing the quality of genotypes.

Statistical analysis. Statistical analyses was mainly performed using Plink V.1.09⁵⁰ with maximum 10,000 permutations. Significance of the association was evaluated by 3×2 chi square test. Logistic regression analysis was performed using SPSS V.23 in order to obtain corrected odds ratio (OR), confidence interval (CI) and p-value from confounding factors like age and BMI. The SNP-SNP interaction to analyze the synergic effect of significantly associated SNPs was performed by using Multifactor Dimensionality Reduction (MDR) software⁵¹. Population attribution risk (PAR) percentage was also estimated for significantly associated risk variants by using adjusted OR. Power of the study was estimated by using CATS online calculator⁵².

Putative analysis of SNPs. The study also interpreted the candidate SNPs in regulatory region assembled in Epigenome (<https://www.roadmappigenomics.org>)⁵³, Encyclopedia of DNA Elements (ENCODE) tool from UCSC Genome Browser (<https://www.UCSCgenomebrowser>)⁵⁴. The present study also investigated the data of several cell lines (H3K4Me3, H3K27Ac, and H3K4Me1) with special emphasis on H1-hESC, GM12878, K562, HUVEC, HSMM, NHLF, and NHEK cells. Further, the study examined DNaseI Hypersensitivity region and transcription factor binding sites including their changed motifs from Haploreg (<https://www.haploreg.org>)³³ data 55 in 125 cell types. Genotype-Tissue Expression (GTEx) portal was used for the identification of gene expression of our candidate SNPs of various genes in ovarian tissue (<https://www.gtexportal.org>)⁵⁵.

Received: 6 May 2020; Accepted: 28 October 2020

Published online: 03 December 2020

References

- Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J. Clin.* **68**, 394–424. <https://doi.org/10.3322/caac.21492> (2018).
- Statistics. *National Institute of Cancer Prevention and Research.* (2018).
- WHO, I. A. f. R. o. C. *GLOBOCON, 2018 Globocon.*
- Cancer Statistics in India. <http://cancerindia.org.in/cancer-statistics/>.
- Bhasin, M. K. & Nag, S. A demographic profile of the people of Jammu and Kashmir 3. Estimates, trends and differentials in mortality. *J. Hum. Ecol.* **13**, 113–140. <https://doi.org/10.1080/09709274.2002.11905526> (2002).
- Mastana, S. S. & Papiha, S. S. in *Genetic Disorders of the Indian Subcontinent* (ed Dhavendra Kumar) 57–80 (Springer Netherlands, 2004).
- Wani, F.J., Khan, N.A., Pandita, K.K., Khurshid, R., Khan, S.H. Cancer trends in Kashmir; common types, site incidence and demographic profiles: National Cancer Registry 2000–2012. *Indian J. Cancer* (2014).
- Sharma, I. *et al.* Ancient Human Migrations to and through Jammu Kashmir- India were not of Males Exclusively. *Scientific reports* **8**, 851. <https://doi.org/10.1038/s41598-017-18893-8> (2018).
- Verma, S. *et al.* DNA base excision repair genes variants rs25487 (X-ray repair cross-complementing 1) and rs1052133 (human 8-oxoguanine glycosylase 1) with susceptibility to ovarian cancer in the population of the Jammu region, India. *J. Cancer Res. Ther.* **15**, 1270–1275. https://doi.org/10.4103/jcrt.JCRT_65_18 (2019).
- Fasching, P. A. *et al.* Role of genetic polymorphisms and ovarian cancer susceptibility. *Mol. Oncol.* **3**, 171–181. <https://doi.org/10.1016/j.molonc.2009.01.008> (2009).
- Razi, S. *et al.* The incidence and mortality of ovarian cancer and their relationship with the Human Development Index in Asia. *ecancermedicalscience* **10**, 628. <https://doi.org/10.3332/ecancer.2016.628> (2016).
- Benson, J. R. & Jatoi, I. The global breast cancer burden. *Future Oncol.* **8**, 697–702. <https://doi.org/10.2217/fon.12.61> (2012).
- McLemore, M. R., Miaskowski, C., Aouizerat, B. E., Chen, L.-M. & Dodd, M. J. Epidemiological and genetic factors associated with ovarian cancer. *Cancer Nurs.* **32**, 281–290. <https://doi.org/10.1097/NCC.0b013e31819d30d6> (2009).
- Dong, A., Lu, Y. & Lu, B. Genomic/epigenomic alterations in ovarian carcinoma: Translational insight into clinical practice. *J. Cancer* **7**, 1441–1451. <https://doi.org/10.7150/jca.15556> (2016).
- Yang, X. *et al.* Association of the functional BCL-2 rs2279115 genetic variant and small cell lung cancer. *Tumour Biol.* **37**, 1693–1698. <https://doi.org/10.1007/s13277-015-3934-9> (2016).
- Li, X., Zhang, R., Liu, Z., Li, S. & Xu, H. The genetic variants in the PTEN/PI3K/AKT pathway predict susceptibility and CE(A)F chemotherapy response to breast cancer and clinical outcomes. *Oncotarget* **8**, 20252–20265. <https://doi.org/10.18632/oncotarget.15690> (2017).
- Mostowska, A., Sajdak, S., Pawlik, P., Lianeri, M. & Jagodzinski, P. P. DNMT1, DNMT3A and DNMT3B gene variants in relation to ovarian cancer risk in the Polish population. *Mol. Biol. Rep.* **40**, 4893–4899. <https://doi.org/10.1007/s11033-013-2589-0> (2013).
- Pongsavee, M. & Wisuwan, K. ERCC5 rs751402 polymorphism is the risk factor for sporadic breast cancer in Thailand. *Int. J. Mol. Epidemiol. Genet.* **9**, 27–33 (2018).
- Sang, L., Lv, Z., Sun, L. P., Xu, Q. & Yuan, Y. Impact of SNP-SNP interactions of DNA repair gene ERCC5 and metabolic gene GSTP1 on gastric cancer/atrophic gastritis risk in a Chinese population. *World J. Gastroenterol.* **24**, 602–612. <https://doi.org/10.3748/wjg.v24.i5.602> (2018).
- Lu, J. J. *et al.* Lack of association between ERCC5 gene polymorphisms and gastric cancer risk in a Chinese population. *Genet. Mol. Res. GMR* **15**, <https://doi.org/10.4238/gmr.15027779> (2016).
- Yoshihama, T. *et al.* GSTP1 rs1695 is associated with both hematological toxicity and prognosis of ovarian cancer treated with paclitaxel plus carboplatin combination chemotherapy: A comprehensive analysis using targeted resequencing of 100 pharmacogenes. *Oncotarget* **9**, 29789–29800. <https://doi.org/10.18632/oncotarget.25712> (2018).
- Han, L. Y., Liu, K., Lin, X. L., Zou, B. B. & Zhao, J. S. Lack of any association of GST genetic polymorphisms with susceptibility to ovarian cancer—A meta-analysis. *Asian Pac. J. Cancer Prevent.* **15**, 6131–6136 (2014).
- Xu, P. *et al.* Genetic variation in BCL2 3'-UTR was associated with lung cancer risk and prognosis in male Chinese population. *PLoS ONE* **8**, e72197. <https://doi.org/10.1371/journal.pone.0072197> (2013).
- Shi, H. *et al.* Prognostic impact of polymorphisms in the MYBL2 interacting genes in breast cancer. *Breast Cancer Res. Treat.* **131**, 1039–1047. <https://doi.org/10.1007/s10549-011-1826-2> (2012).
- He, J. *et al.* Genetic variants in the nucleotide excision repair pathway genes and gastric cancer susceptibility in a southern Chinese population. *Cancer Manag. Res.* **10**, 765–774. <https://doi.org/10.2147/cmar.s160080> (2018).
- Zhu, Y., Guo, L., Wang, S., Yu, Q. & Lu, J. Association of smoking and XPG, CYP1A1, OGG1, ERCC5, ERCC1, MMP2, and MMP9 gene polymorphisms with the early detection and occurrence of laryngeal squamous carcinoma. *J. Cancer* **9**, 968–977. <https://doi.org/10.7150/jca.22841> (2018).
- Chen, J. *et al.* Association of ERCC1 polymorphisms with the risk of colorectal cancer: A meta-analysis. *Crit. Rev. Eukaryot. Gene Expr.* **27**, 267–275. <https://doi.org/10.1615/CritRevEukaryotGeneExpr.2017019713> (2017).
- Sun, Y., Tan, L., Li, H., Qin, X. & Liu, J. Association of NER pathway gene polymorphisms with susceptibility to laryngeal cancer in a Chinese population. *Int. J. Clin. Exp. Pathol.* **8**, 11615–11621 (2015).
- Zhao, Z. *et al.* The association of polymorphisms in nucleotide excision repair genes with ovarian cancer susceptibility. *Biosci. Rep.* **38**, doi:<https://doi.org/10.1042/bsr20180114> (2018).
- Assis, J. *et al.* Genetic variants as ovarian cancer first-line treatment hallmarks: A systematic review and meta-analysis. *Cancer Treat. Rev.* **61**, 35–52. <https://doi.org/10.1016/j.ctrv.2017.10.001> (2017).
- Verma, S. *et al.* Genetic variants of DNAH11 and LRFN2 genes and their association with ovarian and breast cancer. *Int. J. Gynaecol. Obstet.* <https://doi.org/10.1002/ijgo.12997> (2019).
- The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585. <https://doi.org/10.1038/ng.2653> (2013).
- Ward, L. D. & Kellis, M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–934. <https://doi.org/10.1093/nar/gkr917> (2012).
- Wang, L.-E. *et al.* Roles of genetic variants in the PI3K and RAS/RAF pathways in susceptibility to endometrial cancer and clinical outcomes. *J. Cancer Res. Clin. Oncol.* **138**, 377–385. <https://doi.org/10.1007/s00432-011-1103-0> (2012).
- Lin, L., Zhang, Z., Zhang, W., Wang, L. & Wang, J. Roles of genetic variants in the PI3K/PTEN pathways in susceptibility to colorectal carcinoma and clinical outcomes treated with FOLFOX regimen. *Int. J. Clin. Exp. Pathol.* **8**, 13314–13322 (2015).
- Wang, C. *et al.* The study of the relation of DNA repair pathway genes SNPs and the sensitivity to radiotherapy and chemotherapy of NSCLC. *Sci. Rep.* **6**, 26526–26526. <https://doi.org/10.1038/srep26526> (2016).
- Pu, X. *et al.* PI3K/PTEN/AKT/mTOR pathway genetic variation predicts toxicity and distant progression in lung cancer patients receiving platinum-based chemotherapy. *Lung cancer (Amsterdam, Netherlands)* **71**, 82–88. <https://doi.org/10.1016/j.lungcan.2010.04.008> (2011).

38. Campbell, I. G. *et al.* Mutation of the PIK3CA gene in ovarian and breast cancer. *Can. Res.* **64**, 7678–7681. <https://doi.org/10.1158/0008-5472.can-04-2933> (2004).
39. Li, L., Plummer, S. J., Thompson, C. L., Tucker, T. C. & Casey, G. Association between phosphatidylinositol 3-kinase regulatory subunit p85alpha Met326Ile genetic polymorphism and colon cancer risk. *Clin. Cancer Res.* **14**, 633–637. <https://doi.org/10.1158/1078-0432.ccr-07-1211> (2008).
40. Spurdle, A. B. *et al.* Polymorphisms at the glutathione S-transferase GSTM1, GSTT1 and GSTP1 loci: risk of ovarian cancer by histological subtype. *Carcinogenesis* **22**, 67–72 (2001).
41. Delort, L. *et al.* Association between genetic polymorphisms and ovarian cancer risk. *Anticancer Res.* **28**, 3079–3081 (2008).
42. Oliveira, C. *et al.* Polymorphisms of glutathione S-transferase Mu 1 (GSTM1), Theta 1 (GSTT1), and Pi 1 (GSTP1) genes and epithelial ovarian cancer risk. *Dis. Markers* **33**, 155–159. <https://doi.org/10.3233/dma-2012-0920> (2012).
43. Yang, Z., Fang, X., Pei, X. & Li, H. Polymorphisms in the ERCC1 and XPF genes and risk of breast cancer in a Chinese population. *Genet. Test. Mol. Biomark.* **17**, 700–706. <https://doi.org/10.1089/gtmb.2013.0122> (2013).
44. Zhu, J. *et al.* Association studies of ERCC1 polymorphisms with lung cancer susceptibility: A systematic review and meta-analysis. *PLoS ONE* **9**, e97616–e97616. <https://doi.org/10.1371/journal.pone.0097616> (2014).
45. Verma, S. *et al.* DNA base excision repair genes variants rs25487 (X-ray repair cross-complementing 1) and rs1052133 (human 8-oxoguanine glycosylase 1) with susceptibility to ovarian cancer in the population of the Jammu region, India. *J. Cancer Res. Ther.* **15**, https://doi.org/10.4103/jcrt.JCRT_65_18 (2019).
46. Jacquemin, G. *et al.* Granzyme B-induced mitochondrial ROS are required for apoptosis. *Cell Death Differ.* **22**, 862–874. <https://doi.org/10.1038/cdd.2014.180> (2015).
47. Gabriel, S., Ziaugra, L. & Tabbaa, D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. in *Current Protocols in Human Genetics* Chapter 2, Unit 2.12, <https://doi.org/10.1002/0471142905.hg0212s60> (2009).
48. Lalrohlu, F. *et al.* Genotyping of T2D susceptible genes in a high risk North-East Indian population. *Obes. Med.* **17**, 100162. <https://doi.org/10.1016/j.obmed.2019.100162> (2020).
49. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat. Protoc.* **5**, 1564–1573. <https://doi.org/10.1038/nprot.2010.116> (2010).
50. Shaun Purcell, C. C. (2015).
51. Motsinger, A. A. & Ritchie, M. D. Multifactor dimensionality reduction: an analysis strategy for modelling and detecting gene-gene interactions in human genetics and pharmacogenomics studies. *Hum. Genomics* **2**, 318–328. <https://doi.org/10.1186/1479-7364-2-5-318> (2006).
52. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213. <https://doi.org/10.1038/ng1706> (2006).
53. Satterlee, J. S. *et al.* The NIH common fund/roadmap epigenomics program: Successes of a comprehensive consortium. *Sci. Adv.* **5**, eaaw6507, <https://doi.org/10.1126/sciadv.aaw6507> (2019).
54. Rosenbloom, K.R.S.C., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C., Maddren, M., Fang, R., Heitner, S.G., Lee, B.T., Barber, G.P., Harte, R.A., Diekhans, M., Long, J.C., Wilder, S.P., Zweig, A.S., Karolchik, D., Kuhn, R.M., Haussler, D., Kent, W.J. ENCODE annotation data. *Nucleic Acids Res. (Database issue):D56–D63. ENCODE Data in the UCSC Genome Browser: Year 5 Update.* .
55. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. bioRxiv:787903, <https://doi.org/10.1101/787903> (2019).

Acknowledgements

The authors gratefully acknowledge the financial support for purchasing consumables and equipment's provided by the Indian Council of Medical Research (5/10/15/CAR-SMVDU/2018-RBMCH), Department of Science and Technology-SERB (YSS/2014/000659) Government of India, and DST, Government of India grant (SSTP/J&K/459). Authors also acknowledge Dr. Swarkar Sharma for technical advice provided during the planning and implementation of study.

Author contributions

S.V. and R.K. primarily planned the study. S.V., A.B., G.R.B., B.S., D.B. and A.N. collected the samples. S.V., A.B., R.S., I.S. and V.S. performed experiment in lab. S.V., V.S. and I.S. performed statistical analysis and drafted manuscript. A.W. provided samples for the study. R.K., A.B. and R.S. gave critical comments in manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-76491-7>.

Correspondence and requests for materials should be addressed to S.V. or R.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020