Current Epidemiology Reports (2020) 7:49-57
https://doi.org/10.1007/s40471-020-00227-4

GENETIC EPIDEMIOLOGY (C AMOS, SECTION EDITOR)

Functional Landscape of Common Variants Associated with Susceptibility to Epithelial Ovarian Cancer

Paulo Cilas Morais Lyra Jr[1] · Letícia B. Rangel[1] · Alvaro N. A. Monteiro[2]

Published online: 14 January 2020

Abstract

Purpose of the Review To date, genome-wide association studies (GWASs) have identified 39 genomic loci associated with risk of epithelial ovarian cancer at genome-wide significance level ($p \leq 5 \times 10^{-8}$) and 13 additional loci using less strict thresholds. Follow-up functional dissection of these loci to uncover the underlining mechanisms driving cancer susceptibility has been challenging.

Recent Findings In a manner similar to how post-linkage studies led the characterization of then poorly understood cellular pathways, functional analysis of GWAS loci is revealing new mechanisms of ovarian cancer.

Summary Here, we review recent methodological and conceptual progress relevant to the understanding of how common genetic variation influences the risk of epithelial ovarian cancer.

Keywords GWAS · TWAS · Functional analysis · Ovarian cancer · Common variants · SNPs · Transcription

Abbreviations

CCOC	Clear cell ovarian carcinoma
DDR	DNA damage response
ENOC	Endometrioid ovarian carcinoma
EOC	Epithelial ovarian cancer
eQTL	Expression quantitative trait loci
GWAS	Genome-wide association studies
HGSOC	High-grade serous ovarian carcinoma
LD	Linkage disequilibrium
LGSOC	Low-grade serous ovarian carcinoma
MOC	Mucinous ovarian carcinoma
MAF	Minor allele frequency
OCAC	Ovarian Cancer Association Consortium
PARP	Poly ADP ribosyl polymerase
S/MAR	Substrate/matrix attachment region
SNP	Single nucleotide polymorphism

This article is part of the Topical Collection on *Genetic Epidemiology*

✉ Alvaro N. A. Monteiro
alvaro.monteiro@moffitt.org

[1] Biotechnology/RENORBIO Program, Federal University of Espírito Santo, Vitória, ES, Brazil

[2] Cancer Epidemiology Program, H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Drive, Tampa, FL, USA

Introduction

In 2018 alone, over 295,000 new cases and 180,000 deaths were due to ovarian cancer, the eighth leading cause of cancer mortality among women [1]. Epithelial ovarian cancer (EOC), the most prevalent type of ovarian cancer, is detected at later stages in more than 50% of the cases with poor prognosis and low survival rates [2]. Invasive EOC is classified into five major histological types, high-grade serous (HGSOC), low-grade serous (LGSOC), mucinous (MOC), endometrioid (ENOC), and clear cell (CCOC) ovarian carcinoma, and two borderline disease histological types, serous and mucinous [3]. These EOC histological subtypes reflect different cell types, many of them from non-ovarian tissues, from which each tumor originates [4]. For example, most HGSOC, the most common EOC type, likely originates from secretory epithelial cells (or their progenitors) in distal fallopian tube precursor lesions [5, 6].

Analysis of mutations in tumor tissue has largely confirmed the histological subtypes, which can be grouped into types I and II [7]. Less aggressive type I tumors have slow growth and include low-grade serous, low-grade endometrioid, mucinous, and clear cell carcinomas and display mutations in *KRAS*, *BRAF*, and *PIK3CA* but not in *TP53* [8]. Highly aggressive type II tumors are characterized by *TP53* mutations and include high-grade serous, high-grade endometrioid tumors, and

undifferentiated carcinomas, but *TP53* mutation prevalence may vary with subtype [9–13]. Type II tumors, exemplified by HGSOC, display a high prevalence of defects in double-strand DNA break repair pathways [13–17].

Reproductive history, age, environmental, and lifestyle factors influence ovarian cancer risk, but genetic factors are also important contributors [18–21]. Inherited rare pathogenic variants in several genes conferring either high (RR > 4; *BRCA1* and *BRCA2*) or moderate (2 < RR ≤ 4; *ATM*, *BRIP1*, *MLH1*, *MSH2, MSH6, PALB2*, *PMS2*, *RAD51C*, and *RAD51D*) lifetime EOC risk have been identified but explain less than a third of the excess familial risk, indicating the existence of additional unidentified genetic factors [22–28].

Because a significant part of inherited susceptibility to cancer is likely to be explained by common alleles with low penetrance [2–5], investigators have expanded their search to identify common alleles (typically minor allele frequency [MAF] > 1%) associated with cancer risk in a significant fraction of the population [29–31]. In ovarian cancer, this effort has been spearheaded by the Ovarian Cancer Association Consortium (OCAC), and has led to the identification of many EOC susceptibility loci. Collectively, loci associated with invasive EOC in women of European ancestry are estimated to explain 6.4% of the polygenic risk [32••].

In this mini-review, we discuss the progress on the functional dissection of common loci identified to date to explore the biology of EOC. In the same manner that dissecting the biological role of *BRCA1* and *BRCA2* led to understanding of their central role in the DNA damage response (DDR) and the discovery of synthetic lethal approaches using inhibitors of poly(ADP-ribose) polymerase 1 (PARP1) [33–38], we expect that identification of mechanisms driving cancer susceptibility will open new prevention and therapeutic options. Recently, Kar et al. published an excellent review of the state of common variation in ovarian cancer, and therefore, we will only briefly summarize those studies covered by Kar et al. and discuss new developments in more detail [39].

Genome-Wide Association Studies in Epithelial Ovarian Cancer

Currently, 38 ovarian cancer risk loci, defined by common single nucleotide polymorphisms (SNPs), have been identified by genome-wide association studies (GWAS) reaching genome-wide significance ($p \leq 5 \times 10^{-8}$) (Fig. 1; dark blue boxes) [32••, 40–52]. Thirty were either uncovered ($n = 12$) or confirmed ($n = 18$) in the Oncoarray meta-analysis in a large GWAS of women of European ancestry [32••]. Two loci were confirmed only in women of Han Chinese ancestry (Fig. 1). A recent ovarian cancer risk GWAS performed with East Asian women identified one new locus at genome-wide significance [52].

The conventional threshold for statistical significance used in GWAS is justified by a simple multiple testing correction based on the number of SNPs interrogated in the genotyping chip, typically $p \leq 5 \times 10^{-8}$ (0.05/1,000,000) [53]. However, the threshold is arbitrary, and additional true risk loci may remain hidden. Although large sample sizes of ovarian cancer cases and controls have been assembled, analyses large enough to satisfy Bonferroni corrections for multiple testing remain challenging. Alternative procedures that incorporate biological or clinical information are needed. An alternative approach was used by a recent GWAS with women of East Asian ancestry that revealed three additional loci using a Bayesian false discovery probability (BFDP) < 10% (Fig. 1; light blue boxes) [52, 54].

Similarly, using a less strict threshold ($p < 1 \times 10^{-6}$) as suggestive evidence of association with EOC, 10 new loci emerged in a GWAS of women of African ancestry, which also replicated five previously identified in women of European ancestry [55]. An additional new risk locus, 22q13.1, was identified in a GWAS of Japanese women ($p = 1.05 \times 10^{-7}$), bringing the total number of established and suggested EOC risk loci to 52 (Fig. 1).

These SNPs are located in the X chromosome and distributed across 17 autosomal chromosomes. Most EOC risk-associated SNPs have small effects (typically per-allele odds ratio (OR) < 2.0) creating a challenging scenario for their use for risk stratification. Importantly, carriers of multiple common risk alleles may be at risk similar to carriers of *BRCA* pathogenic variants for which increased screening and surveillance are recommended [56, 57]. In addition to contributing to risk stratification, GWAS-identified risk loci are expected to reveal novel biological processes involved in cancer etiology [58].

In order to exploit these discoveries, it is critical to identify which SNP(s), among a set of credible candidates, are mechanistically driving susceptibility at the locus. During the past several years, guiding principles for functional analysis of GWAS susceptibility loci have emerged [58–69]. The early observation that most associated SNPs were not in gene-coding regions led to the overall hypothesis that trait-associated alleles exert their effects by changing the transcriptional output of target genes [58]. This hypothesis is largely being confirmed, and functional analysis has identified several mechanisms by which gene expression can be modified at each locus (Fig. 1).

Post-GWAS Analysis: Establishing the Functional Chain of Causation

One of the guiding principles of post-GWAS functional analysis is an ideal progression from statistical *association* to *functionality* to *causality*. Risk-associated SNPs represent a set of many SNPs in the locus, one (or some) of which is

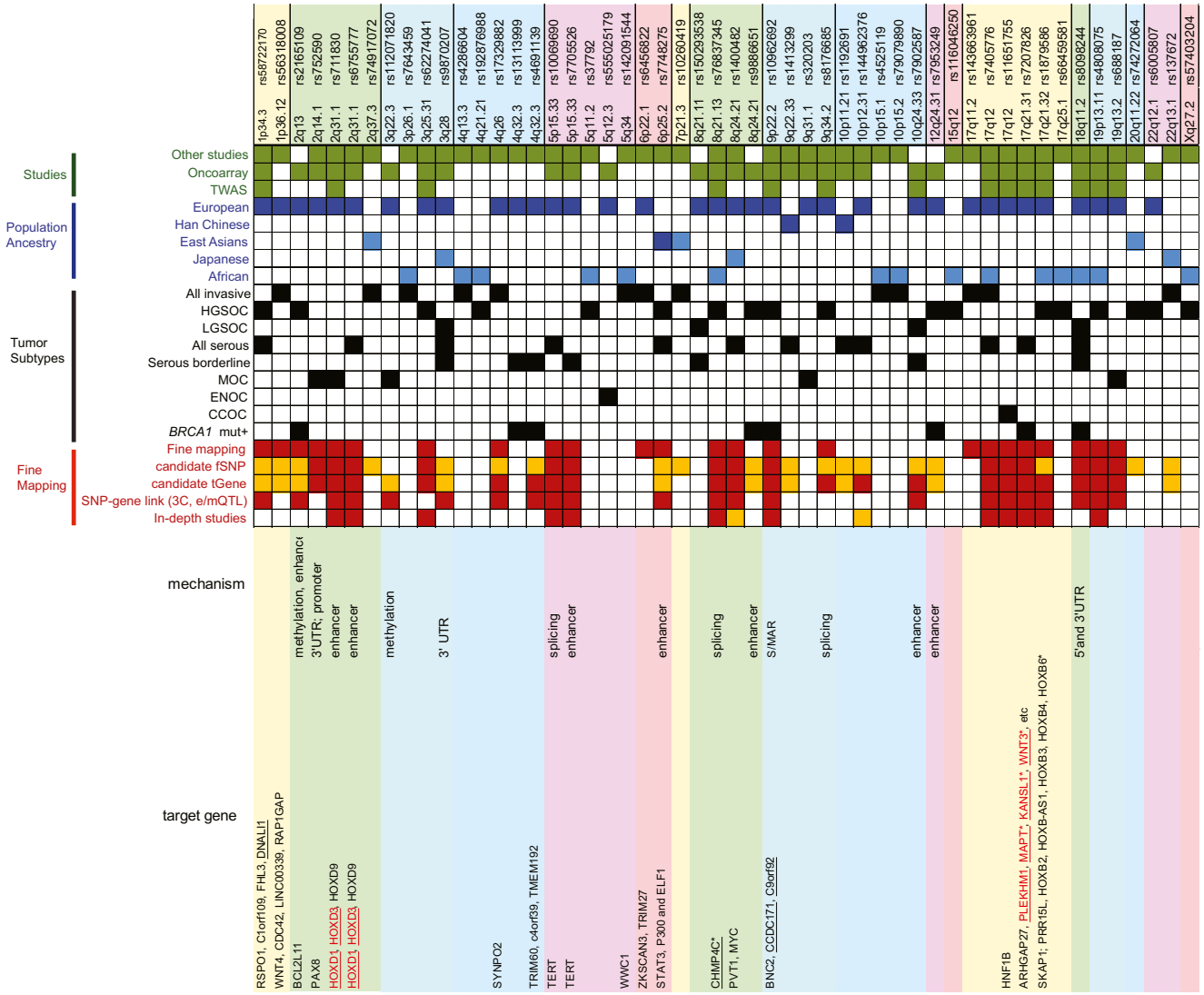

Fig. 1 The functional landscape of common EOC risk loci. Data compilation of all information available for 52 independent common SNP loci (MAF > 1%) associated with EOC risk. Each locus has a lead SNP identified by the different studies (green boxes), the populations in which the association was detected (dark blue boxes [$p < 5 \times 10^{-8}$]; light blue boxes [BFDRP < 10%; or $p < 1 \times 10^{-6}$ for novel and $p < 0.05$ for replicating loci in women of African Ancestry; or $p = 1.05 \times 10^{-7}$ for loci identified in Japanese women]) and in which in tumor subtypes, the association was detected (black boxes). Post-GWAS functional follow-up information is shown in red boxes when finalized and yellow when in progress. Mechanisms by which gene expression can be modified at each locus are shown as the candidate target genes' driving risk at the locus (underlined—TWAS find; red underlined—TWAS find and GWAS functional; asterisk indicates splicingTWAS). S/MAR scaffold/matrix attachment region, UTR untranslated region

hypothesized to be causally linked to disease. Theoretically, the identification of *causal* SNPs should be the first step to define the molecular mechanisms of risk enhancement. However, the statistical and experimental requirements to determine whether a SNP is causal are still ill defined. Here, we refer to *functional* SNP when it displays an allele-specific functional effect and to *credible causal* SNPs when enough collective evidence has been gathered to support a mechanistic link but recognize that these are evolving operational definitions.

Genotyping chip designs can minimize the number of SNPs tested by choosing tag (or index) SNPs to represent large regions of high linkage disequilibrium (haplotype blocks). A drawback of this design is that very few SNPs located at a locus are directly interrogated for association, and it is likely that the tag SNP does not represent the functional SNP causally related to cancer risk. Therefore, fine mapping by dense genotyping and imputation is instrumental to ensure that we are capturing most, if not all, causal SNPs.

The first step is to identify the set of candidate SNPs. When fine mapping data is available, this set is obtained by retaining the most highly significant SNPs at the locus. Fine mapping is also critical to perform conditional analysis and determine whether there are multiple independent signals at the locus.

Alternatively, when fine mapping data is not available, the tag SNP at the locus is used to retrieve all other SNPs in high linkage disequilibrium (LD; typically $r^2 > 0.8$). However, the likelihood of missing a true causal variant, in particular if it is rare, is significantly higher in the latter approach.

With the candidate set defined, SNPs are then inspected for their location in the genome in relation to chromosomal features such as promoters, enhancers, coding regions, and splicing acceptor sites. Some features can be directly deduced by the genetic code (e.g., coding regions, splicing acceptor/donor site) while others are inferred using chromatin modifications previously shown to be enriched in the feature (e.g., enhancers) [70]. This can be manually done for individual loci using the Human Genome Browser, but several available pipelines allow for automatic queries of multiple SNPs [71–74]. All SNPs located in any biological feature are retained for further analysis.

The next series of analyses are centered on testing whether the SNP displays allele-specific changes in the biological feature. For example, causal SNPs in coding regions are expected to show allele-specific differences in assays for protein stability or function (e.g., kinase activity). Causal SNPs in enhancers are expected to show allele-specific differences in a transcription reporter assay or electrophoretic mobility shift assay [75, 76•].

The set of functional SNPs (those that showed allele specific effects) and their location allow for the generation of specific gene target hypothesis at the locus. SNPs in gene features (e.g., coding region, splicing site, 5′ and 3′ untranslated region (UTR)) strongly suggest that the gene in which the SNP lies is the target of regulation. However, caution should be exercised as some genic features could also act as distant enhancers to a different target gene [77]. For SNPs in intergenic features (enhancers), two approaches have proven powerful to connect regulatory elements with their target genes. The first, eQTL (expression quantitative trait loci) analysis, allows for testing risk-associated alleles for their association with expression levels and has been instrumental in identifying regulatory region-promoter links in ovarian cancer [76•, 78]. The other approach probes the proximity between two linearly distant DNA regions in the 3D chromatin. They include chromosome conformation capture and ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) techniques [79, 80]. Because promoter-enhancer cooperativity is achieved through physical proximity, enhancers are expected to be close to their target gene promoter.

Finally, one must establish how the target gene (or its regulation) impinges on the phenotype. This can be initially established using adequate in vitro models that can modulate expression of the candidate gene and be evaluated for oncogenic phenotypes (increased cell growth, decreased apoptosis, multi-nucleated cells, anchorage-independent growth) [47]. Organoids and in vivo models provide opportunities to conduct in-depth experiments on the role of these risk-associated regions [81].

Transcriptome-Wide Association Studies

Transcriptome-wide association studies (TWASs) are a gene-centric approach that uses eQTL to impute gene expression onto genotyped individuals followed by testing the association with disease risk [82–84]. First, eQTL reference panels of samples with associated genotype and gene expression data are used to train a predictive model of expression in the gene's vicinity. Next, using the predictive model, expression is imputed in individuals of known genotype, for example, from GWAS cohorts. Finally, statistical association between predicted gene expression and the trait is tested. From the standpoint of uncovering biology, TWAS has the advantage to directly identify the mechanism (in this case mediator genes) by which genetic variation modifies the phenotype.

The first TWAS for EOC was conducted with high-density genotyping data and RNA sequencing transcriptome data from 53 tissues from GTEx Project (https://gtexportal.org/home/) to train predictive models of genetically regulated expression for 17,121 genes [85, 86]. Data from 97,898 women including 29,396 HGSOC cases were analyzed, and a total of 35 genes spanning 14 loci were associated with risk ($p < 2.21 \times 10^{-6}$) [86]. Of those, 34 were within 1 Mb from previously identified EOC susceptibility variants. Importantly, 11 genes across eight loci corroborated data from previous functional annotation, bioinformatic prediction, or in vitro cellular models (Fig. 1). High expression of *FZD4* at the 11q14.2 locus, the only gene not located within 1 Mb from GWAS-identified EOC risk locus, was associated with increased risk of HGSOC and could constitute a novel risk locus [86].

Of the 34 genes within 1 Mb of previously identified EOC GWAS SNP, only three (*DNALI1*, *HOXD3*, and *CCDC171*) remained statistically significant after conditioning for the top EOC GWAS SNP. This indicates that expression for the vast majority of genes identified by TWAS is regulated by previously identified GWAS SNPs for EOC. Interestingly, the strength of association was only attenuated, suggesting that the SNP(s) with the largest contribution to the regulation of these genes have not yet been identified. At several loci, multiple genes reaching significant association were found to be co-regulated, and further analysis to determine their individual contribution to the phenotype will be needed [86].

Recently, more than 2000 eQTL samples of primary HGSOC, EOC precursor tissues (ovarian and fallopian epithelial cells) and other hormonal-related cancers (breast and prostate cancer) were applied in a multi-tissue TWAS using the largest ovarian cancer GWAS available with 13,037 HGSOC cases and 40,941 controls from OCAC [87]. To train the prediction model, different panels were constructed (each panel is

defined as a tissue-state-phenotype triplet) from 84 breast normal, 654 breast tumors, 70 ovarian normal FT, 115 ovarian normal OS, 201 ovarian tumor HGSOC, and 376 prostate tumors. After Bonferroni correction for 66,764 total tests (reflecting the number of genes and gene-junction models tested), a total of 32 associations for 18 unique genes were detected. Twenty-one out of 32 associations exhibited strong evidence for a single shared causal variant, and only four genes presented the possibility of joint causal variants [87].

The authors expanded their study by performing a splice-transcriptome-wide association study (spTWAS) by integrating splicing quantitative trait loci (QTLs), thus enabling the detection of 74 junction-level models significantly associated with risk. Nine genes were identified in both TWAS studies providing additional support for their role [86, 87].

What Have We Learned?

Approximately 10 years have elapsed since the identification of the first genome-wide significant risk locus at 9p22.2 for EOC. Data and tools for functional dissection of these loci have been slowly built and developed by the effort of large cancer site-specific consortia (such as OCAC) as well as the GAME-ON (Genetic Associations and Mechanisms in Oncology) umbrella consortium which improved the conceptual and technical cross-fertilization among groups working on different cancers. Although only a few loci have been functionally dissected and important challenges still remain, the following themes have emerged.

The *HOX* Axis

The homeobox (HOX) gene family encodes transcription factors that contain homeodomains and function in defining the metazoan body plan, normal vertebrate limb, and organ development [88]. Several *HOX* genes have emerged as strong target gene candidates in ovarian cancer risk. *HOXD9* is the most likely candidate at the 2q31.1, associated with serous and mucinous subtypes [51, 78]. In addition, *HOXD3* emerged in the recent TWAS as independent of previously identified GWAS signals and could constitute a novel risk locus [86].

Interestingly, in the female reproductive organs, several *HOX* genes are expressed along the Mullerian duct tissues. For example, *HOXA9* is expressed in the fallopian tubes while the uterus expresses *HOXA10* during development [89]. Expression of specific *HOX* genes has been implicated in the differentiation of cells in the reproductive tract. Inappropriate ectopic expression of *Hoxa9*, *Hoxa10*, and *Hoxa11* in tumorigenic OSE epithelial cells in mice led to tumors with a papillary serous, endometrioid, and mucinous phenotypes and may constitute an early step in EOC oncogenesis [90].

Using data from only 12 genome-wide significant loci known at the time, a *HOX*-centric network connecting genes from five of the 12 serous EOC risk emerged after gene set enrichment analysis. Six networks centered at 2q31 and 17q21.32 were significantly enriched in genes implicated in EOC. They were centered on *HOXD1*, *HOXD3, HOXB2*, *HOXB5*, *HOXB6*, and *HOXB7* [91]. Remarkably, this network also connected with three other loci through a small number of genes, almost all of which represent the strongest causal gene identified in each locus (*BNC2* at 9p22, *HNF1B* at 17q12, and *ABHD8* at 19p13) [46, 47, 76•, 91].

In addition to *HOX*, other transcription factors have emerged as target genes for several loci such as *PAX8*, *MYC*, and *BNC2* (Fig. 1). Many have been implicated in cell differentiation, morphogenesis, and organogenesis of tissues and organs in the reproductive tract, suggesting that changes in developmental program of precursor cell types might underlie EOC susceptibility at some loci. On the other hand, removal of an 8q24 colorectal cancer risk locus containing a *Myc* enhancer led to a mouse with normal intestinal morphology with no major impact on intestinal cell differentiation, despite being resistant to intestinal tumors [81]. Further studies will be necessary to test this hypothesis in ovarian cancer.

Borderline Serous Subtype, Telomerase, and Telomeres

Based on post-GWAS functional follow-up findings, borderline serous EOC etiology is associated with mechanisms regulating telomere length. SNP rs7705526 intronic to *TERT* is associated with borderline serous EOC at the 5p15.33 locus ($p = 1.3 \times 10^{-15}$) and longer telomeres in leukocytes ($p = 2.3 \times 10^{-14}$) [49]. Furthermore, in vitro luciferase reporter assays to identify transcriptional regulatory regions in ovarian show allele-specific activity suggesting that the mechanism underlying risk at this locus is through changes in an enhancer that acts on the *TERT* promoter [49]. *TERT* encodes the catalytic subunit of telomerase, responsible to maintain telomere length after replication. *TERT* activity is not detectable or low in normal somatic tissue, and *TERT* reactivation has long been proposed as an oncogenic process [92].

Along similar lines, a SNP at the locus 10q24.33 is associated with telomere length and with borderline serous EOC [32••, 44, 93]. eQTL analysis indicates that *OBFC1* is the target gene (Fig. 1). *OBFC1* encodes an accessory factor that stimulates the activity of DNA polymerase-alpha-primase that initiates DNA replication [94]. OBFC1 also participates in telomere-associated complex that binds telomeric single-stranded DNA [95]. The causal relevance of telomere length for risk of cancer was tested through a Mendelian randomization study, which showed strong association for serous low-malignant-potential ovarian cancer with 4.35-fold risk [95% CI; 2.39–7.94] increase (OR [95% CI] per 1-SD change in genetically increased telomere length) [96].

Not All Mechanisms Are Created Equal

At EOC risk loci for which functional analysis has been performed, intragenic and gene proximal SNPs seem to act through changes in promoters, splicing sites, or that affect the stability of the transcript (5′ and 3′ UTR). For intergenic SNPs, allele-specific changes at enhancer elements seem to be an important underlying mechanism (Fig. 1). The recent functional analysis of the 9p22.2 cancer susceptibility locus results supports the original hypothesis that risk-associated SNPs in intergenic regions act by modifying the regulation of target genes [58, 76•]. Surprisingly, regulation of *BNC2* at the locus is likely mediated by a distinct mechanism of cis-regulation by a scaffold/matrix attachment region (S/MAR) which is likely to modulate the local chromatin environment [76•, 97]. Although the identification of the functional SNP that can modulate chromatin architecture at the S/MAR may need further studies, the finding of credible risk SNPs associated to S/MAR raises the possibility that such regions and elements that modify chromatin 3D architecture may contribute to the underlying mechanism of risk at other cancer susceptibility loci.

Where Do We Go from Here?

Complementary approaches such as GWAS and TWAS, combined with ever-improving haplotype imputation, are likely to continue to drive the discovery of additional genomic loci influencing EOC susceptibility in the near future. In particular, effort should be directed at filling the knowledge gap between populations of European and non-European ancestry. This gap is most significant for populations of Hispanic ancestry and filling it expected to have many benefits beyond Hispanic populations for discovery of loci involved in complex traits using multi-ethnic cohorts [98]. In addition, several case-control studies based on germline genome sequencing are now underway and are also likely to identify additional genetic loci implicated in EOC risk.

To cope with the likely increase in the pace of discovery, functional analysis must overcome three significant challenges. The organizational challenge is to create a network of laboratories performing standardized high-throughput assays in an integrated fashion to scale up basic analysis for hundreds of loci at a time. Although large coordinated efforts have their drawbacks, such as difficulty obtaining sustainable funding and issues with allocation of credit, they are also extremely rewarding as the scientific environment is extremely dynamic and the rate of progress is much higher than what any individual lab could accomplish over many years as exemplified by the GAME-ON consortium (https://epi.grants.cancer.gov/gameon/).

The technical challenge is to generate more complex and genetically defined co-cultures, organoid, and in vivo models. It is plausible that some risk loci act in a non-cell autonomous fashion. For example, the changes in gene expression led by risk alleles may prime the organism for cancer not by disturbing the developmental program of a precursor cell but rather by changing how lymphocytes may suppress a tumor in its early stages or through communication between stroma and tumor cells. Co-cultures, organoids, and in vivo models will be required to tease out these paracrine effects. Furthermore, these models, which will also be instrumental to model gene-environment interactions in the lab, will need to take into account genetic ancestry to correctly model risk effects in multi-ethnic cohorts. Given the present lack of ancestral genetic diversity in cancer cell lines [99], the development of more complex representative models is going to be extremely difficult without a concentrated effort.

Finally, a conceptual challenge is to continually expand the repertoire of biologically plausible testable hypothesis for each locus and develop rigorous experimental standards to ensure that the most credible causal SNPs and genes have been found [58]. In addition to the value of GWAS for risk stratification at the population level, functional dissection will add value to the individual because uncovering new mechanisms will open new avenues for prevention and treatment that could be rapidly translated to the clinic to improve EOC outcomes.

Acknowledgments We thank all the women who have donated their time and samples to make possible this work.

Funding Information Funding for ovarian cancer research in the Monteiro Lab came from NIH awards U19 CA148112, R01 CA116167, and U54 CA163068 and from the Rivkin Center and the Moffitt Foundation. This study was funded in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES)—Finance Code 001 and Fundação de Amparo à Pesquisa do Estado do Espírito Santo (FAPES).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Human and Animal Rights This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.

2. SEER Stats fact sheets: ovary cancer: National Cancer Institute; 2019. Available from: https://seer.cancer.gov/statfacts/html/ovary.html.
3. Prat J. Ovarian carcinomas: five distinct diseases with different origins, genetic alterations, and clinicopathological features. Virchows Archiv. 2012;460(3):237–49.
4. Karnezis AN, Cho KR, Gilks CB, Pearce CL, Huntsman DG. The disparate origins of ovarian cancers: pathogenesis and prevention strategies. Nat Rev Cancer. 2016;17:65.
5. Kim J, Park EY, Kim O, Schilder JM, Coffey DM, Cho CH, et al. Cell origins of high-grade serous ovarian cancer. Cancers (Basel). 2018;10(11).
6. Labidi-Galy SI, Papp E, Hallberg D, Niknafs N, Adleff V, Noe M, et al. High grade serous ovarian carcinomas originate in the fallopian tube. Nat Commun. 2017;8(1):1093.
7. Shih Ie M, Kurman RJ. Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. Am J Pathol. 2004;164(5):1511–8.
8. Kurman RJ, Shih IM. Pathogenesis of ovarian cancer: lessons from morphology and molecular biology and their clinical implications. Int J Gynecol Pathol. 2008;27(2):151–60.
9. Kurman RJ, Shih IM. The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. Am J Surg Pathol. 2010;34(3):433–43.
10. Teer JK, Yoder S, Gjyshi A, Nicosia SV, Zhang C, Monteiro ANA. Mutational heterogeneity in non-serous ovarian cancers. Sci Rep. 2017;7(1):9728.
11. Ahmed AA, Etemadmoghadam D, Temple J, Lynch AG, Riad M, Sharma R, et al. Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. J Pathol. 2010;221(1):49–56.
12. Ho ES, Lai CR, Hsieh YT, Chen JT, Lin AJ, Hung MH, et al. p53 mutation is infrequent in clear cell carcinoma of the ovary. Gynecol Oncol. 2001;80(2):189–93.
13. Kuo KT, Mao TL, Jones S, Veras E, Ayhan A, Wang TL, et al. Frequent activating mutations of PIK3CA in ovarian clear cell carcinoma. Am J Pathol. 2009;174(5):1597–601.
14. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609–15.
15. Bonome T, Lee JY, Park DC, Radonovich M, Pise-Masison C, Brady J, et al. Expression profiling of serous low malignant potential, low-grade, and high-grade tumors of the ovary. Cancer Res. 2005;65(22):10602–12.
16. Bast RC Jr, Hennessy B, Mills GB. The biology of ovarian cancer: new opportunities for translation. Nat Rev Cancer. 2009;9(6):415–28.
17. Patch A-M, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole–genome characterization of chemoresistant ovarian cancer. Nature. 2015;521:489.
18. Reid BM, Permuth JB, Sellers TA. Epidemiology of ovarian cancer: a review. Cancer Biol Med. 2017;14(1):9–32.
19. Webb PM, Jordan SJ. Epidemiology of epithelial ovarian cancer. Best Pract Res Clin Obstet Gynaecol. 2017;41:3–14.
20. Pharoah PD, Ponder BA. The genetics of ovarian cancer. Best Pract Res Clin Obstet Gynaecol. 2002;16(4):449–68.
21. Stratton JF, Pharoah P, Smith SK, Easton D, Ponder BA. A systematic review and meta-analysis of family history and risk of ovarian cancer. Br J Obstet Gynaecol. 1998;105(5):493–9.
22. Jones MR, Kamara D, Karlan BY, Pharoah PDP, Gayther SA. Genetic epidemiology of ovarian cancer and prospects for polygenic risk prediction. Gynecol Oncol. 2017;147(3):705–13.
23. Lilyquist J, LaDuca H, Polley E, Davis BT, Shimelis H, Hu C, et al. Frequency of mutations in a large series of clinically ascertained ovarian cancer cases tested on multi-gene panels compared to reference controls. Gynecol Oncol. 2017;147(2):375–80.
24. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer. 2004;4(11):850–60.
25. Song H, Dicks E, Ramus SJ, Tyrer JP, Intermaggio MP, Hayward J, et al. Contribution of germline mutations in the RAD51B, RAD51C, and RAD51D genes to ovarian cancer in the population. J Clin Oncol. 2015;33(26):2901–7.
26. Ramus SJ, Song H, Dicks E, Tyrer JP, Rosenthal AN, Intermaggio MP, et al. Germline mutations in the BRIP1, BARD1, PALB2, and NBN genes in women with ovarian cancer. J Natl Cancer Inst. 2015;107(11).
27. Jervis S, Song H, Lee A, Dicks E, Tyrer J, Harrington P, et al. Ovarian cancer familial relative risks by tumour subtypes and by known ovarian cancer genetic susceptibility variants. J Med Genet. 2014;51(2):108–13.
28. Lu Y, Ek WE, Whiteman D, Vaughan TL, Spurdle AB, Easton DF, et al. Most common 'sporadic' cancers have a significant germline genetic component. Hum Mol Genet. 2014;23(22):6112–8.
29. Amos CIDJ, Wang Z, Byun J, Schumacher FR, Gayther SA, Hunter DJ, et al. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. Cancer Epidemiology Biomarkers & Prevention. 2016; under review.
30. Bahcall OG. iCOGS collection provides a collaborative model. Foreword Nature genetics. 2013;45(4):343.
31. Houlston RS, Peto J. The search for low-penetrance cancer susceptibility alleles. Oncogene. 2004;23(38):6471–6.
32.•• Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nature genetics. 2017;49(5):680–91 **This paper describes the largest GWAS analysis of Epithelial Ovarian Cancer risk using samples from the Ovarian Cancer Association Consortium (OCAC) and the Consotium of Investigators of Modifiers of BRCA1/2 (CIMBA). It identifies 12 new susceptibility loci.**
33. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005;434(7035):913–7.
34. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature. 2005;434(7035):917–21.
35. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Mol Cell. 2010;40(2):179–204.
36. Kennedy RD, D'Andrea AD. DNA repair pathways in clinical practice: lessons from pediatric cancer susceptibility syndromes. J Clin Oncol. 2006;24(23):3799–808.
37. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med. 2009;361(2):123–34.
38. Fong PC, Yap TA, Boss DS, Carden CP, Mergui-Roelvink M, Gourley C, et al. Poly(ADP)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. J Clin Oncol. 2010;28(15):2512–9.
39. Kar SP, Berchuck A, Gayther SA, Goode EL, Moysich KB, Pearce CL, et al. Common genetic variation and susceptibility to ovarian cancer: current insights and future directions. Cancer Epidemiol Biomarkers Prev. 2017.
40. 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(9):996–1000.
41. Chen K, Ma H, Li L, Zang R, Wang C, Song F, et al. Genome-wide association study identifies new susceptibility loci for epithelial ovarian cancer in Han Chinese women. Nat Commun. 2014;5:4682.

42. Pharoah PD, Tsai YY, Ramus SJ, Phelan CM, Goode EL, Lawrenson K, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat Genet. 2013;45(4):362–70.
43. Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nat Genet. 2010;42(10):874–9.
44. Kuchenbaecker KB, Ramus SJ, Tyrer J, Lee A, Shen HC, Beesley J, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. Nat Genet. 2015;47(2):164–71.
45. Bolton KL, Tyrer J, Song H, Ramus SJ, Notaridou M, Jones C, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. Nat Genet. 2010;42(10):880–4.
46. Lawrenson K, Kar S, McCue K, Kuchenbaeker K, Michailidou K, Tyrer J, et al. Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. Nat Commun. 2016;7:12675.
47. Shen H, Fridley BL, Song H, Lawrenson K, Cunningham JM, Ramus SJ, et al. Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. Nat Commun. 2013;4:1628.
48. Permuth-Wey J, Lawrenson K, Shen HC, Velkova A, Tyrer JP, Chen Z, et al. Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31. Nat Commun. 2013;4:1627.
49. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat Genet. 2013;45(4):371–84.
50. Couch FJ, Wang X, McGuffog L, Lee A, Olswold C, Kuchenbaecker KB, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. PLoS Genet. 2013;9(3):e1003212.
51. Kelemen LE, Lawrenson K, Tyrer J, Li Q, Lee JM, Seo JH, et al. Genome-wide significant risk associations for mucinous ovarian carcinoma. Nat Genet. 2015;47(8):888–97.
52. Lawrenson K, Song F, Hazelett DJ, Kar SP, Tyrer J, Phelan CM, et al. Genome-wide association studies identify susceptibility loci for epithelial ovarian cancer in east Asian women. Gynecol Oncol. 2019;153(2):343–55.
53. Krzywinski M, Altman N. Comparing samples—part II. Nat Methods. 2014;11:355.
54. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet. 2007;81(2):208–27.
55. Manichaikul A, Peres LC, Wang XQ, Barnard ME, Chyn D, Sheng X, et al. Identification of novel epithelial ovarian cancer loci in women of African Ancestry. Int J Cancer Journal international du cancer. 2019.
56. Jervis S, Song H, Lee A, Dicks E, Harrington P, Baynes C, et al. A risk prediction algorithm for ovarian cancer incorporating BRCA1, BRCA2, common alleles and other familial effects. J Med Genet. 2015;52(7):465–75.
57. Turnbull C, Sud A, Houlston RS. Cancer genetics, precision prevention and a call to action. Nat Genet. 2018;50(9):1212–8.
58. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011;43(6):513–8.
59. Monteiro AN, Freedman ML. Lessons from postgenome-wide association studies: functional analysis of cancer predisposition loci. J Intern Med. 2013;274(5):414–24.
60. Sur I, Tuupanen S, Whitington T, Aaltonen LA, Taipale J. Lessons from functional analysis of genome-wide association studies. Cancer Res. 2013;73(14):4180–4.
61. Knight JC. Approaches for establishing the function of regulatory genetic variants involved in disease. Genome Med. 2014;6(10):92.
62. Fachal L, Dunning AM. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. Curr Opin Genet Dev. 2015;30:32–41.
63. Rivandi M, Martens JWM, Hollestelle A. Elucidating the underlying functional mechanisms of breast cancer susceptibility through post-GWAS analyses. Front Genet. 2018;9:280.
64. Cannon ME, Mohlke KL. Deciphering the emerging complexities of molecular mechanisms at GWAS loci. Am J Hum Genet. 2018;103(5):637–53.
65. Nishizaki SS, Boyle AP. Mining the unknown: assigning function to noncoding single nucleotide polymorphisms. Trends Genet. 2017;33(1):34–45.
66. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101(1):5–22.
67. Paul DS, Soranzo N, Beck S. Functional interpretation of noncoding sequence variation: concepts and challenges. Bioessays. 2014;36(2):191–9.
68. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet. 2013;93(5):779–97.
69. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. Am J Hum Genet. 2018;102(5):717–30.
70. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.
71. UCSC GB. Available from: genome.ucsc.edu.
72. Coetzee SG, Rhie SK, Berman BP, Coetzee GA, Noushmehr H. FunciSNP: an R/bioconductor tool integrating functional noncoding data sets with genetic association studies to identify candidate regulatory SNPs. Nucleic Acids Res. 2012.
73. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(Database issue):D930–4.
74. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22(9):1790–7.
75. Buckley M, Gjyshi A, Mendoza-Fandino G, Baskin R, Carvalho RS, Carvalho MA, et al. Enhancer scanning to locate regulatory regions in genomic loci. Nat Protoc. 2016;11(1):46–60.
76.• Buckley MA, Woods NT, Tyrer JP, Mendoza-Fandino G, Lawrenson K, Hazelett DJ, et al. Functional analysis and fine mapping of the 9p22.2 ovarian cancer susceptibility locus. Cancer research. 2019;79(3):467–81 **This paper which describes functional dissection of the 9p22.2 risk locus identifies a scaffold/matrix attachment region as a possible new mechanism for regulation of gene expression in susceptibility loci.**
77. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190–5.
78. Lawrenson K, Li Q, Kar S, Seo JH, Tyrer J, Spindler TJ, et al. Cis-eQTL analysis and functional validation of candidate susceptibility genes for high-grade serous ovarian cancer. Nat Commun. 2015;6:8234.
79. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289–93.
80. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295(5558):1306–11.
81. Sur IK, Hallikas O, Vaharautio A, Yan J, Turunen M, Enge M, et al. Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. Science. 2012;338(6112):1360–3.

82. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47(9):1091–8.
83. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48(3):245–52.
84. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 2018;9(1):1825.
85. Consortium GT. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):648–60.
86. Lu Y, Beeghly-Fadiel A, Wu L, Guo X, Li B, Schildkraut JM, et al. A transcriptome-wide association study among 97,898 women to identify candidate susceptibility genes for epithelial ovarian cancer risk. Cancer Res. 2018;78(18):5419–30.
87. Gusev A, Lawrenson K, Lin X, Lyra PC Jr, Kar S, Vavra KC, et al. A transcriptome-wide association study of high-grade serous epithelial ovarian cancer identifies new susceptibility genes and splice variants. Nat Genet. 2019;51(5):815–23.
88. Hrycaj SM, Wellik DM. Hox genes and evolution. F1000Res. 2016;5.
89. Taylor HS, Vanden Heuvel GB, Igarashi P. A conserved Hox axis in the mouse and human female reproductive system: late establishment and persistent adult expression of the Hoxa cluster genes. Biol Reprod. 1997;57(6):1338–45.
90. Cheng W, Liu J, Yoshida H, Rosen D, Naora H. Lineage infidelity of epithelial ovarian cancers is controlled by HOX genes that specify regional identity in the reproductive tract. Nat Med. 2005;11(5):531–7.
91. Kar SP, Tyrer JP, Li Q, Lawrenson K, Aben KK, Anton-Culver H, et al. Network-based integration of GWAS and gene expression identifies a HOX-centric network associated with serous ovarian cancer risk. Cancer Epidemiol Biomarkers Prev. 2015;24(10):1574–84.
92. Hiyama E, Gollahon L, Kataoka T, Kuroi K, Yokoyama T, Gazdar AF, et al. Telomerase activity in human breast tumors. J Natl Cancer Inst. 1996;88(2):116–22.
93. Levy D, Neuhausen SL, Hunt SC, Kimura M, Hwang SJ, Chen W, et al. Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. Proc Natl Acad Sci U S A. 2010;107(20):9293–8.
94. Casteel DE, Zhuang S, Zeng Y, Perrino FW, Boss GR, Goulian M, et al. A DNA polymerase-{alpha}{middle dot}primase cofactor with homology to replication protein A-32 regulates DNA replication in mammalian cells. J Biol Chem. 2009;284(9):5807–18.
95. Wan M, Qin J, Songyang Z, Liu D. OB fold-containing protein 1 (OBFC1), a human homolog of yeast Stn1, associates with TPP1 and is implicated in telomere length regulation. J Biol Chem. 2009;284(39):26725–31.
96. Telomeres Mendelian Randomization C, Haycock PC, Burgess S, Nounu A, Zheng J, Okoli GN, et al. Association between telomere length and risk of cancer and non-neoplastic diseases: a Mendelian randomization study. JAMA Oncol. 2017;3(5):636–51.
97. Choi J, Brown KM. A dynamic cis-regulation pattern underlying epithelial ovarian cancer susceptibility. Cancer Res. 2019;79(3):439–40.
98. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019;570(7762):514–8.
99. Dutil JCZ, Monteiro AN, Teer JK, Eschrich SA. An interactive resource to probe genetic diversity and estimated ancestry in cancer cell lines. Cancer Res. 2018; Provisionally accepted.