Dear Editor,

Ovarian cancer (OC) is a heterogeneous cancer primarily beginning in an ovary. It often results in tumor cells that can metastasis to invade to other organs of human body. In 2012, it was reported 239,000 new cases and 152,000 deaths worldwide.\textsuperscript{1,2} As the eighth-most common cause of death, OC is regarded as a ‘silent killer’ due to absence of symptoms in early stages.\textsuperscript{1,2} Integrating literature evidence is critical for cross-validation of cancer mechanisms. During the past decade, thousands of small-scale studies have been published regarding the OC-related genes and how they function in various cellular events. Unfortunately, there is no such an integrated resource for OC genes with literature evidence available. Here, we developed OCGene with substantial literature evidences and regulatory data at http://ocgene.bioinfo-minzhao.org/.

The gene collection on the OC were mainly based on 12 data sources, including OMIM, GAD, one online database focusing on OC genes,\textsuperscript{3} one expert review,\textsuperscript{4} gene manually curation from GeneRif database, genome-wide association studies from GWASCatalog and six candidate gene lists produced by various large-scale genomic platforms on OC from the TCGA\textsuperscript{5} (detailed in the Supplementary Methods). In total, 2067 genes associated with 2819 PubMed abstracts were compiled. The comprehensive functional annotations were associated with all the genes, including gene expression profiles, homologous genes in other model species, protein–protein interaction, somatic mutations and methylations.

As shown in Figure 1a, a typical OCGene gene entry contains eight categories information by clicking the label ‘General information,’ ‘Literature,’ ‘Expression,’ ‘lncRNA,’ ‘Regulation,’ ‘Mutation,’ ‘Homolog’ and ‘Interaction’ on the top. Highlighted summaries of curated literature evidence are provided in the ‘Literature’ page. For the precomputed coexpression pattern between OC-related genes and lncRNAs, user can obtain the correlation coefficients and the corrected statistical $P$-values. In addition, the expression plots of the corresponding lncRNAs were also integrated from MiTranscriptome database\textsuperscript{6} (detailed in the Supplementary Methods). The ‘Homolog’ page is used to map those human genes with evidence to other model species, including mouse, rat, zebrafish and fruit fly. Moreover, the ‘Regulation’ page is used to classify regulatory information, including interactions with transcription factors, abundance of post-translational modification information and methylation information.

On the basis of the gene ranking results of ToppGene\textsuperscript{7} (detailed in the Supplementary Methods), the top 100-ranked genes tend to have multiple evidences from different data sources (Figure 1b). These highly overlapping relationship of top-ranked genes support the accuracy of our data. In addition, using reliable pathway-based gene interaction data, we constructed a more comprehensive cellular map for OC based on top 100-ranked genes (Figure 1c). Of the 79 nodes, 74 of them are from our OCGene. The remaining five are the novel genes that may potentially bridge the OC-related gene to fully implement their cellular function. In conclusion, the majority of curated top-ranked OC-related genes are linked to each other in a highly modular structure, which may accelerate the signaling processing for oncogenesis.

In conclusion, we developed a user-friendly literature-based online database for in-depth analysis of OC genes with comprehensive annotation. The features of our database include: (i) manually curated experimentally verified genes from thousands of literatures; (ii) comprehensively integrated from nine reliable data sources; and (iii) precomputed regulatory pattern involved transcription factors, microRNA and long non-coding RNAs. These comprehensive annotations and precomputed regulatory information highlighted the importance of OCGene to elucidating unexplored field of OC development. All the information with functional features are available in a plain text format that is free to download.

Conflict of interest
The authors declare no conflict of interest.
Acknowledgements. We thank Professor Richard Burns for his review of this manuscript.

1. Siegel R et al. CA Cancer J Clin 2012; 62: 10–29.
2. Lengyel E. Am J Pathol 2010; 177: 1053–1064.
3. Kaur M et al. Nucleic Acids Res 2009; 37: D820–D823.
4. Bast RC Jr. et al. Nat Rev Cancer 2009; 9: 415–428.
5. The Cancer Genome Atlas Network. Nature 2011; 474: 609–615.
6. Iyer MK et al. Nat Genet 2015; 47: 199–208.
7. Chen J et al. Nucleic Acids Res 2009; 37: W305–W311.

Supplementary Information accompanies this paper on Cell Death and Disease website (http://www.nature.com/cddis)