Computational and Structural Biotechnology Journal 20 (2022) 1068–1076

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

SDC: An integrated database for sex differences in cancer

Long-Fei Zhao [a,b,c,d,e], Jin-Ge Zhang [a,b,c,d,e], Feng-Yu Qi [a,b,c,d,e], Wei-Yan Hou [f], Yin-Rui Li [a,b,c,d,e], Dan-Dan Shen [a,b,c,d,e], Li-Juan Zhao [a,b,c,d,e], Lin Qi [f,*], Hong-Min Liu [a,b,c,d,e,*], Yi-Chao Zheng [a,b,c,d,e,*]

[a] Key Lab of Advanced Drug Preparation Technologies, Ministry of Education of China, Zhengzhou University, 100 Kexue Avenue, Zhengzhou, Henan 450001, China
[b] State Key Laboratory of Esophageal Cancer Prevention & Treatment, Zhengzhou University, 100 Kexue Avenue, Zhengzhou, Henan 450001, China
[c] Key Laboratory of Henan Province for Drug Quality and Evaluation, Zhengzhou University, 100 Kexue Avenue, Zhengzhou, Henan 450001, China
[d] Institute of Drug Discovery and Development, Zhengzhou University, 100 Kexue Avenue, Zhengzhou, Henan 450001, China
[e] School of Pharmaceutical Sciences, Zhengzhou University, 100 Kexue Avenue, Zhengzhou, Henan 450001, China
[f] School of Information Engineering, Zhengzhou University, 100 Kexue Avenue, Zhengzhou, Henan 450001, China

ARTICLE INFO

Article history:
Received 1 November 2021
Received in revised form 29 January 2022
Accepted 24 February 2022
Available online 26 February 2022

Keywords:
Cancer
Sex
Differences
Database

ABSTRACT

Sex differences are evident in the incidence and mortality of diverse cancers. With the development of personalized approaches in cancer treatment, the impact of sex differences has not been systematically incorporated into preclinical and clinical cancer research. The molecular mechanisms underlying sex differences in cancer have not been elucidated. Here, we developed the first database of Sex Differences in Cancer (SDC), a web-based public database that integrates resources from multiple databases, including The Cancer Genome Atlas (TCGA), Genotype-Tissue Expression Project (GTEx), UCSC Xena, Broad Institute Cancer Cell Line Encyclopedia (CCLE), Genomics of Drug Sensitivity in Cancer (GDSC). SDC contains 27 types of cancers, 6 types of molecular data, more than 10,000 donors, 977 cancer cell lines were used to analyze sex differences among cancers. It provides five main modules: Survival and phenotype, Molecular differences, Signatures and pathways, Therapy response, Download. Users can download the all the visualized results and raw data after analysis. Collectively, SDC is the first integrated database to analyze sex differences in cancer on the web server, which will strengthen our understanding of the role of sex in cancers. It is implemented in Shiny-server and freely available for public use at http://sdc.anticancer.xyz.

1. Introduction

Global cancer statistics [1] (https://gco.iarc.fr) and previous reports [2–5] showed that males have higher incidence, mortality, and poorer outcomes than females in the case of various types of cancer. Moreover, the male gender is an independent risk factor associated with distant metastasis and prognosis [6]. Esophagogastric cancer [7], gastric cancer [8,9], melanoma [10], and lung cancer [11,12] with sex bias have been focused. However, the molecular basis for these observed disparities is poorly understood.

While one study has suggested that sex differences in cancer may arise through the effect of circulating sex hormones [13], it has also been suggested that sex bias is derived from genetic and epigenetic differences arising from the influence of the sex chromosomes independent of sex hormones [14]. For example, the X-linked lysine demethylase 6A and 5C, as well as Y-linked paralogs lysine demethylase 6C and 5D may be regulators of incidence and prognosis for sex-specific cancer [10,15,16]. Moreover, the whole-genome study uncovered sex differences in mutation density, tumor evolution, and mutation signatures on the noncoding autosomal genome [17]. Additionally, sex differences also play a role in distinct immunity as females have stronger innate and adaptive immune responses than males, reducing the risk of mortality due to cancer [18,19]; males are more likely to benefit from immune checkpoint inhibitors [18,20]. All these findings may be related to somatic mutation, immune microenvironment, and the immune system, but they need to be explored further.

Altogether, these studies underlined the prevalence of sex differences at a molecular level, in cancer and suggested that these differences are sex-specific but not lifestyle-dependent. However, in studies of cancer risk, prognosis, and therapeutic response, sex is still underexplored as a relevant variable. Additionally, these

* Corresponding authors at: Key Lab of Advanced Drug Preparation Technologies, Ministry of Education of China, Zhengzhou University, 100 Kexue Avenue, Zhengzhou, Henan 450001, China (H.M. Liu and Y.C. Zheng).
E-mail addresses: ielqi@zzu.edu.cn (L. Qi), liuhm@zzu.edu.cn (H.-M. Liu), yichaozheng@zzu.edu.cn (Y.-C. Zheng).

https://doi.org/10.1016/j.csbj.2022.02.023

studies on sex effect were limited to individual genes, single molecular data types, and single cancer lineages, which lack association and are not attributed to signaling pathways. There is also a lack of exploration of sex differences in common molecular subtypes.

Therefore, building a public database is vital for collecting omics data and conducting integrative and in-depth analysis with sex differences. Here, we developed the SDC, the first web-based public database for accessing reproducible comprehensive differences analysis between the sexes that covers modules including phenotypic, molecular, and therapeutic response. These results can be quickly queried and presented in a customized manner via the website (http://sdc.anticancer.xyz), to utilize sex insights in the study of molecular oncology research and cancer therapy.

2. Methods

Most of the analysis is run in R 3.6, and the code is open-sourced on GitHub (https://github.com/longfei8533/SDC) for reference and repetition.

2.1. Data source

Biological data are mainly from TCGA, GTEx, CCLE and GDSC. Part of the data are from the collation and analysis results of UCSC Xena (http://xena.ucsc.edu/). Cancer incidence and mortality data were manually downloaded from "CANCER TODAY" page at The Global Cancer Observatory (https://gco.iarc.fr). Overall Survival (OS), Progression-Free Interval (PFI), Disease-Free Interval (DFI), and Disease-Specific Survival (DSS) were defined and calculated in previous study [21], and the supplementary data was download for further analysis. Molecular subtype data were obtained from 26 published papers collected by UCSC Xena. Tumor microenvironment (TME) subtypes data obtained from previous study [22]. Cancers with <10 samples in one sex were excluded from analysis.

2.2. Tumor purity

ESTIMATE is a method that uses gene expression profiles of 141 immune genes and 141 stromal genes to infer tumor cellularity [23]. Tumor purity was added as a covariable to the model for correction during mRNA, miRNA, and copy number variation (CNV) analysis.

2.3. Survival analysis and molecular subtypes

Survival analysis was performed with the Log-rank test. Fisher's exact test was performed to test the sex bias of different molecular subtypes.

2.4. mRNA and miRNA

The package DESeq2 [24] provides methods to test the differential expression of gene between female and male, and gives the P and log2FoldChange value based on a model using the negative binomial distribution. The tumor purity was added as a covariable to the model for correction. log2FoldChange <0 indicates that the expression of gene is male-biased; log2FoldChange >0 indicates that the expression of gene is female-biased. Tissue types in the GTEx database were converted into TCGA codes for easy comparison of sex differences between tumors and normal tissues.

The target genes of differentially expressed miRNA were predicted by TargetScanHuman [25]. Counts per million (CPM) was computed and used to normalize mRNA and miRNA data. Then, the Pearson correlation coefficient was used to analyze the correlations.

2.5. DNA methylation

The ChAMP package [26] was used for the analysis of Illumina Methylation beadarray data (TCGA 450 k). Differential methylation probes, differential methylation regions and differential methylation were shown individually.

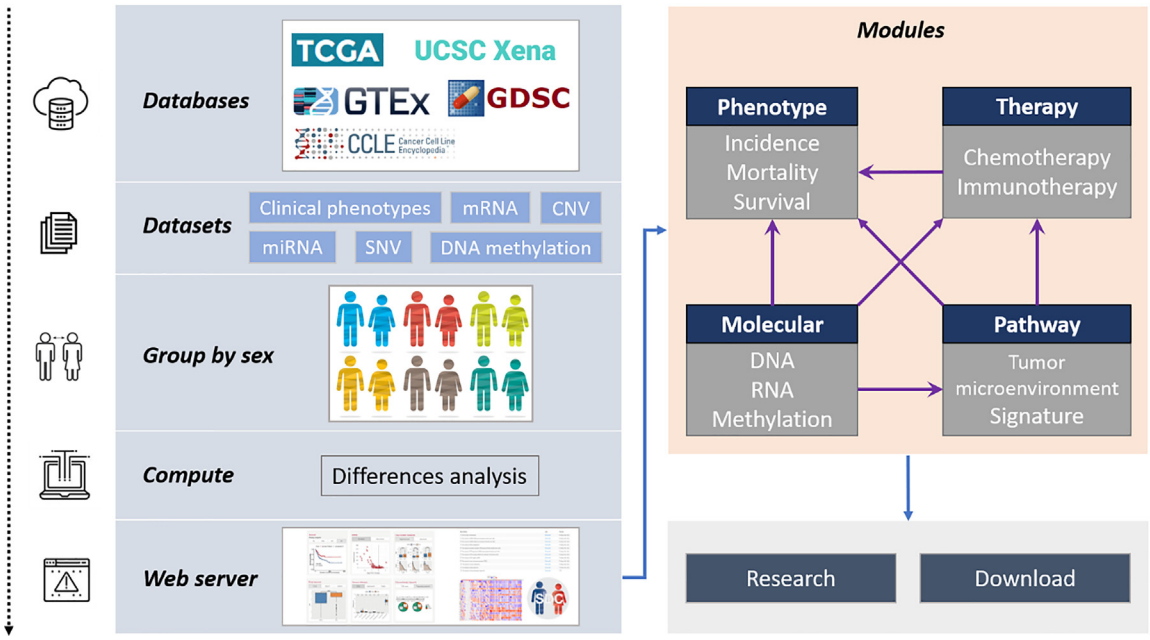
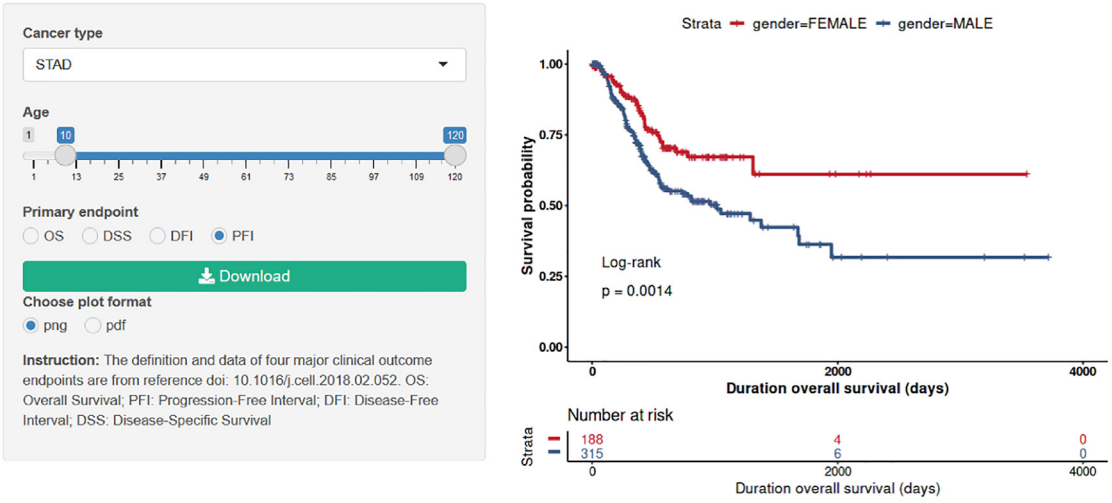


Fig. 1. The design of SDC.

Fig. 2. Home page function. A. Module descriptions and shortcuts are provided on the home page. Estimated age-standardized incidence and mortality rates (World) in 2020 for multiple types of cancers are displayed at the bottom of the page. **B.** Multiple analysis results can be quickly viewed by selecting one cancer type in the red box on the "Quick view" panel.

A



B

	FEMALE	MALE	p
n	188	315	
Age (median [IQR])	68.50 [60.00, 74.00]	66.00 [58.00, 73.00]	0.055
Race (%)			0.462
---- [Not Evaluated]	1 (0.5)	2 (0.6)	
---- [Unknown]	21 (11.2)	40 (12.7)	
---- ASIAN	32 (17.0)	67 (21.3)	
---- BLACK OR AFRICAN AMERICAN	3 (1.6)	11 (3.5)	
---- NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	0 (0.0)	1 (0.3)	
---- WHITE	131 (69.7)	194 (61.6)	
Clinical stage (%)			0.107
---- [Discrepancy]	1 (0.5)	11 (3.5)	
---- [Unknown]	16 (8.5)	16 (5.1)	
---- Stage I	0 (0.0)	2 (0.6)	

Fig. 3. Survival and phenotype module. A. Kaplan–Meier survival analysis between sexes. **B.** Patient clinical characteristics.

2.6. Somatic mutation and copy number variation

Fisher's exact test was performed on 2×2 contingency table generated from female and male cohorts to find differentially mutated genes. The total mutation burden was defined as the number of single nucleotide variant (SNV) and INDEL in the whole genome. Significance was calculated by Wilcox rank-sum test. In gene level, the Gistic2 copy number was used to estimate copy number differences between female and male. Further, CNV burdens including broad CNV scores (BCS), focal CNV scores (FCS) and global CNV scores (GCS), which were computed by CNApp [27]. Finally, we recalculated copy number segments in the 1 M interval to obtain the region profile of CNV. Suppose there are N segments on the 1 M interval, with length L bases and copy number C.

$$Averagecopynumber = \sum_{n=1}^{N}(Cn \times Ln)/\Sigma_{n=1}^{N}(Ln)$$

2.7. Tumor microenvironment

We used the IOBR R package [28] to decode tumor microenvironment (TME) contexture using 7 published methodologies CIBERSORT [29], ESTIMATE [23], quanTIseq [30], TIMER [31], IPS [32], MCPCounter [33], EPIC [34]. These methods are based on computational inference of gene expression profiles.

2.8. Signatures and pathways

255 published signature gene sets were collected by IOBR, involving tumor microenvironment, tumor metabolism, m6A, exosomes, microsatellite instability, and tertiary lymphoid structure. We calculated the signature scores for all samples in the TCGA using three methodologies of ssGSEA, PCA and Z-score.

The PARADIGM algorithm integrates pathway, expression and copy number data to infer the activation of pathway features within a superimposed pathway network structure. This dataset is ssGSEA scores for 1387 constituent pathways, then Z transformed [35].

2.9. Treatment response

Drug sensitivity data in cancer cell lines from the CCLE and GDSC were integrated, and these cancer cell lines were grouped by sex. Then, IC_{50} of the individual drug to male or female were

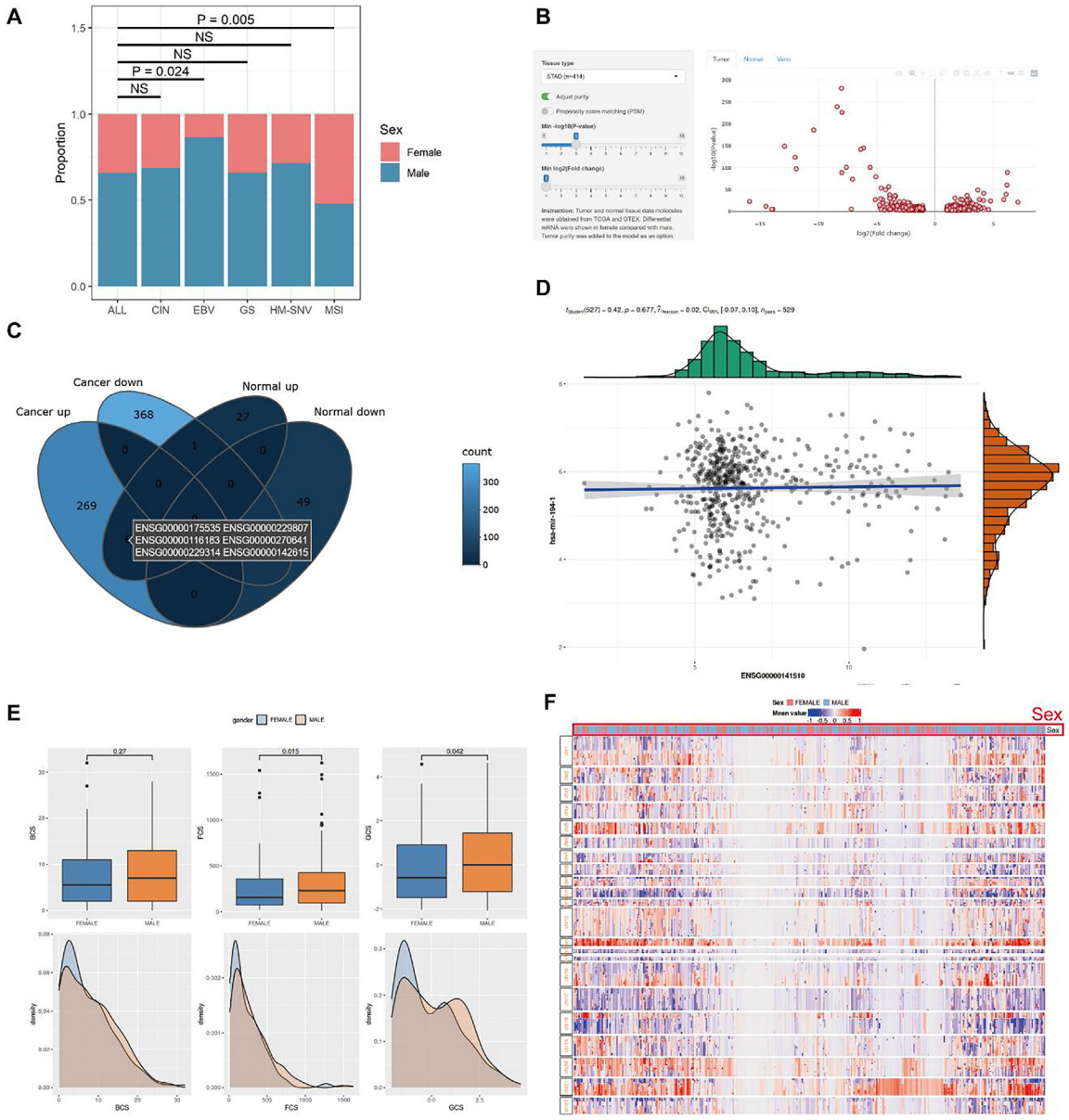


Fig. 4. Molecular differences module. A. Sex bias of molecular subtypes. **B.** Differential expressed genes were filtered and shown as a scatter plot. **C.** Intersection of differential genes in tumor and normal tissue. **D.** Correlation between miRNA and target mRNA expression. **E.** The differences of CNV scores for broad, focal and global CNV burdens. **F.** Heatmap of the CNV region profile in all samples with sex annotation.

compared. Tumor Immune Dysfunction and Exclusion (TIDE) [36] was used to predict the immunotherapy response from the gene expression profiles of cancer tissue samples. Conserved pan-cancer microenvironment subtypes can predict response to immunotherapy [22]. Molecular profiles of cancer can be clustered into four distinct microenvironments termed (1) immune-enriched, fibrotic (IE/F); (2) immune-enriched, non-fibrotic (IE); (3) fibrotic (F); and (4) immune-depleted (D).

3. Implementation and results

3.1. The database overall design

The web was developed with R and Shiny following a modular and robust design. The data were preprocessed locally with some of the results loaded into memory for speed and another part stored in SQLite database (Fig. 1). SDC uses a series of user-friendly interfaces to display results. Because of the Bootstrap and Shiny dependency, the interfaces are dynamic and interactive on a variety of devices with different screen sizes, and five main modules were provided, including (i) Survival and phenotype, (ii) Molecular differences, (iii) Signatures and pathways, (iv) Therapy response, (v) Download. All customized resulting images can be downloaded directly in PNG or PDF format. For "Download" module, all pre-processed data can be downloaded, most of which are saved as R data, and can be quickly loaded into the R environment by the R function "readRDS". All the raw data used for analysis is also linked for easy access.

3.2. Web interface and usage

The home page provides a module description and a quick jump (Fig. 2A). The global multi-cancer morbidity and mortality histogram is displayed at the bottom of the page, providing a quick

and intuitive view of gender differences. The "Quick View" function is for the quickly selection of 1 in 27 tumor types for preview (Fig. 2B).

For "Survival and phenotype" module, users can select different age stratifications and different endpoints to study differences in survival in specific cancers (Fig. 3A). Statistical analysis was performed on clinical characteristics of cancer patients and presented in the table (Fig. 3B). Findings from Fig. S1-S4 showed that there was no significant difference in survival between females and males in lung cancer, while, in esophageal and gastric cancers, females have a better prognosis than males).

Six data categories of analysis are included in the "Molecular differences" module. (i) For molecular subtypes, fisher's exact test results revealed sex-biased molecular subtypes (Fig. 4A). (ii) For mRNA, users can screen differentially expressed mRNAs under tumor purity correction. By inputting the minimum P value and fold change, users can further filter the highly reliable differential genes, and the results can be showed as a scatter plot (Fig. 4B). Data in mRNA level from normal GTEx tissues were calculated in the same way, so it is possible to discover which differential expression of gene are inherent before tumor development and which are caused after tumor development. Meanwhile, the results are displayed in the "Venn" panel (Fig. 4C). The detailed description of all genes can be jumped to GeneCard (https://www.genecards.org/) with the aid of the hyperlink. (iii) miRNAs are filtered and presented in the same way. The score for target gene prediction of miRNA is shown in the table in the "Target" panel, and the correlation between genes and miRNAs can be further viewed on the "Correlation" panel (Fig. 4D). (iv) Differential DNA methylation in three levels, including CpG island, methylation regions and methylation blocks, were shown in the scatter plot and table. (v) Somatic mutation (SNV and INDEL) results that mainly include differences in mutation frequency of gene and total tumor mutation burden. (vi) For CNV, the Gistic2 copy number was used to estimated copy number differences between females and males in gene level. And, CNV burdens including BCS, FCS and GCS, which has been reported to be potential biomarkers for prognosis and immunotherapy [37,38], were calculated with purity correction (Fig. 4E). Further, we recalculated copy number segments (log2(tumor/normal)) in the 1 M interval to obtain the region profile of CNV (Fig. 4F).

For "Signatures and pathways" module, "Tumor microenvironment", "Signatures" and "Pathways" three functions can be used. (i) Due to the importance of tumor-infiltrating immune cells in cancer treatment efficacy and patient prognosis [39], score for the tumor-infiltrating immune cells was calculated using 7 pub-

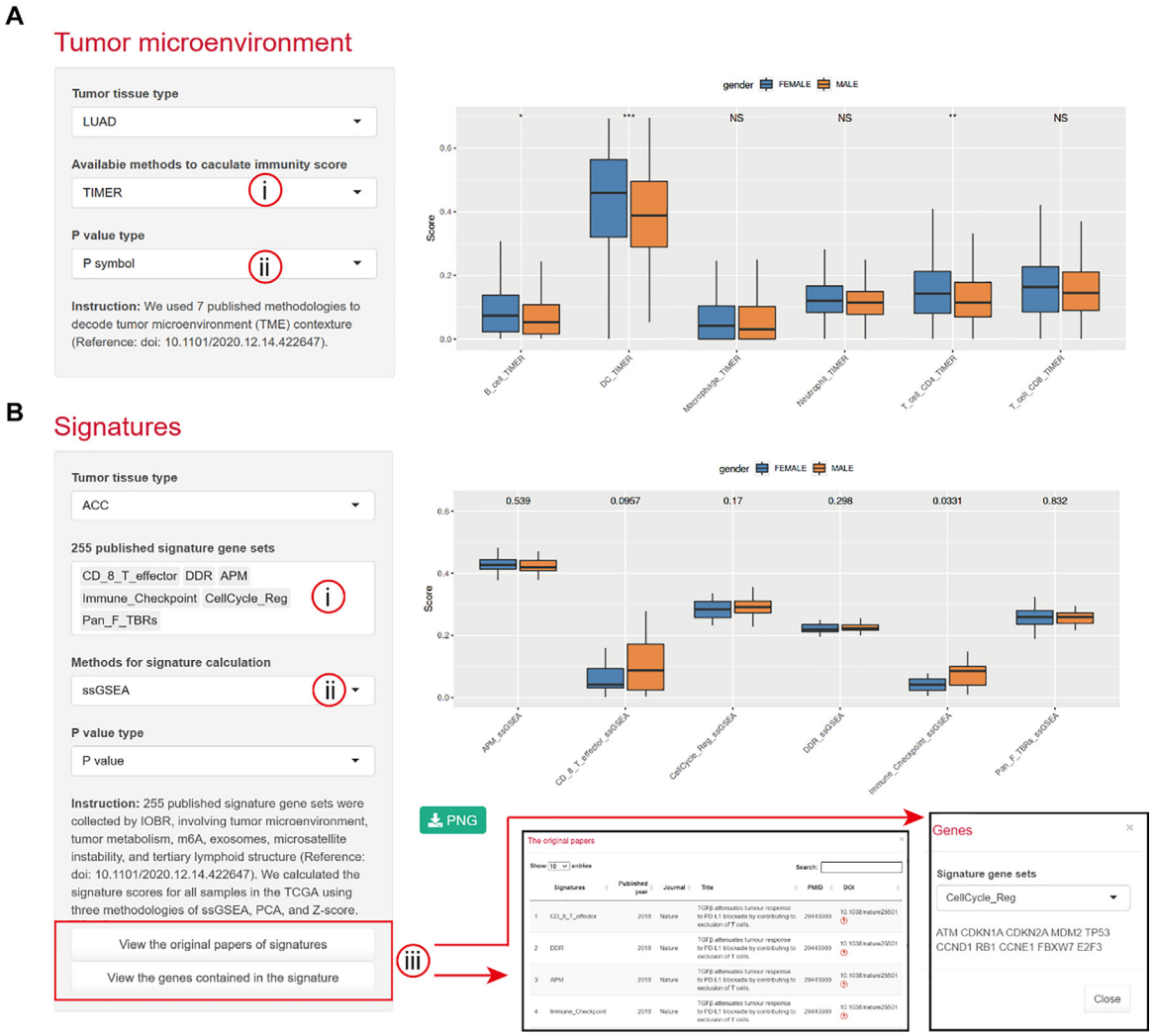


Fig. 5. Signatures and Pathways module. A. Analysis of differences in the tumor microenvironment. (i) is to select one of the seven published methods. (ii) can convert whether the plot is displayed as a p-value or asterisks. **B.** Analysis of the signal signature difference. (i) indicates that tumor-related signatures can be added and removed for presentation on the same plot. (ii) is three methods for signature calculation. (iii) is the two buttons to display the signature source and genes within the signature.

lished methods [23,29–34]. Different analysis was then performed based on the score and users can choose to display the exact P value on the box plot or a more intuitive symbol (Fig. 5A). (ii) In the "Signatures" panel, 255 tumor-associated signatures scores were calculated using ssGSEA, PCA and Z-score methods and compared between the sexes. Users can click on the two buttons at the bottom of the page to find the original article and signature genes (Fig. 5B). (iii) In the "Pathways" panel, different analysis was performed on pathway activity scores of 1387 constituent PARADIGM pathways.

For "Therapy response" module, "Chemotherapy response" and "Immunotherapy response" panels were included. In the "Chemotherapy response" panel, users can select one of three databases and then select a drug or target to explore the sensitivity of tumor cell lines to the drug. The results were presented as a box plot and volcano plot (Fig. 6A). In the "Immunotherapy response" panel, patients with a TIDE score <0 were defined as positive responders to immunotherapy (Fig. 6B). Patients with IE and IE/F subtypes are more likely to benefit from immunotherapy, when using tumor microenvironment subtypes to predict immunotherapy response.

4. Future developments and discussion

As we have found that there were significant phenotypic differences between the two sexes that were disruptive to causal inferences, more covariates will be added to the model in the future or pre-treat the population with propensity score matching, so that to obtain a better and unbiased estimation. Our next plan is to integrate more samples and more types of biological data (e.g. immunotherapy patients, proteomics and epigenetics data) and mine relationships between modules (As shown in Fig. 1), the purple line represents the possible association) to provide more comprehensive and accurate insights. In the coming future, users' data uploading function and analysis service will be provided to assess gender bias and generate reports.

Overall, SDC is the first comprehensive and user-friendly web database to study molecular differences, pathway differences, and therapeutic response differences between males and females in cancer. Based on this web database, we aim to integrate more comprehensive data and provide more evidence to explain the sex differences in cancer systematically, so that to facilitate antitu-

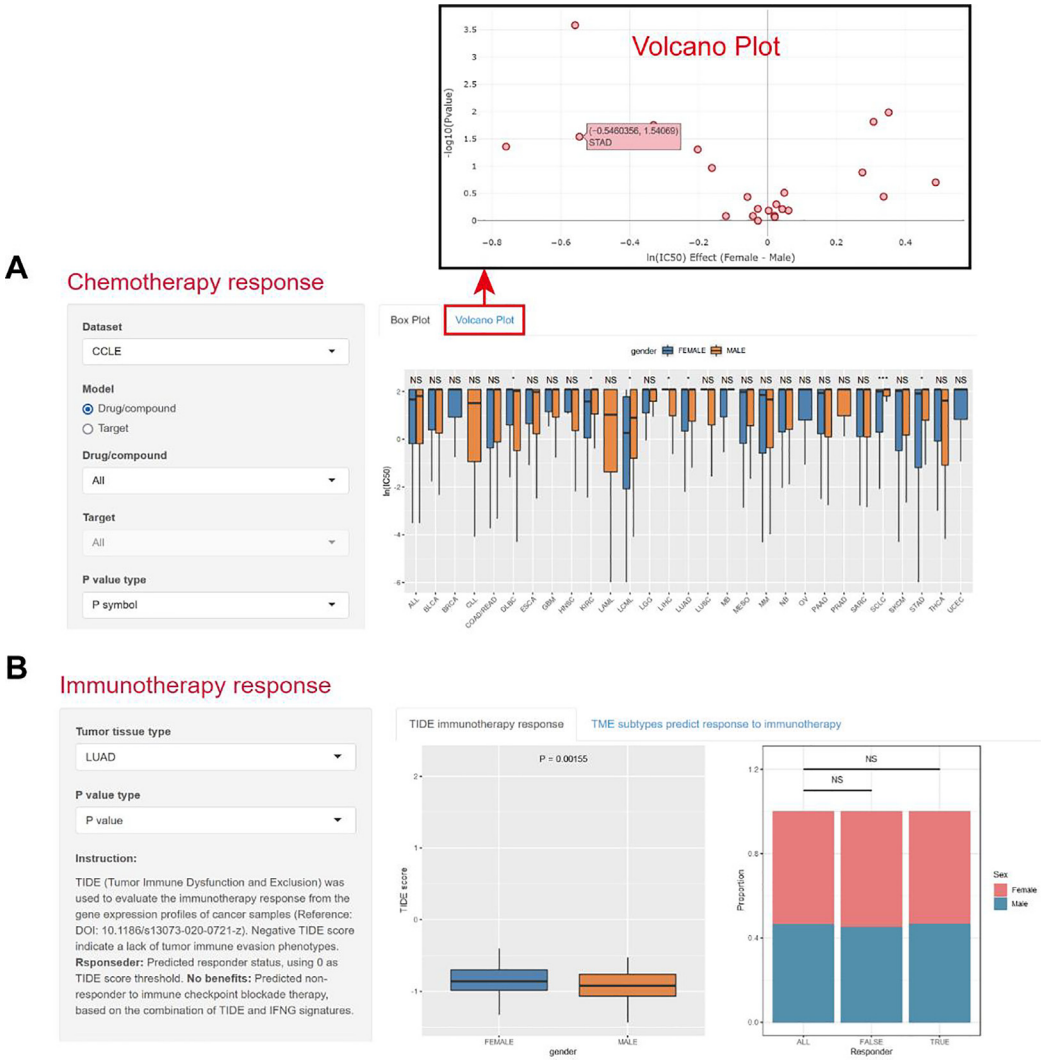


Fig. 6. Therapy response module. A. Chemotherapy response results. **B.** Immunotherapy response results. The TIDE scores and response status between male and female were compared and displayed as box plot and histogram plot, respectively.

mor studies and clinical practices. Fully understanding the mechanisms underlying sexual dimorphism in human cancers will be benefit to both basic cancer research and translational application for sex-specific diagnosis, prognosis, and treatment of cancer, which will contribute to personalized precision medicine.

Data and code availability

This study did not produce any raw data. Processed data, including reanalysis of published datasets are available in the download page at www.anticancer.xyz. Scripts used to generate results are available at https://github.com/longfei8533/SDC.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank TCGA, GTEx, UCSC Xena, CCLE, and GDSC teams. This work was supported by the National Key Research Program (No. 2018YFE0195100); the National Natural Science Foundation of China (No. 82020108030 and No. U21A20416); Science and Technology Innovation Talents of Henan Provincial Education Department (No. 19IRTSTHN001, China); Basic and Frontier Technology Research Project of Henan Province (No. 212102310313, China); Youth Supporting Program from Henan Province (No. 2021HYTP060); Basic Research of the Key Project of the High Education from the Education Department of Henan Province (No. 22ZX008, China).

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, et al. GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 2018;68(2018):394–424. https://doi.org/10.3322/caac.21492.

[2] Zheng D, Trynda J, Williams C, Vold JA, Nguyen JH, Harnois DM, et al. Sexual dimorphism in the incidence of human cancers. BMC Cancer 2019;19:684. https://doi.org/10.1186/s12885-019-5902-z.

[3] Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, et al. Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries, The Lancet. Global Health 2018;6:e555–67. https://doi.org/10.1016/S2214-109X(18)30127-X.

[4] Radkiewicz C, Johansson ALV, Dickman PW, Lambe M, Edgren G. Sex differences in cancer risk and survival: A Swedish cohort study. Eur. J. Cancer 2017;84:130–40. https://doi.org/10.1016/j.ejca.2017.07.013.

[5] Innos K, Padrik P, Valvere V, Aareleid T. Sex differences in cancer survival in Estonia: a population-based study. BMC Cancer 2015;15:72. https://doi.org/10.1186/s12885-015-1080-9.

[6] Wang Y, Zeng Z, Tang M, Zhang M, Bai Y, Cui H, et al. Sex Disparities in the Clinical Characteristics, Synchronous Distant Metastasis Occurrence and Prognosis: A Pan-cancer Analysis. J. Cancer 2021;12:498–507. https://doi.org/10.7150/jca.50536.

[7] Athauda A, Nankivell M, Langley RE, Alderson D, Allum W, Grabsch HI, et al. Impact of sex and age on chemotherapy efficacy, toxicity and survival in localised oesophagogastric cancer: A pooled analysis of 3265 individual patient data from four large randomised trials (OE02, OE05, MAGIC and ST03). Eur. J. Cancer 2020;137:45–56. https://doi.org/10.1016/j.ejca.2020.06.005.

[8] Li H, Wei Z, Wang C, Chen W, He Y, Zhang C. Gender Differences in Gastric Cancer Survival: 99,922 Cases Based on the SEER Database. J. Gastrointest. Surg. 2020;24:1747–57. https://doi.org/10.1007/s11605-019-04304-y.

[9] Kim HW, Kim J-H, Lim BJ, Kim H, Kim H, Park JJ, et al. Sex Disparity in Gastric Cancer: Female Sex is a Poor Prognostic Factor for Advanced Gastric Cancer. Ann. Surg. Oncol. 2016;23:4344–51. https://doi.org/10.1245/s10434-016-5448-0.

[10] Emran AA, Nsengimana J, Punnia-Moorthy G, Schmitz U, Gallagher SJ, Newton-Bishop J, et al. Study of the Female Sex Survival Advantage in Melanoma-A Focus on X-Linked Epigenetic Regulators and Immune Responses in Two Cohorts. Cancers (Basel) 2020;12. https://doi.org/10.3390/cancers12082082.

[11] Hanna D, Sugamori KS, Bott D, Grant DM. The impact of sex on hepatotoxic, inflammatory and proliferative responses in mouse models of liver carcinogenesis. Toxicology 2020;442:. https://doi.org/10.1016/j.tox.2020.152546152546.

[12] Wang S, Zhang J, He Z, Wu K, Liu X-S. The predictive power of tumor mutational burden in lung cancer immunotherapy response is influenced by patients' sex. Int. J. Cancer 2019;145:2840–9. https://doi.org/10.1002/ijc.32327.

[13] Clocchiatti A, Cora E, Zhang Y, Dotto GP. Sexual dimorphism in cancer. Nat. Rev. Cancer 2016;16:330–9. https://doi.org/10.1038/nrc.2016.30.

[14] Credendino SC, Neumayer C, Cantone I. Genetics and Epigenetics of Sex Bias: Insights from Human Cancer and Autoimmunity. Trends Genet. 2020;36:650–63. https://doi.org/10.1016/j.tig.2020.06.016.

[15] Tricarico R, Nicolas E, Hall MJ, Golemis EA. X- and Y-Linked Chromatin-Modifying Genes as Regulators of Sex-Specific Cancer Incidence and Prognosis. Clin. Cancer Res. 2020;26:5567–78. https://doi.org/10.1158/1078-0432.CCR-20-1741.

[16] Dunford A, Weinstock DM, Savova V, Schumacher SE, Cleary JP, Yoda A, et al. Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat. Genet. 2017;49:10–6. https://doi.org/10.1038/ng.3726.

[17] Li CH, Prokopec SD, Sun RX, Yousif F, Schmitz N, Boutros PC. Sex differences in oncogenic mutational processes. Nat. Commun. 2020;11:4330. https://doi.org/10.1038/s41467-020-17359-2.

[18] Conforti F, Pala L, Bagnardi V, de Pas T, Martinetti M, Viale G, et al. Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis. Lancet Oncol 2018;19:737–46. https://doi.org/10.1016/S1470-2045(18)30261-4.

[19] Irelli A, Sirufo MM, D'Ugo C, Ginaldi L, de Martinis M. Sex and Gender Influences on Cancer Immunotherapy Response. Biomedicines 2020;8. https://doi.org/10.3390/biomedicines8070232.

[20] Castro A, Pyke RM, Zhang X, Thompson WK, Day C-P, Alexandrov LB, et al. Strength of immune selection in tumors varies with sex and age. Nat. Commun. 2020;11:4128. https://doi.org/10.1038/s41467-020-17981-0.

[21] Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell 2018;173:400–416.e11. https://doi.org/10.1016/j.cell.2018.02.052.

[22] Bagaev A, Kotlov N, Nomie K, Svekolkin V, Gafurov A, Isaeva O, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell 2021;39:845–865.e7. https://doi.org/10.1016/j.ccell.2021.04.014.

[23] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612. https://doi.org/10.1038/ncomms3612.

[24] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.

[25] Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015;4. https://doi.org/10.7554/eLife.05005.

[26] Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics 2014;30:428–30. https://doi.org/10.1093/bioinformatics/btt684.

[27] Franch-Expósito S, Bassaganyas L, Vila-Casadesús M, Hernández-Illán E, Esteban-Fabró R, Díaz-Gay M, et al. CNApp, a tool for the quantification of copy number alterations and integrative analysis revealing clinical implications. Elife 2020;9. https://doi.org/10.7554/eLife.50267.

[28] D. Zeng, Z. Ye, G. Yu, J. Wu, Y. Xiong, R. Zhou, W. Qiu, N. Huang, L. Sun, J. Bin, Y. Liao, M. Shi, W. Liao, IOBR: Multi-omics Immuno-Oncology Biological Research to decode tumor microenvironment and signatures, 2020.

[29] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453–7. https://doi.org/10.1038/nmeth.3337.

[30] Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. Genome Med 2019;11:34. https://doi.org/10.1186/s13073-019-0638-6.

[31] Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res. 2020;48:W509–14. https://doi.org/10.1093/nar/gkaa407.

[32] Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017;18:248–62. https://doi.org/10.1016/j.celrep.2016.12.019.

[33] Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol 2016;17:218. https://doi.org/10.1186/s13059-016-1070-5.

[34] Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. Elife 2017;6:. https://doi.org/10.7554/eLife.26476e26476.

[35] Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell 2018;173:291–304.e6. https://doi.org/10.1016/j.cell.2018.03.022.

[36] Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, et al. Large-scale public data reuse to model immunotherapy response and resistance. Genome Med. 2020;12:21. https://doi.org/10.1186/s13073-020-0721-z.
[37] Hieronymus H, Murali R, Tin A, Yadav K, Abida W, Moller H, et al. Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. Elife 2018;7. https://doi.org/10.7554/eLife.37294.
[38] Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Ann. Oncol. 2019;30:44–56. https://doi.org/10.1093/annonc/mdy495.
[39] Li B, Severson E, Pignon J-C, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol 2016;17. https://doi.org/10.1186/s13059-016-1028-7.