

# Advancing personalized cancer therapy: Onko\_DrugCombScreen—a novel Shiny app for precision drug combination screening

Jingyu Yang <sup>1</sup>, Meng Wang<sup>2</sup>, Jürgen Dönitz <sup>1,3,4</sup>, Björn Chapuy<sup>2,5</sup>, Tim Beißbarth <sup>1,3,4,\*</sup>

<sup>1</sup>Medical Bioinformatics, University Medical Center Göttingen, Göttingen, 37077, Germany

<sup>2</sup>Department of Hematology, Oncology, and Cancer Immunology, Charité—University Medical Center Berlin, Campus Benjamin Franklin, Berlin, 12203, Germany

<sup>3</sup>Campus Institute Data Science (CIDAS), University of Göttingen, Göttingen, 37073, Germany

<sup>4</sup>Comprehensive Cancer Center Niedersachsen (CCCN), Göttingen, 37075, Germany

<sup>5</sup>Department of Hematology and Medical Oncology, Georg-August University Göttingen, Göttingen, 37075, Germany

\*To whom correspondence should be addressed. Email: [tim.beissbarth@bioinf.med.uni-goettingen.de](mailto:tim.beissbarth@bioinf.med.uni-goettingen.de)

## Abstract

Identifying and validating genotype-guided drug combinations for a specific molecular subtype in cancer therapy represents an unmet medical need and is important in enhancing efficacy and reducing toxicity. However, the exponential increase in combinatorial possibilities constrains the ability to identify and validate effective drug combinations. In this context, we have developed Onko\_DrugCombScreen, an innovative tool aiming at advancing precision medicine based on identifying significant drug combination candidates in a target cancer cohort compared to a comparison cohort. Onko\_DrugCombScreen, inspired by the molecular tumor board process, synergizes drug knowledgebase analysis with various statistical methodologies and data visualization techniques to pinpoint drug combination candidates. Validated through a TCGA-BRCA case study, Onko\_DrugCombScreen has demonstrated its proficiency in discerning established drug combinations in a specific cancer type and in revealing potential novel drug combinations. By enhancing the capability of drug combination discovery through drug knowledgebases, Onko\_DrugCombScreen represents a significant advancement in personalized cancer treatment by identifying promising drug combinations, setting the stage for the development of more precise and potent combination treatments in cancer care. The Onko\_DrugCombScreen Shiny app is available at [https://rshiny.gwdg.de/apps/onko\\_drugcombscreen/](https://rshiny.gwdg.de/apps/onko_drugcombscreen/). The Git repository can be accessed at [https://gitlab.gwdg.de/MedBioinf/mtb/onko\\_drugcombscreen](https://gitlab.gwdg.de/MedBioinf/mtb/onko_drugcombscreen).

## Introduction

Cancer treatment is an intricate field, with the ongoing quest to develop therapies that effectively target the disease while minimizing side effects. Synergistic drug combinations aim to reduce the concentration of each drug while achieving the same therapeutic effect, thereby minimizing side effects. This represents a key advancement in cancer therapy [1]. This approach aims to overcome the limitations of single-agent treatments by improving efficacy and reducing the likelihood of drug resistance [2]. However, the task of identifying optimal drug combinations is complicated by the significant variability in tumor types and patient responses, along with the complexities of cancer biology, the high dimensionality of data, and the number of drug combinations far beyond what is possible for clinical testing [3, 4]. These challenges make it difficult to predict which combinations are most likely to be effective in specific cancer molecular type cohorts compared to other molecular subtypes, necessitating advanced computational tools and extensive experimental validation to navigate the vast landscape of potential drug combinations and tailor treatments to cohort patients' needs [5].

Molecular tumor boards (MTBs) are crucial in personalizing cancer treatment, integrating multidisciplinary expertise to interpret genetic data and guide treatment decisions based on a patient's unique tumor characteristics [6–8]. Drawing inspiration from the methodologies employed by MTBs, the utilization of drug databases alongside detailed drug's level of evidence information emerges as a crucial strategy in advancing patient-specific treatment recommendations [6–8]. This approach not only facilitates the identification of the most suitable drugs for individual patients based on their genetic profiles but also sets the foundation for drug combination prediction based on patient cohorts. Applying statistical methods based on drug databases to the set of recommended drugs for these patient cohorts enables researchers and clinicians to predict more accurately effective drug combinations. This strategy underscores a significant shift toward a data-driven and evidence-based framework to optimize combination therapy for cancer patients, leveraging the increasingly available genetic and pharmacological data to enhance treatment efficacy and patient outcomes.

In recent decades, an increasing application of computational approaches has been developed for the prediction of

Received: July 18, 2024. Revised: December 24, 2024. Editorial Decision: January 6, 2025. Accepted: January 13, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of NAR Genomics and Bioinformatics.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [reprints@oup.com](mailto:reprints@oup.com) for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

drug combinations and their effects. Preuer *et al.* developed DeepSynergy, a deep learning-based approach that accurately predicts drug combination synergies for cancer treatments, significantly surpassing traditional performance methods [9]. Similarly, Wang *et al.* introduced DeepDDS, a deep learning model that employs graph neural networks and attention mechanisms to precisely predict and prioritize synergistic drug combinations for cancer treatments, achieving the advantage of enhanced interpretability through chemical substructure analysis [10]. Cheng *et al.* demonstrated that a network-based methodology, concentrating on the relative configuration of drug–target modules in connection to disease modules, can effectively prioritize potentially efficacious drug combinations for complex diseases such as cancer [11]. GAECDS, presented by Li *et al.*, is an innovative approach combining graph autoencoders and convolutional neural networks to accurately predict drug synergy, showing superior performance in identifying efficacious drug combinations [12]. Concurrently, numerous classical machine learning (ML) models have also exhibited performance comparable to deep learning methods, demonstrating their robustness and utility in this complex domain. Gayvert *et al.* showcased that a random forest model, utilizing single drug dose responses as features, could accurately predict drug pair synergy and effectiveness in mutant *BRAF* melanomas [13]. Janizek *et al.* introduced TreeCombo, an XGBoost-based approach that leverages the power of gradient boosting to improve predictive accuracy, outperforming DeepSynergy by using drug physiochemical features and cancer cell line gene expression data. The use of XGBoost, which combines multiple decision trees to make robust predictions, demonstrated comparable efficacy to deep learning on medium-scale datasets, while offering the additional benefits of reduced complexity in hyperparameter tuning and enhanced interpretability through TreeSHAP, a feature attribution method that identifies the contribution of each variable in a clear and consistent manner [14]. However, current pre-clinical screenings primarily focus on the synergistic effects of drug combinations, often overlooking key factors for clinical success, such as potential toxicity and selective efficacy against tumors [3]. At the same time, there is a clear lack of innovative computational solutions to demonstrate their feasibility and benefits in translational applications, especially in the field of cancer, where there is an urgent need to identify combination therapies suitable for specific cancer group patients based on patient-specific biomarkers [15, 16].

In this paper, we present an Onko\_DrugCombScreen Shiny app designed to address this gap in cancer therapy, which could predict the significant drug combination candidates based on the target patient cohort statistical analysis against the comparison cohort. The primary goal of Onko\_DrugCombScreen extends beyond merely providing treatment recommendations based on drug databases such as GDKD [17], CIViC [18], and OncoKB [19]. It integrates statistical methods and data visualization to analyze recommendations based on extensive drug databases within the target cancer cohort and comparison cohort genetic data, thereby uncovering potential drug combinations and mapping them onto cell line data, providing a robust basis for clinical drug screening. Based on the drug evidence levels in the knowledge database for medications, one can directly ascertain whether the variant mapping drugs are selective at the cancer types in the target patient cohort, and the previous studies collected in the database can save workload on drug toxicity analysis.

**Table 1.** Contingency table for Fisher’s exact test analysis

	Target	Comparison	Row total
Drug1 + Drug2	$a$	$b$	$a + b$
Non-Drug1 + Drug2	$c$	$d$	$c + d$
Column total	$a + c$	$b + d$	$a + b + c + d$ ( $n$ )

In this table,  $a$  represents the number of patients in the target tumor cohort receiving Drug1 + Drug2 co-recommendation,  $b$  is the number of comparison cohort patients receiving Drug1 + Drug2 co-recommendation,  $c$  is the number of patients in the target tumor cohort not receiving Drug1 + Drug2 co-recommendation, and  $d$  is the number of comparison cohort patients not receiving Drug1 + Drug2 co-recommendation.

This brings renewed hope for the clinical translation of cancer-type-specific drug combination therapies.

## Materials and methods

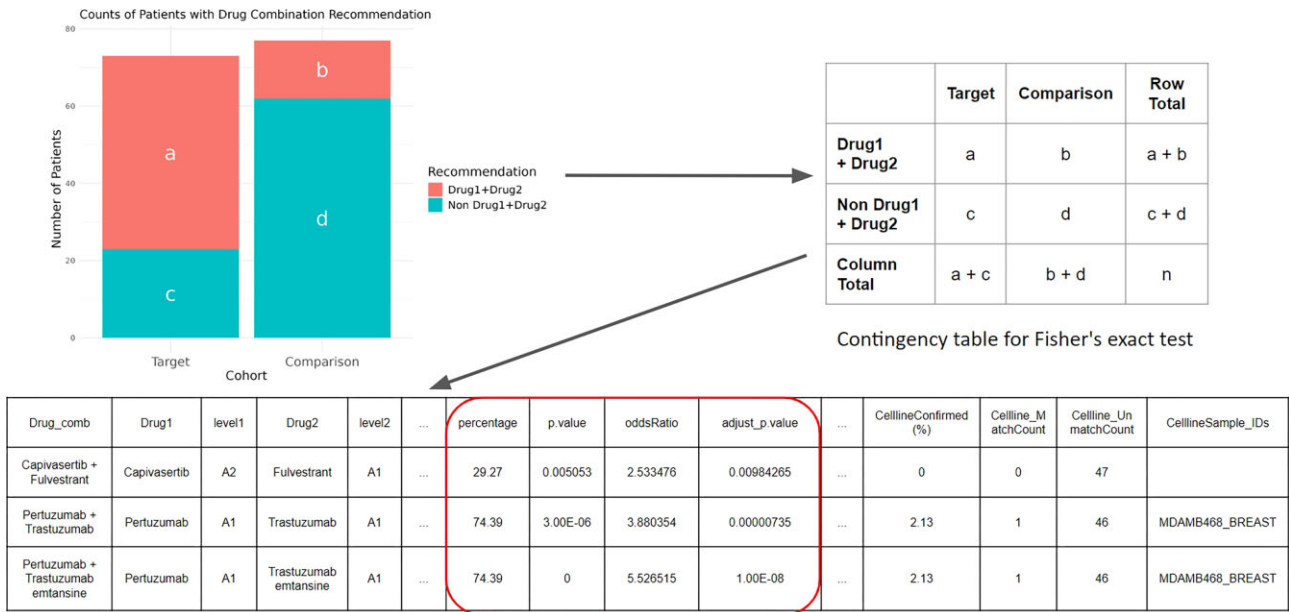
### Fisher’s exact test in cancer subtype recommendations

Besides predicting single drugs, clinicians and researchers are interested in determining whether two drugs are simultaneously recommended for the target tumor type and exhibit significant differences compared to the comparison tumor group. Here, we defined co-recommended drugs as candidate drug combinations that are presented in the Drug\_comb column in the DrugComb analysis table. We then counted the number of patients in the target tumor cohort and the comparison cohort for each candidate drug combination. Subsequently, we used these four counts to construct a contingency table (Table 1) and performed a Fisher’s exact test for each candidate drug combination. By analyzing the  $P$ -value and odds ratio that are circled with a red rectangle in Fig. 1 results obtained from Fisher’s exact test, we can determine whether the occurrence of a candidate drug combination is significantly different and assess the magnitude of this difference. The  $P$ -values were adjusted using the Benjamini–Hochberg method, as reflected in the `adjust_p.value` column, to account for multiple hypothesis testing and control the false discovery rate. Additionally, we report drug combination candidate recommendations for cell line data in the final four columns in the DrugComb analysis table to assist with wet-lab validation (Fig. 1).

Here, Fisher’s exact test serves as a robust statistical method to determine the significance of the association between the candidate drug combination and the tumor type. Fisher’s exact test is particularly suitable for small sample sizes and for datasets where the assumptions of chi-squared tests are not met.

### Drug level of evidence

Here, we adopted the MTB drug’s level of evidence category approach proposed by Perera-Bel *et al.* [6]. As shown in Table 2. “A” signifies evidence for the same cancer type, while “B” indicates evidence for any other cancer type. Horizontally, Level 1 represents evidence supported by regulatory agencies or clinical guidelines. Level 2 includes evidence from clinical trials. Finally, Level 3 consists of preclinical trial evidence. Therefore, based on the different target cancer types of drugs and their respective clinical evidence, six levels of drug evidence are derived: A1, A2, A3, B1, B2, and B3. With this drug level of evidence, the selection of recommended drugs for specific



**Figure 1.** A “drug co-recommendation” refers to the recommendation of two drugs for a single patient, exemplified as Drug1 + Drug2. The bar plot illustrates the counts of such co-recommendations within the target cancer subtype cohort compared to the comparison cancer subtype cohort, leading to the construction of the contingency table on the right. Fisher’s exact test was applied to each drug co-recommendation to assess statistical significance, resulting in the DrugComb analysis table displayed below. The percentage column (percentage) indicates the proportion of drug co-recommendations, with the *P*-value and adjusted *P*-value columns (p.value, adjust\_p.value) reflecting the significance level. The odds ratio (oddsRatio) provides a measure of the effect size in comparison to the control group. The final four columns offer details on the cell line drug recommendation status, including specific cell line IDs (individual\_id) used for wet-lab validation.

**Table 2.** Drugs obtained from the drug knowledge database are classified into clinically relevant categories using a system of six levels of evidence

	Approved	Clinical	Preclinical
Same cancer	A1	A2	A3
Other cancer	B1	B2	B3

cancer types and their clinical strength can be clearly defined, which can guide the clinical decision.

**Analysis tools**

Onko\_DrugCombScreen was implemented using R (v.4.3.1) and R Shiny (v.1.8.0). This Shiny app integrated a variety of R programming language packages for comprehensive bioinformatics analysis. For parsing and generating data structures, we utilized readxl v1.4.1 [20, 21]. To facilitate data manipulation and transformation, we employed packages such as reshape2 v1.4.4 [22], tidyr v1.2.1 [21], and dplyr v1.0.10 [23]. We applied packages such as maftools v2.12.0 [24], clusterProfiler v4.4.4 [25], and VariantAnnotation v1.42.1 [26] for the analysis of somatic variants, functional profiles of genes. For data visualization, we used packages such as circlize v0.4.15 [27] for circular visualizations, ggalluvial v0.12.3 [28, 29] for alluvial diagrams, ggrepel v0.9.2 [30] for label clarity, ComplexHeatmap v2.12.1 [31] for sophisticated heatmaps, and ggplot2 v3.4.0 [32] for creating customizable static plots.

**Data source**

The harmonized drug database, derived from open-source drug knowledge databases, including GDKD [17], CIViC [18], and OncoKB [19], utilizes the DrugBank Vocabulary dataset from DrugBank [33] to standardize drug synonyms. TCGA-

BRCA data and breast cancer cell line data used in the case study were collected from UCSC Xena hubs [34].

**Results**

**Case study: application and validation using TCGA-BRCA data**

**Dataset selection and processing**

In this case study, the Onko\_DrugCombScreen was applied to the TCGA-BRCA dataset to validate its efficacy in identifying effective drug combinations for breast cancer. The TCGA-BRCA dataset, derived from the TCGA Pan-Cancer (PAN-CAN) initiative, was chosen for its comprehensive genetic profiling, including extensive data on copy number variations, single nucleotide variations, and molecular subtype profiles [35]. This dataset provides a broad coverage of genetic variations, making it an ideal resource for this analysis. Additionally, cell line data from the Cancer Cell Line Encyclopedia (Breast) were incorporated to complement the analysis [36]. All of the above datasets are available in UCSC Xena hubs [34].

In the preprocessing phase, somatic mutation data from the PANCAN was converted into a compatible CSV file for analysis by the Onko\_DrugCombScreen. This process involved filtering the dataset to isolate BRCA cancer data and further stratifying it into molecular subtypes: luminal A/B, HER2+, normal-like, and basal-like. In this case study, we used normal-like breast cancer as the comparison cohort, and HER2+ and basal-like subtypes as the target cohorts, respectively, to analyze and validate the efficacy of Onko\_DrugCombScreen. Additionally, by integrating cell line data, Onko\_DrugCombScreen provided guidance on suitable cell lines for subsequent experimental validation.

## Validation and results

The drug co-recommendation comparison analysis revealed significant disparities between the three BRCA subtypes (HER2+ and basal-like) and the normal-like BRCA data. Significant drug co-recommendations extracted from Onko\_DrugCombScreen were compared with combinational therapies in Wang and Minden's review [37], as well as FDA-approved drug combinations, to validate the effectiveness. As [Supplementary Table S1](#) shows, the "adjust\_p.value" and "OR" (odds ratio), obtained from the Fisher's exact test, indicate the significance and magnitude of the drug combination in the target cohort compared to the comparison cohort, and the "Percentage" depicts the proportion of the drug combination recommended in the target cohort. Setting the threshold at  $\text{adjust\_p.value} < .05$ ,  $\text{OR} > 1$ , and  $\text{Percentage} > 50\%$  retains around 30% of the significant candidate drug combinations (30 250/111 987 in HER2+ versus normal-like and 48 348/112 069 in basal-like versus normal-like). Notably, these stringent criteria preserved almost all approved and clinical trial drug combinations, including the approved combinational therapy of pertuzumab + trastuzumab for the HER2+ subtype and pembrolizumab + paclitaxel for triple-negative breast cancer. These results highlight Onko\_DrugCombScreen's accuracy in identifying clinically relevant drug combinations, confirming its effectiveness. Besides, upon comparison with the DrugComb.org database [38], it was found that none of the approved and currently in clinical trial drug combinations of breast cancer had any recorded synergy scores.

The validation analysis demonstrates that the Onko\_DrugCombScreen is adept at identifying established breast cancer drug combinations in the BRCA subtypes such as HER2+ and basal-like when compared to normal-like BRCA subtype. This finding not only validates the tool's effectiveness but also highlights its potential in discovering novel drug combinations for various cancer types. Consequently, the case study accentuates the utility of the Onko\_DrugCombScreen in providing targeted and efficacious drug recommendations.

## Data analysis workflow of Onko\_DrugCombScreen

The data analysis workflow of Onko\_DrugCombScreen is depicted in [Fig. 2](#): Variant data such as single nucleotide variants (SNVs) and copy number variants (CNVs) from both the cancer subtype cohort and the comparison cancer subtype cohort are preprocessed and converted into variant tables compatible with Onko\_DrugCombScreen. These patient variant data are then mapped to public drug databases (CIViC, GDKD, OncoKb) after integration with variant interpretation annotations and drug evidence levels for drug recommendations. The resulting drug recommendations are subjected to statistical analysis, focusing on the statistical differences in drug combination candidates observed between the target cancer subtype group and the comparison cancer subtype group. To identify drug combination candidates that are significantly and frequently recommended in the target group compared to the comparison group, Fisher's exact test is applied. Subsequently, the selected drug combination candidates undergo an integrated analysis with cell line data to identify available cell line samples, facilitating wet-lab validation. Additionally, all analysis results are visualized, making the findings clearer and more intuitive. The integration of these processes is crucial for

confirming drug combination recommendations for the cancer type of interest. The final validation stage may include conducting wet-lab drug screenings to confirm the analysis results and deepen the understanding of the underlying biological mechanisms ([Fig. 2](#)). To assist users in becoming familiar with the analysis workflow of Onko\_DrugCombScreen, [Supplementary File S1](#) is provided as a test dataset for practice and exploration.

## Data preprocessing

SNVs and CNVs are typically stored in formats such as VCF, MAF, TXT, or Excel. A preprocessing step is necessary to convert these various formats into CSV format ([Fig. 2](#)). These data frames are then suitable for use in the knowledge-based drug recommendation analysis within Onko\_DrugCombScreen. To assist users, we have provided an example preprocessing script along with a detailed instruction markdown file for the TCGA-BRCA case study in the GitLab repository under the "test\_data" directory. Users can modify the provided example preprocessing script according to their needs to convert their own variant data into a compatible input format ([Fig. 2](#)).

## Matching rule between variant annotations in patients' data and database

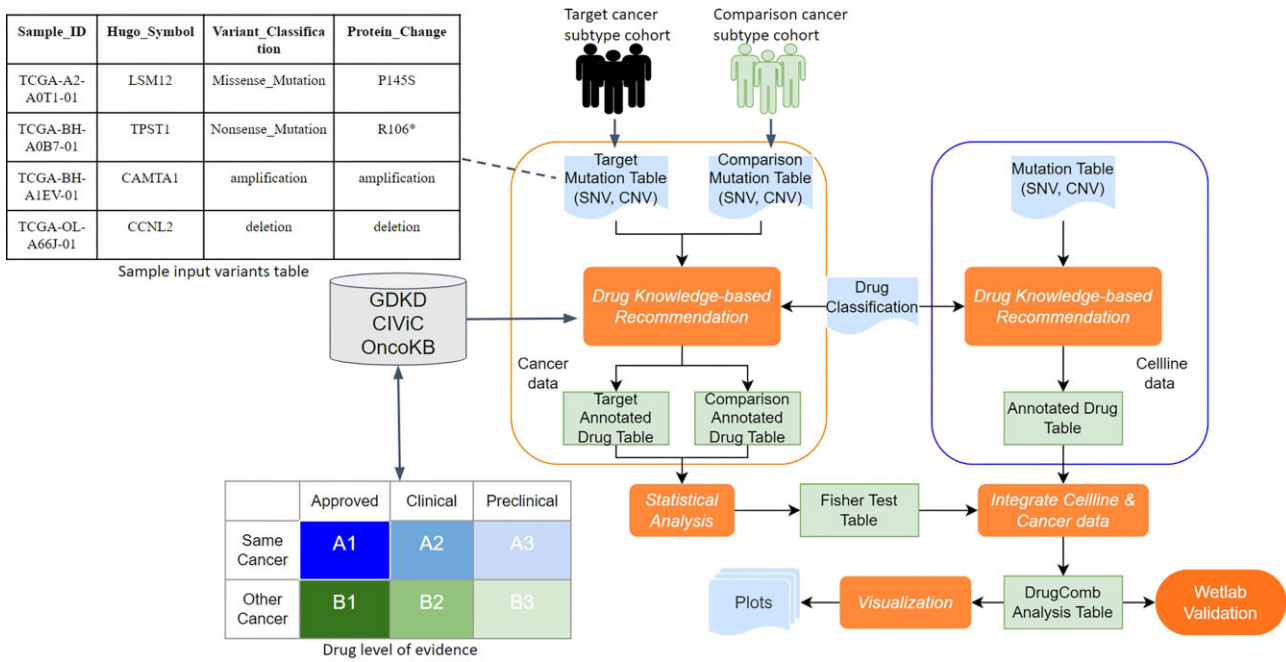
Due to the different annotation descriptions of variants in the three drug databases (GDKD [17], CIViC [18], and OncoKB [19]) and original patients' variant data, we harmonized the three drug databases and designed a matching rule based on the interpretation of biological significance ([Table 3](#)). All variant classes or effects map to the biological interpretations of "loss," "gain," or "mutation." We can then associate the original variants ([Table 4](#)) with the information in the knowledge database based on biological interpretations and obtain the relevant target drug information.

## Drug recommendation annotation

After the matching rule is defined, the drug knowledge-based analysis was performed to export the drug recommendation tables across all target and comparison subtype data. However, due to discrepancies in drug nomenclature across the three drug databases, we employed the "DrugBank Vocabulary" dataset from DrugBank to standardize synonymous drug names. Subsequently, each drug name was annotated to its final drug class. This annotation is stored in the columns "Origin\_Drug\_Name" and "Classified\_Drug\_Name" of the DrugComb analysis table. Additionally, other useful information such as variant type is annotated in the "mutType" column, and variant match status—which indicates whether the amino acid change in the raw data exactly matches the database records or not—is saved in the "Match\_Sign" column.

## Data visualization

Onko\_DrugCombScreen provides a variety of charts for visual analysis results, allowing users to understand data more intuitively ([Fig. 3](#)). The application integrates multiple plotting functions, including volcano plots, heatmaps, Circos plots, alluvial diagrams, UpSet plots, and bar charts. These visualization tools help to easily identify recommended drugs or candidate drug combinations for subsequent wet-lab analysis and validation. Users can configure settings in the left panel of Onko\_DrugCombScreen and customize the resolution for PDF file export.



**Figure 2.** The workflow for Onko\_DrugCombScreen drug combination data analysis. After the recommendation process based on drug knowledge, SNVs and CNVs are merged into an annotated drug table. Following statistical analysis and integration of cell line data, the final DrugComb analysis table will be used for visualization and wet-lab validation.

**Table 3.** Biological interpretation of variant annotations in drug databases

Database	Variant	Interpretation
GDKD/CIViC/OncoKB	“splice”	Loss
GDKD/CIViC/OncoKB	“delins”	Loss
GDKD/CIViC/OncoKB	“ins”	Insertion
GDKD/CIViC	“del”	Deletion
GDKD	“indel”	Loss
GDKD/CIViC	“fs”	Loss
GDKD/CIViC/OncoKB	“deletion”	Loss
GDKD/CIViC/OncoKB	“amplification”	Gain
GDKD	mut	Mutation
GDKD	any	Mutation
CIViC	loss/loss-of-function	Loss
CIViC	“mutation”	Mutation
CIViC	“^expression”	Gain
CIViC	“Overexpression”	Gain
CIViC	“Underexpression”	Loss
OncoKB	Truncating mutations	Loss
OncoKB	Oncogenic mutations	Mutation
CIViC	“FRAMESHIFT”	Loss
CIViC	“FRAME SHIFT”	Loss
OncoKB/CIViC	Exon 17 mutations	Mutation (exact match)
CIViC	Exon 19 deletion	Loss (exact match)
CIViC	Exon 14 skipping mutation	Mutation (exact match)

**Discussion**

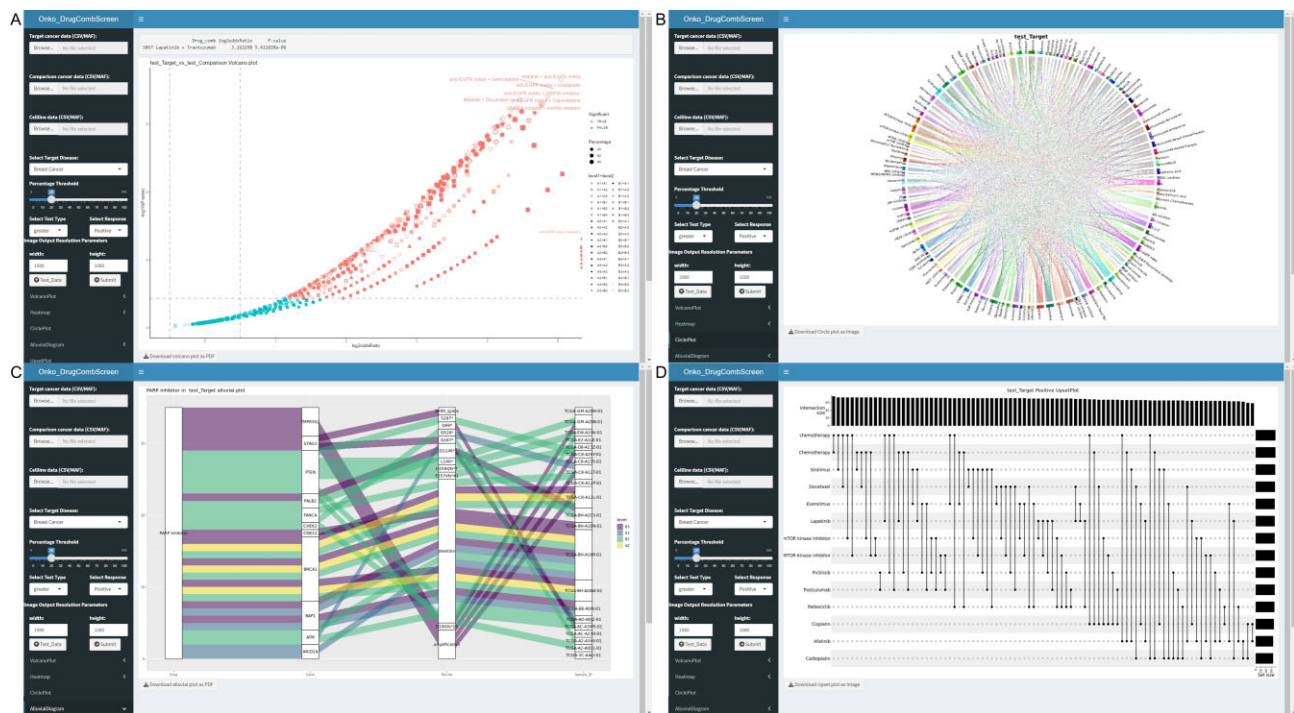
Drug combinations are widely recognized for their benefits in cancer therapy. Here, we developed the Onko\_DrugCombScreen Shiny app integrated statistical analysis to identify the most significant candidate drug combinations for a target tumor type cohort. We utilized drug

**Table 4.** Biological interpretation of variant annotations in patients’ variants

Variant	Interpretation
In_Frame_Ins	Insertion (exact match)
In_Frame_Del	Deletion (exact match)
Frame_Shift_Ins	Loss
Frame_Shift_Del	Loss
Splice_site	Loss
amplification	Gain
deletion	Loss
Missense_Mutation	Mutation
Nonsense_Mutation	Loss
Nonstop_Mutation	Exact match
Translation_Start_Site	Exact match

knowledgebase recommendations derived from mutation data of the targeted cancer patient cohort and the comparison cohort to identify drug co-recommendations. This is complemented by integrating cell line data to assist in the validation of biological experiments. Onko\_DrugCombScreen’s ability to identify effective drug combinations, as demonstrated in the TCGA-BRCA case study, suggests a promising way toward more tumor type tailored and effective cancer combination therapy.

In contrast to current computational methods that mainly focus on synergy and dose-response matrices, Onko\_DrugCombScreen is a drug knowledge-based analysis approach. It not only provides therapeutic recommendations but also offers guidance for clinical research, thereby integrating more closely with clinical applications. Moreover, all drug recommendations can be traced back to patient genetics and variants through the visual alluvial diagram of Onko\_DrugCombScreen. Utilizing the drug database that is also employed by the MTB-Report [6], each recommended drug’s level of evidence and response are explicitly defined.



**Figure 3.** Visualization of the Onko\_DrugCombScreen. **(A)** Volcano plot identifying significant drug combinations. **(B)** Circos plot depicting the most proportional drug co-recommendations. **(C)** Alluvial diagram tracing mutations back to recommended single drugs. **(D)** UpSet plot showing the top-recommended single drugs and their intersections.

This clarity effectively aids in addressing issues of selectivity in the recommended drug combinations, issues that are often overlooked in previous methods. Moreover, the utilization of drug databases for the recommendation of candidate drug combinations based on patient gene mutation profiles can potentially reduce the effort required for toxicity analysis [39, 40]. These databases provide valuable information on the relationships between individual drugs and specific gene mutations or molecular targets, which can guide the selection of drug combinations with potentially favorable efficacy profiles. Furthermore, the drugs included in these databases are often approved or under clinical trials, meaning that their toxicity profiles have been extensively studied and characterized [41, 42]. These existing safety data can serve as a foundation for assessing the toxicity of drug combinations, as it provides insights into the common adverse events, dose-limiting toxicities, and recommended dosing schedules of the individual drugs. By leveraging this information, researchers can streamline the toxicity assessment process and make more informed decisions when designing drug combination studies. However, the toxicity of a drug combination is not simply the sum of individual drug toxicities. Comprehensive safety assessments are still necessary, considering factors such as drug–drug interactions, dosing, scheduling, and specific patient populations.

Onko\_DrugCombScreen also offers flexibility for application across various cancer types beyond breast cancer, provided there is sufficient sample size and well-defined molecular stratification within the target and comparison cohorts. For example, in the TCGA-BRCA case study, the target cohorts included 280 patients with the basal-like subtype and 82 patients with the HER2+ subtype, while the comparison cohort (normal-like subtype) consisted of 143 patients. These cohort sizes provide adequate statistical power for

Fisher’s exact test [43, 44], which remains robust even with smaller sample sizes, such as when the sample size is below 5. However, when any expected cell counts in a contingency table are very low, statistical interpretation may become limited. Therefore, we recommend caution when analyzing smaller cohorts for drug combination candidate identification, as reduced statistical robustness can affect result interpretation. The tool’s ability to analyze stratified cohorts makes it versatile across different cancer types, but its success depends on sufficient sample sizes and well-defined molecular subtypes.

Looking forward, the potential for further development of Onko\_DrugCombScreen is substantial. While the current version does not yet incorporate detailed drug synergy, dose–response, or combination toxicity analyses, we recognize these as crucial factors for clinical implementation of combination therapies. Future iterations of Onko\_DrugCombScreen could integrate drug synergy data from sources such as DrugComb [45, 46], where available, and refine recommendations using dose–response metrics, particularly during validation in cell line or patient-derived model systems. Additionally, toxicity analysis can be streamlined by leveraging existing toxicity profiles from drug knowledgebases, though a more comprehensive assessment of combination-specific toxicity will be necessary. To address these gaps, we propose a possible solution that incorporates artificial intelligence (AI) and ML techniques to enhance the tool’s capabilities. Specifically, ML models such as graph convolutional neural networks [47], random forest, or boosted models could analyze large-scale patient variant data and drug interactions to more accurately predict combination efficacy and minimize adverse effects. By incorporating these AI-driven approaches, Onko\_DrugCombScreen will evolve into a more robust and clinically relevant tool,

capable of offering precision-guided, synergistic drug combinations tailored to individual patients.

## Conclusion

In conclusion, the Onko\_DrugCombScreen Shiny app represents an innovative tool in the field of precision cancer therapy, offering a novel drug knowledge-based approach to drug combination screening. This application leverages drug knowledge database analysis along with advanced statistical and visualization techniques to identify effective drug combinations. It effectively utilizes drug recommendations from targeted cancer cohort and comparison cohort, combined with cell line data, to provide prominent drug co-recommendations for targeted cancer type. Validated through a TCGA-BRCA case study, the application has demonstrated its potential in accurately identifying both existing and novel drug combinations, aligning with the evolving field of precision oncology. Looking ahead, the integration of AI and ML technologies holds the promise of further enhancing its predictive capabilities, making it a valuable tool in the quest for more targeted and effective cancer treatment approaches.

## Acknowledgements

The authors would like to acknowledge the Volkswagen Foundation's support in MTB-Report project. J.Y. was supported by the Ph.D. program "Genome Science"—International Max Planck Research School, part of the Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences. J.D. and T.B. are members of the Göttingen Campus Institute Data Science.

*Author contributions:* J.Y., M.W., B.C., J.D., and T.B. designed the study. J.Y. and B.C. provided major contributions to the computational method and visualization. J.Y. designed, developed, and implemented the Shiny app. M.W. tested the Shiny app and helped identify problems and bugs. J.Y. wrote the manuscript. All authors critically reviewed the content and approved the final manuscript.

## Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

## Conflict of interest

None declared.

## Funding

This project was supported by the Volkswagen Foundation within research project MTB-Report (ZN3424).

## Data availability

The Onko\_DrugCombScreen source code is available on the GitHub repository at [https://gitlab.gwdg.de/MedBioinf/mtb/onko\\_drugcombscreen](https://gitlab.gwdg.de/MedBioinf/mtb/onko_drugcombscreen), and also on Zenodo at <https://doi.org/10.5281/zenodo.14614900>. Test datasets are provided in Supplementary File S1. The application is accessible at [https://rshiny.gwdg.de/apps/onko\\_drugcombscreen/](https://rshiny.gwdg.de/apps/onko_drugcombscreen/).

## References

- Duarte D, Vale N. Evaluation of synergism in drug combinations and reference models for future orientations in oncology. *Curr Res Pharmacol Drug Discov* 2022;3:100110. <https://doi.org/10.1016/j.crphar.2022.100110>
- Delou JM, Souza AS, Souza LC *et al.* Highlights in resistance mechanism pathways for combination therapy. *Cells* 2019;8:1013. <https://doi.org/10.3390/cells8091013>
- Kong W, Midena G, Chen Y *et al.* Systematic review of computational methods for drug combination prediction. *Comput Struct Biotechnol J* 2022;20:2807–14. <https://doi.org/10.1016/j.csbj.2022.05.055>
- Narayan RS, Molenaar P, Teng J *et al.* A cancer drug atlas enables synergistic targeting of independent drug vulnerabilities. *Nat Commun* 2020;11:2935. <https://doi.org/10.1038/s41467-020-16735-2>
- Sarmah D, Meredith WO, Weber IK *et al.* Predicting anti-cancer drug combination responses with a temporal cell state network model. *PLoS Comput Biol* 2023;19:e1011082. <https://doi.org/10.1371/journal.pcbi.1011082>
- Perera-Bel J, Hutter B, Heining C *et al.* From somatic variants towards precision oncology: evidence-driven reporting of treatment options in molecular tumor boards. *Genome Med* 2018;10:18. <https://doi.org/10.1186/s13073-018-0529-2>
- Luchini C, Lawlor RT, Milella M *et al.* Molecular tumor boards in clinical practice. *Trends Cancer* 2020;6:738–44. <https://doi.org/10.1016/j.trecan.2020.05.008>
- Kurz NS, Perera-Bel J, Höltermann C *et al.* Identifying actionable variants in cancer—the dual web and batch processing tool MTB-Report. *Stud Health Technol Inform* 2022;296:73–80.
- Preuer K, Lewis RP, Hochreiter S *et al.* DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018;34:1538–46. <https://doi.org/10.1093/bioinformatics/btx806>
- Wang J, Liu X, Shen S *et al.* DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform* 2022;23:bbab390. <https://doi.org/10.1093/bib/bbab390>
- Cheng F, Kovács IA, Barabási AL. Network-based prediction of drug combinations. *Nat Commun* 2019;10:1197. <https://doi.org/10.1038/s41467-019-09186-x>
- Li H, Zou L, Kowah JA *et al.* Predicting drug synergy and discovering new drug combinations based on a graph autoencoder and convolutional neural network. *Interdiscip Sci* 2023;15:316–30. <https://doi.org/10.1007/s12539-023-00558-y>
- Gayvert KM, Aly O, Platt J *et al.* A computational approach for identifying synergistic drug combinations. *PLoS Comput Biol* 2017;13:e1005308. <https://doi.org/10.1371/journal.pcbi.1005308>
- Janizek JD, Celik S, Lee SI. Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine. bioRxiv, <https://doi.org/10.1101/331769>, 27 May 2018, preprint: not peer reviewed.
- Cui W, Aouidate A, Wang S *et al.* Discovering anti-cancer drugs via computational methods. *Front Pharmacol* 2020;11:733. <https://doi.org/10.3389/fphar.2020.00733>
- Tan AC, Bagley SJ, Wen PY *et al.* Systematic review of combinations of targeted or immunotherapy in advanced solid tumors. *J Immunother Cancer* 2021;9:e002459.
- Dienstmann R, Jang IS, Bot B *et al.* Database of genomic biomarkers for cancer drugs and clinical targetability in solid tumors. *Cancer Discov* 2015;5:118–23. <https://doi.org/10.1158/2159-8290.CD-14-1118>
- Griffith M, Spies NC, Krysiak K *et al.* CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* 2017;49:170–4. <https://doi.org/10.1038/ng.3774>
- Chakravarty D, Gao J, Phillips S *et al.* OncoKB: a precision oncology knowledge base. *JCO Precis Oncol* 2017;1:PO.17.00011.

20. Wickham H, Bryan J. readxl: read Excel files. 2023. <https://readxl.tidyverse.org>, <https://github.com/tidyverse/readxl>
21. Wickham H, Averick M, Bryan J *et al.* Welcome to the Tidyverse. *J Open Source Softw* 2019;4:1686. <https://doi.org/10.21105/joss.01686>
22. Wickham H. Reshaping data with the reshape package. *J Stat Softw* 2007;21:1–20. <https://doi.org/10.18637/jss.v021.i12>
23. Wickham H, François R, Henry L *et al.* dplyr: a grammar of data manipulation. 2023. <https://dplyr.tidyverse.org/>. (15 July 2024, date last accessed).
24. Mayakonda A, Lin DC, Assenov Y *et al.* Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 2018;28:1747–56. <https://doi.org/10.1101/gr.239244.118>
25. Yu G, Wang LG, Han Y *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284–7. <https://doi.org/10.1089/omi.2011.0118>
26. Obenchain V, Lawrence M, Carey V *et al.* VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics* 2014;30:2076–8. <https://doi.org/10.1093/bioinformatics/btu168>
27. Gu Z, Gu L, Eils R *et al.* ‘Circlize’ implements and enhances circular visualization in R. 2014. <https://doi.org/10.1093/bioinformatics/btu393>. (15 July 2024, date last accessed).
28. Brunson JC, Read QD. ggalluvial: alluvial plots in ‘ggplot2’. R package version 0.12.5. 2023. <https://github.com/tidyverse/dplyr>, <https://dplyr.tidyverse.org>. Date accessed 15 July 2024.
29. Brunson JC. ggalluvial: layered grammar for alluvial plots. *J Open Source Softw* 2020;5:2017. <https://doi.org/10.21105/joss.02017>
30. Slowikowski K. ggrepel: automatically position non-overlapping text labels with ‘ggplot2’. 2024. <https://ggrepel.slowkow.com/>, <https://github.com/slowkow/ggrepel>. (15 July 2024, date last accessed).
31. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;32:2847–9. <https://doi.org/10.1093/bioinformatics/btw313>
32. Villanueva RAM, Chen ZJ. ggplot2: elegant graphics for data analysis. 2019;17:160–167. <https://doi.org/10.1080/15366367.2019.1565254>. (15 July 2024, date last accessed).
33. Wishart DS, Feunang YD, Guo AC *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;46:D1074–82. <https://doi.org/10.1093/nar/gkx1037>
34. Goldman MJ, Craft B, Hastie M *et al.* Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* 2020;38:675–8. <https://doi.org/10.1038/s41587-020-0546-8>
35. Weinstein JN, Collisson EA, Mills GB *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013;45:1113–20. <https://doi.org/10.1038/ng.2764>
36. Barretina J, Caponigro G, Stransky N *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;483:603–7. <https://doi.org/10.1038/nature11003>
37. Wang Y, Minden A. Current molecular combination therapies used for the treatment of breast cancer. *Int J Mol Sci* 2022;23:11046. <https://doi.org/10.3390/ijms231911046>
38. Zheng S, Aldahdooh J, Shadbahr T *et al.* DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal. *Nucleic Acids Res* 2021;49:W174–84. <https://doi.org/10.1093/nar/gkab438>
39. Vo AH, Van Vleet TR, Gupta RR *et al.* An overview of machine learning and big data for drug toxicity evaluation. *Chem Res Toxicol* 2019;33:20–37. <https://doi.org/10.1021/acs.chemrestox.9b00227>
40. Galletti C, Bota PM, Oliva B *et al.* Mining drug–target and drug–adverse drug reaction databases to identify target–adverse drug reaction relationships. *Database* 2021;2021:baab068. <https://doi.org/10.1093/database/baab068>
41. Guengerich FP. Mechanisms of drug toxicity and relevance to pharmaceutical development. *Drug Metab Pharmacokinet* 2011;26:3–14.
42. Toropov AA, Toropova AP, Raska I Jr *et al.* Comprehension of drug toxicity: software and databases. *Comput Biol Med* 2014;45:20–5.
43. Finney DJ. The Fisher–Yates test of significance in 2×2 contingency tables. *Biometrika* 1948;35:145–56.
44. Kim HY. Statistical notes for clinical researchers: chi-squared test and Fisher’s exact test. *Restor Dent Endod* 2017;42:152–5.
45. Zagidullin B, Aldahdooh J, Zheng S *et al.* DrugComb: an integrative cancer drug combination data portal. *Nucleic Acids Res* 2019;47:W43–51.
46. Liu H, Zhang W, Zou B *et al.* DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res* 2020;48:D871–81.
47. Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. arXiv, <https://arxiv.org/abs/1606.09375>, 30 June 2016, preprint: not peer reviewed.