Extending the Boundaries of Cancer Therapeutic Complexity with Literature Data Mining

Authors names

Danna Niezni[1], Hillel Taub-Tabib[2], Yuval Harris[1], Hagit Sason-Bauer[1], Yakir Amrusi[1], Dana Azagury[1], Maytal Avrashami[1], Shaked Launer-Wachs[1], Jon Borchardt[3], M Kusold[3], Aryeh Tiktinsky[2], Tom Hope[2,5], Yoav Goldberg[2,4], *Yosi Shamay[1]

Author Affiliation

[1] Faculty of Biomedical Engineering, Technion – Israel Institute of Technology, Haifa, Israel

[2]Allen Institute for AI, Tel Aviv, Israel

[3]Allen Institute for AI, Seattle, USA

[4]Bar-Ilan University, Ramat-Gan, Israel

[5]The Hebrew University, Jerusalem, Israel

Contact Information

Dr. Yosi Shamay

Cancer Nanomedicine and Nanoinformatics Lab,

Biomedical Engineering Faculty, Technion – Israel Institute of Technology, Haifa, Israel

Tel +972-(0)4-829-1203

Email: yshamay@technion.ac.il

Abstract

Even though there is high motivation to develop effective monotherapies for cancer, combination drug therapy is still one of the main pillars of oncological treatments. The formation of an effective combinatorial standard of care (SOC) can take many years and its length of development is increasing with complexity of treatment. In this paper, we develop a path to extend the boundaries of complexity in combinatorial cancer treatments using text data mining (TDM). We first used TDM to characterize the current boundaries of cancer treatment complexity. We found that the current complexity limit for clinical trials is 6 drugs per plan and for pre-clinical research is 10. We then developed a TDM based assistive technology, cancer plan builder (CPB), that allows experts to create high complexity combination treatment (HCCT) plans of significantly larger size. We evaluated CPB by comparing it to the standard of care, to fully automated plans and to HCCT plans created by experts without the tool. We found that by average, 86% of elements in the plans generated with CPB are either compatible with standard of care or FDA approved for the correct indication. In addition, CPB assisted plans were superior in every metric to plans generated either automatically or manually. To conclude, we propose a new tool and workflow to handle the vast complexity space of combinatorial cancer treatment facilitate further evaluation of HCCT in experimental cancer research.

Introduction

Combination Therapy in Cancer

The number of available cancer treatments is increasing in size and in diversity, involving a wide variety of modalities such as surgery, chemotherapy, targeted therapy, cell therapy, radiation therapy, immunotherapy, etc. Oncology relies heavily on pharmaceutical intervention and drug combinations are used in most cancers, even though there is a strong incentive to develop monotherapies as a substitute. The most ideal example of such successful transition is the approval of imatinib monotherapy for chronic myeloid leukemia (CML) and gastrointestinal stromal tumors (GIST) over chemotherapy cocktails[1]. On the other hand, for some cancers it seems that the current optimal solution is adding more drugs to the standard of care (SOC) instead of replacing them. This approach is well demonstrated in the case of Diffuse Large B-cell Lymphoma (DLBCL) [2]. The history of the SOC in DLBCL can teach us a lot on how complex treatments are formed. After testing multiple chemotherapeutic drugs from the 1980s until early 1990s, a four-drug regiment named CHOP (Cyclophosphamide, Hydroxydaunorubicin, vincristine and Prednisone or Prednisolone) was selected as the SOC. in the late 1990s, Rituximab, a CD20 antibody was tested as monotherapy in DLBCL with the aim to replace CHOP. Even though Rituximab had an impressive activity, it was not superior to CHOP, and eventually was tested in combination with CHOP [3]. The 5-drug combo RCHOP was approved in 1999 and in 2015 a kinase inhibitor, ibrutinib was tested as a monotherapy replacement [4]. Again, even though it showed activity it was not superior to the SOC and was tested in combination to form a 6-drug regimen– RCHOP-I. The fact that it takes 40 years to optimize this 6-drug regimen as SOC indicates that there is an unmet need to accelerate discoveries of complex drug regiments and treatment plans. We propose to exploit the scientific literature generated by decades of trials and errors. One can imagine the following theoretical thought experiment- If a clinician, for example an oncologist, could read, remember and understand all the papers published in medical history, would she/he propose a more complex treatment plan? what is the resolution and detail such a plan can contain? and what would be the burden of proof for the new plan?

There are numerous reports in the literature on anecdotal optimization details for example: i) some RTK (Receptor Tyrosine Kinase) inhibitors work better when given at night [5,6]. ii) if anti-vascular therapy is given exactly one hour before chemotherapy a surprising synergy and efficacy emerges[7], iii) focused ultrasound improves penetration of nanoparticles into brain tumors[8,9] There are countless examples as such, representing the high amount of details that can be incorporated in future cancer treatment plans.

Text Data Mining (TDM) for High Complexity Combination Therapy (HCCT).

The resolution of details and the complexity of treatment protocols are not only limited by what has been tested and published, but also by the limit of knowledge that human scientists and clinician can process. Technological developments in text and data mining (TDM) have opened up new possibilities for researchers, enabling them to analyze textual information in ways that were not previously feasible. In biomedicine, automated literature search, natural language processing (NLP) and TDM were used successfully to identify relationships between drugs, diseases, proteins and genes, saving both time and effort for researchers. We would like to propose a path towards TDM based navigation, organization and planning within the grand space of possible combinatorial treatments and exploit cancer complexity into what we call high complexity combination therapy (HCCT).

The scientific literature can be exploited in multiple dimensions of the complex plan: if a drug matches a relevant mutation, what drugs can synergize with it? Which formulation can enhance the performance drug combination? What other treatment (e.g., focused ultrasound or radiation) can increase the distribution of the drug formulations? What is the temporal sequence to administer that combination? Is there a diet that improve the treatment or a sport?

These types of questions, if expanded and repeated while answered properly, could yield a highly complex *but reasonable* medical hypothesis, which translates to a treatment plan in the form of HCCT. In this study, we first characterize the current and possible landscape of combination therapy in cancer and unravel the boundaries of treatment complexity. We then hypothesized that treatment elements can be easily assembled into a complex plan with a dedicated cancer plan builder (CPB) tool which provides literature-based evidence the user assembling the plan. We then present new approaches to generate personalized HCCT plans for different cancer types with defined genetic and protein findings. We show that CPB facilitates the generation of HCCT beyond the current complexity boundaries. We then compare the CPB approach to manual plan generation via crowdsourcing tasks. Finally, we developed scores and metrics to facilitates the comparison, ranking and prioritization of the different HCCTs.

Results

Characterizing the current complexity of cancer treatments.

Before exploring the generation of complex treatment plans, we sought to characterize and analyze the status of drug combination therapy in research and clinical trials. We used two different methods to capture the number of pharmaceutical interventions applied in different cancers to quantitively

characterize the level of complexity in combination therapy (**Figure 1a**). We mined drug combinations in research from PubMed abstracts and clinicaltrials.gov for clinical stage combinations. Our text mining approach used both explicit and implicit drug combinations. In the explicit search, specific drug names were mentioned in the described treatment (e.g. "a combination of dolutegravir and lamivudine was administered …"), while in the implicit searches, the number of drugs in a combination is mentioned without specifying the drug names (e.g. "the 2 drug regimen proved effective …"). To accurately capture the syntax of implicit combinations, we used the SPIKE engine for syntactic structure search[10] and ranked the frequency of implicit drug combination phrases (**Supplementary Fig S1**). The most frequent phrasings were 'X-drug combination' and 'X-drug regimen'. When the two approaches were plotted, it was clearly evident that there is a logarithmic decline in frequency (string negative correlation) of each additional drug to the combination, ending at 6 drugs when the drugs are mentioned explicitly by name (**Fig 1b**). Additionally, we found that when drugs are explicit, the number of drugs in the combination drops down in the exact same manner in both research and clinical trials **(Figure 1b**). But when the drugs are implicit, the decline from two to four drugs is moderate. This can be explained by PubMed abstract's word limit, where describing a 4-6 drug combination by the drug names will use valuable space. Furthermore, there are no more than six drugs mentioned explicitly in any combination, while when the drugs are implicit, combinations of up to 10 drugs arise from the search. Additionally, we can learn that combinations in clinical trials are mostly studied with 2-3 drugs while combinations of 4 drugs represents less than 10% and 5 drugs less than 1% of clinical trials. The feasible limit of drug cocktails has only a handful of clinical trials with 6 drugs per treatment. Though it is possible that there are clinical trials with more than 6 drugs that were missed due to the explicit search, it is not likely to post a clinical trial without specifying the drugs especially in clinical trials. An additional search for explicit drug combinations in any disease type revealed that there are >50 fold more publications in the 6-drug combination category mainly to anti-HIV cocktails related research as well as antibiotic cocktails.

Next, we wanted to quantify the combination space for different sizes of treatment plans and understand what effect it has on the number of possible plans (**Fig 1c**). The number of possible combinations was calculated according to **eq. 1**, this is the equation for calculating the number of possible permutations to organize 'n' drugs in groups of 'r' with no repetitions in the same combination. For example, there are around 100 possible 2-drug combinations for a cancer type with 20 optional drugs to choose from, but for a cancer with 40 available drugs the number of possible 2-drug plans is almost 1000. We used SPIKE to quantify the number of drugs studied in each cancer by capturing co-occurrence at the sentence level. Since different cancer types are studied in a different extent, the combinatorial space for each cancer type is different. Thus, the number of possible plans

might be limited for less studied cancers such as uveal melanoma compared to the highly investigated lung cancer (**Figure 1d**).

Eq. 1- the number of possible 'r-drug' combinations for a cancer type with 'n' studied drugs. $\frac{n!}{r!(n-r)!}$

We then evaluated the distribution of complexity in different cancer types in the scientific literature. To this end, we captured the number of implicit combinations (such as 'five drug regimens') in PubMed and sorted them by cancer type (**Fig 1e**). Most cancer types had combination of up to four drugs per treatment, and it was quite uncommon to see six or more drugs combined. Lung cancer, breast cancer, leukemia and colorectal cancer don't have treatments that combine more than five drugs. There might be some bias against combinations of two drugs as many studies use the explicit drug names. This can be seen in some cancers like lymphoma and melanoma where the four-drug combination is more common than combinations of two or three drugs. The upper limit for drug combination, was identified in glioblastoma (GBM) with a combination of nine drugs. This shows that five or six limit in clinical trials is being pushed in pre-clinical research. We can speculate that as GBM has poor prognosis, multiple drug combinations are more likely to be tested than cancers that reached a good prognosis with the SOC. As was mentioned in the introduction, we can see 6-drug combos in clinical trials for DLBCL which took over 3 decades to develop.

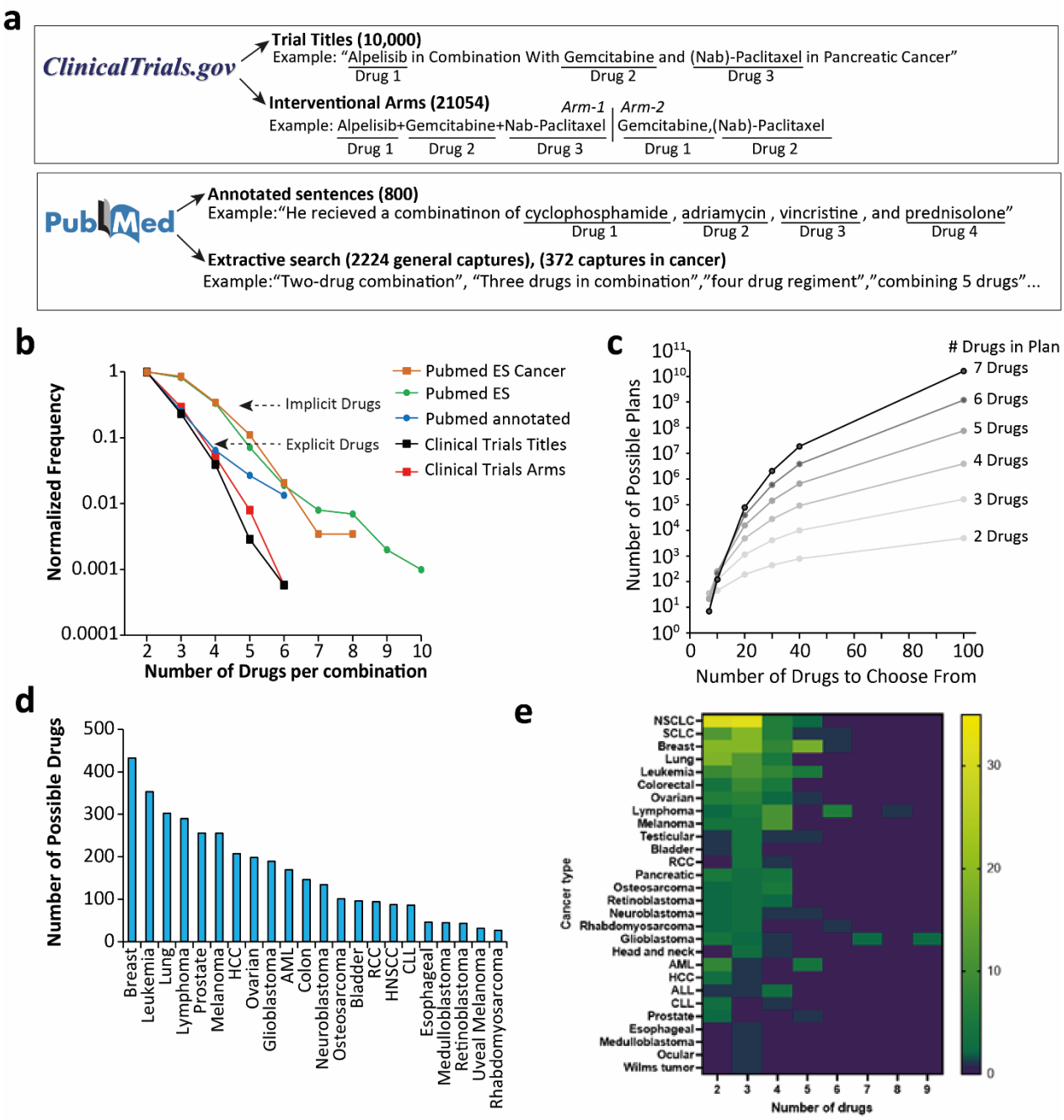


Figure 1. **The current and possible landscape of combination therapy in cancer. a)** approaches to characterize combination therapy from clinicaltrials.gov (top) or PubMed (bottom). **b)** Normalized frequency of drug combinations in treatments for cancers derived from clinialtrial.gov, PubMed and manual annotations. **c)** Number of possible plans according to the number of possible drugs in plan and drugs to choose from. **d)** Number of possible drugs per cancer type, extracted from PubMed abstracts via SPIKE. **e)** Frequency of implicit drug combinations sorted by cancer type, extracted from PubMed via SPIKE.

Cancer Plan Builder tool -CPB

When approaching complex treatment planning, we hypothesize that a simple workflow for generating plans can be derived from the following assumptions:

1) Drugs can be ranked and suggested to the user from the scientific literature based on their co-occurrence score with disease name, associated genes/proteins and additional keywords.

2) The user can choose the most relevant drugs based on both the provided score and the highlighted evidence from the literature.

3) Addition of drugs to the plan can be made in a stepwise manner based on drug synergy evidence between all the drugs in the plan and not necessarily based on disease relevance.

With these assumptions in mind, planners could in principle employ a search tool like PubMed to search for papers with mesh headings relating to the disease conditions and to therapeutic agents. Planners could then screen the results to create a short list of first line treatments for the cancer, they can then run additional PubMed queries to obtain papers which reference these agents, and screen them for additional agents which synergistically combine with the existing ones. This process can then proceed in an iterative fashion, continuously adding treatments which are synergistic to existing ones (**Fig 2a**).

While this workflow is feasible, the need to screen the returned abstracts at each stage of the process does not allow treatment plans to scale beyond a handful of treatments. For example, an initial PubMed search for "(Melanoma[MeSH Terms]) AND (therapeutic uses[MeSH Terms])" returns 14459 results. Thus, the screening required to establish a comprehensive list of promising first line treatments may already be prohibitively time consuming, even without considering additional searches which could identify synergistic treatments.

To address these limitations, we designed and implemented a dedicated cancer plan builder application (CPB), which streamlines the above workflow and dramatically reduces the screening time. Treatment planning with the app proceeds as follows: (**Fig 2b**)- (i) the planner inputs the cancer type, the patient genetic history and biopsy findings, and optionally other keywords of value (e.g., metastasis if the cancer is at that advanced stage, etc.) (ii) the app performs an initial search and displays a list of treatments ranked by their relevance to the input conditions. Each treatment is associated with evidence from the literature showing its relation to the input conditions. (iii) the planner reviews the treatment and evidence and moves one or more treatments into a planning canvas showing the current plan (iv)

the planner can click any treatment on the canvas to trigger a search for treatments which combine with it synergistically (v) as before, the planner sees the evidence for the pairwise synergy before deciding whether to add a synergistic treatment to the plan (vi) the planner can connect an added treatment to other treatments in the plan and inspect evidence for potential synergies (or adverse reactions) further increasing the cohesiveness of the plan. The process continues in an iterative fashion where the planner can experiment with adding and removing treatments based on the evidence.

Note that in contrast to using a standard search engine like PubMed, planners don't need to screen large lists of papers in order to identify candidate treatments. The app automatically extracts treatments and ranks them by the strength of their association to the input conditions (for first line treatments) or to other treatments which are already in the plan (for additional treatments. See the Methods section for the details of this process). Planners still need to validate short evidence snippets associated with the treatments, but as we'll demonstrate, this process is an order of magnitude faster than traditional screening and allows plans to scale to a significantly larger number of components while staying reasonable and consistent.

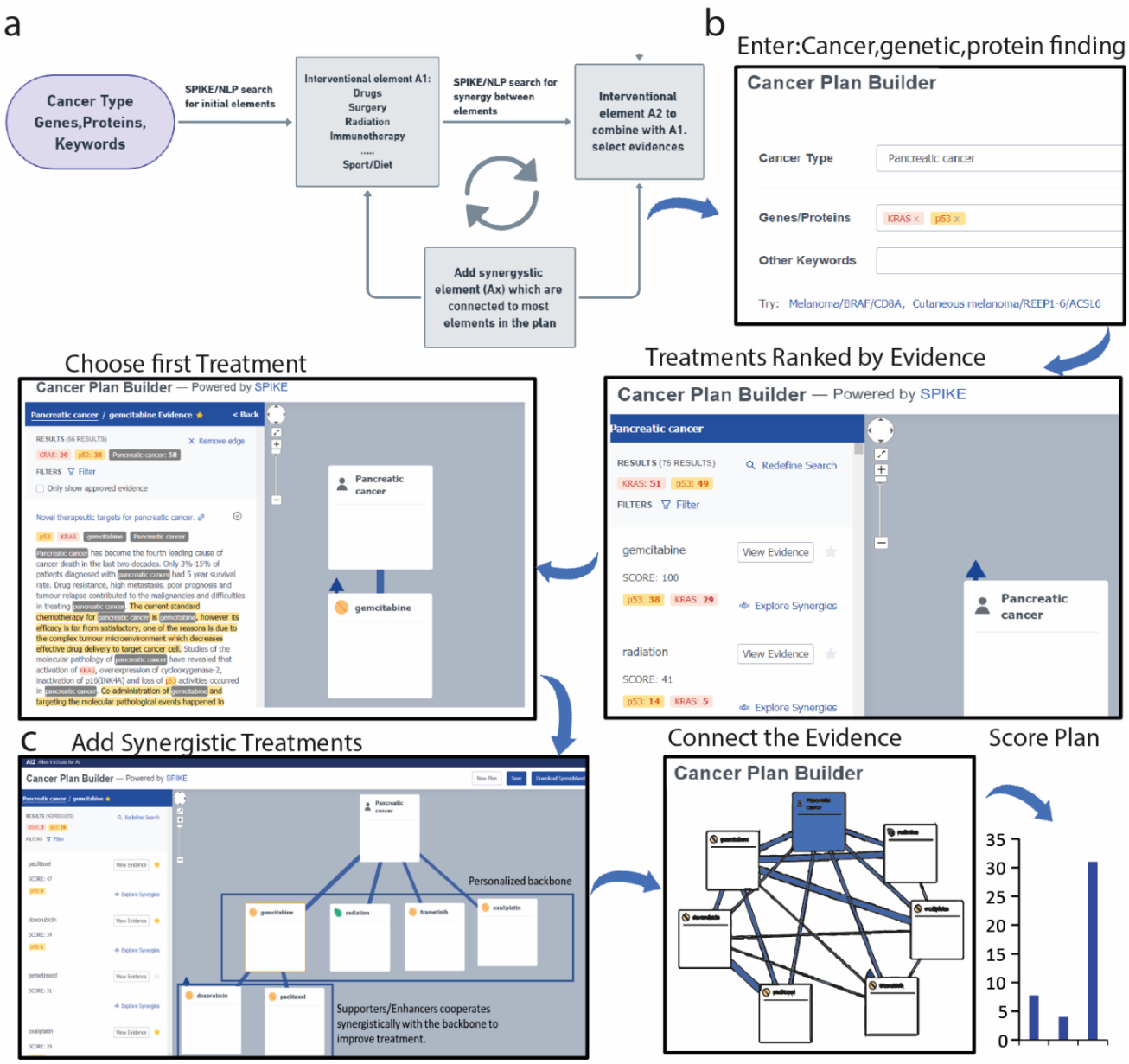


Figure 2. Workflow of building a plan with CPB. a) thought process and fallback strategy. **b)** an example of the process and its result, the user built the personalized backbone treatment with drugs that are directly associated with pancreatic cancer, added the supporting drugs and finally, found the interconnectivity and exported the score of the plan.

Evaluating against the Standard of Care

To evaluate whether the app produces reasonable treatment options, we compared the ranked drug suggestion to the standard of care (first line therapy) as well as number of FDA approvals beyond the standard of care such as second- and third-line therapy. Since the app ranks 100 drugs per case we looked at the top 10 drugs (10% of total suggestions). In each evaluation, we divided cases into 6

general cancer types (with no actionable genetic information), and 6 'personalized type' which include the cancer name together with a known targetable mutation (hallmark for personalized medicine in cancer). We found that for vast majority of personalized and general cancer types, the entire possible standard of care appeared in the top 10 suggestions (**Fig 3a**). Furthermore, we asked how many drugs in the top 10 suggestions are FDA approved for the same indication and found that indeed, more than 75% of top 10 suggested drugs approved for the relevant cancer type (**Fig 3b**). Interestingly, for liver cancer, only 25% of drugs were FDA approved, which can be explained by the lack of approved options for hepatocellular carcinoma, the most common cancer of the liver[11]. When comparing personalized and general cancer types, there are more FDA approved drug in the personalized cases (82.5% vs. 74.2%, respectively), while there is no significant difference between the two in the standard of care (80% for both) This data supports our assumption that including case related keywords, can increase the quality of drug suggestions which facilitates generation of reasonable plans.

Evaluating against Fully Automated Plans

To further evaluate the plans generated with CPB, we defined several relevant parameters that could be comparable between different plans. We consider the level of confidence P, as the amount of published data between the drug and disease or drug to drug. This can also be regarded as the weights of the graph. The number of elements in the plan, N, can be regarded as the plan complexity index. The inter-connectivity of the plan, R, can be regarded as the level of inner-synergism within the drugs. A strong plan is likely to have many interconnections between the different nodes (interventions in the plan). This will probably result in high R/N which, when multiplied by the amount of information (P) would lead to the total score reflecting the importance of these variables. Thus, our score is defined as (R/N)*P or (weights/nodes) *edges (**Table 1)**. The weakness of this scoring system are difficulties in comparing between different cancer types and newly approved drugs, regardless of their advantages, which are less studied and will lead to lower scores due to lack of evidence. But nevertheless, within the same cancer type it can serve as a good benchmark to choose between possible plans.

Based on the proposed scoring system, we tested the generation of four 6-element plans for DLBCL, our inspiration for the project (**Fig 3c**). We compared a plan generated by a human expert planner aided by CPB to a plan, with the first top 6 drugs suggested automatically by the app. As controls, we generated one plan which included 6 random drugs from top 30 suggestions (Rand30) and one with 6 random drugs from a list of 2000 FDA approved drugs (Rand2000). The results clearly show that both human and the top 6 were identical and received identical score that was more than 8-fold higher than

Rand30 control and orders of magnitudes higher than Rand2000 (**Fig 3d**). Interestingly, the RCHOPI which shows the best survival benefit in clinical trials[12] was the top 6 plan and human generated plan.

We then generated plans with 7 elements for three case studies and evaluated their score accordingly (**Fig 3e-f**). We chose one adenocarcinoma (colorectal cancer), one sarcoma (osteosarcoma) and one hematological cancer (multiple myeloma). As expected, plans designed by choosing seven random drugs (Rand2000) resulted in scores close to 0, even plans with seven random drugs chosen from top 30 (Rand30) received relatively low scores (<100). Differently from the DLBCL case, the highest score in this evaluation was for human generated plans followed by the top7 suggestion (**Fig 3f**). We noticed that in both human and top7 the plans for osteosarcoma were higher than multiple myeloma which was higher than colorectal cancer. This observation might reflect the nature of the disease and treatment.

Another possible level of plan evaluation is to identify drugs which are strongly linked to the disease itself (backbone drugs), and how many elements are strongly linked to other drugs (supporter drugs). To this end, we devised a radiometric scoring system that shows within a certain plan what is the distribution of backbone and supporter drugs with monoscore and comboscore respectively. The scores are a simple normalized ratio between the amount of evidence between the disease (backbone) vs the amount of evidence related to other drugs (supporter). We tested these scores on colorectal cancer plan and showed that two drugs are supporters (gemcitabine and paclitaxel), and 5 drugs are backbone with oxaliplatin and irinotecan being the highest-ranking backbone drugs (**Fig 3g**). This metric can facilitate decisions on drug administration scheduling and timing as supporters should be administrated with backbone drugs but backbone they can be staggered.

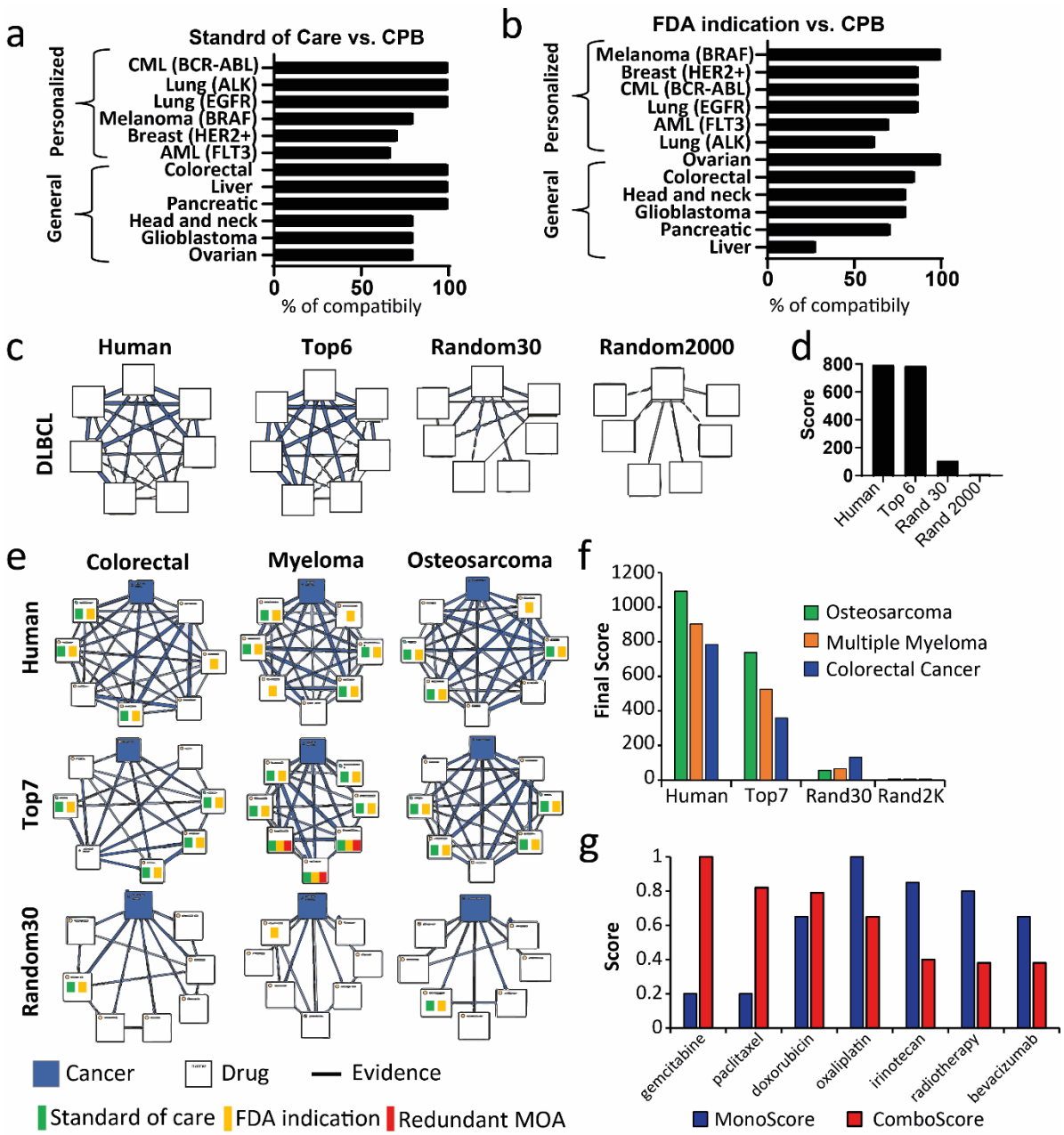


Figure 3. Evaluation of CPB app. a) % of Standard of care drugs in the top 10 ranked drug by the app from 2500. **b)** % of FDA approved drug for the specified indication in the top 10 ranked drugs by the app. **c**) 6 element plans for DLBCL. All plans were generated with the app either with a human choosing drugs based on evidence, drugs chosen from the top 7 ranked drugs, or random pick from top 30 or 2000 FDA approved drugs. **d)** plan evaluation score using with (weights/nodes) *edges. **e)** 7 element treatment plans for 3 cancer types (Colorectal, Multiple Myeloma and Osteosarcoma). All plans were generated with the app either with a human choosing drugs based on evidence, drugs chosen from the top 7 ranked drugs, or random pick from top 30. The drugs were verified for correct FDA indication approval, standard of care and mechanism of action (MOA). **f)** plan evaluation score with (weights/nodes) *edges. **g)** scoring each element within plan as mono vs combo therapy. top ranked elements were normalized to 1.

Evaluating against Manually Constructed Plans

For further evaluation of the app's ability to generate complex treatment plans, we compared it with an equivalent task performed by students using standard literature search methods. We generated fictional cancer patient case studies that were given to undergraduate students as part of a final assignment in a graduate course titled 'advanced methods in treating cancer'. The cases included the patient's cancer type, genetic finding, protein findings and more personalized characteristics. The task was to generate a complex treatment plan comprising of a minimum of 6 drugs or additional support from any type of non-pharmaceutical interventions such as radiotherapy. The timeframe for task completion was 1 month and at the end, we received 23 different treatment plans, each plan is a personalized and complex treatment (**Fig 4a**). We show an example summary of a student project outcome in Ewing's Sarcoma in **Supplementary Fig S2**. We analyzed the projects and compared the students' plans with known possible treatments for the same disease from two other resources, Wikipedia.com and the cancer.org. These resources contain summarized common treatment options in a clear and accessible way. As seen in **Fig 4b and supplementary Fig S3**, in all cases, the students offered more complex plans, with more diverse types of elements. The complexity is expressed mostly by incorporating more drugs and wider use of immunotherapy in the treatments. Also, the students included integrative medicine and chronotherapy that are usually not included in standard treatments.

In attempt to reach wider conclusions from the student plans, we searched for an agreement between the students regarding the timing element in their plans (**Fig 4c**). For example, we saw that in the majority of the plans (56.25%) the students suggested to first use chemotherapy and then to apply immunotherapy. This is reasonable, since some chemotherapy causes immunogenic cell death, which can educate the immune system to attack the dead cancer cells, resulting in a more efficient attack on living cancer cells following the immunotherapy[13]. The chronology of radiotherapy and immunotherapy was also widely agreed upon, the majority (75%) suggested the use of radiotherapy first, this is very reasonable biologically since radiotherapy induces an inflammation reaction that will be enhanced later by the immunotherapy and is more likely to reach improved outcomes[14]. Most of the plans concluded that chemotherapy should be used before (46.67%) or at the same time (40%) as radiotherapy. Finally, we have noticed most of the plans (60.87%) did not include specific tyrosine kinase inhibitors (TKIs), from those who did include TKIs in their treatment plan the main conclusion was to prescribe them with either radiotherapy or immunotherapy (21.72% from all plans).

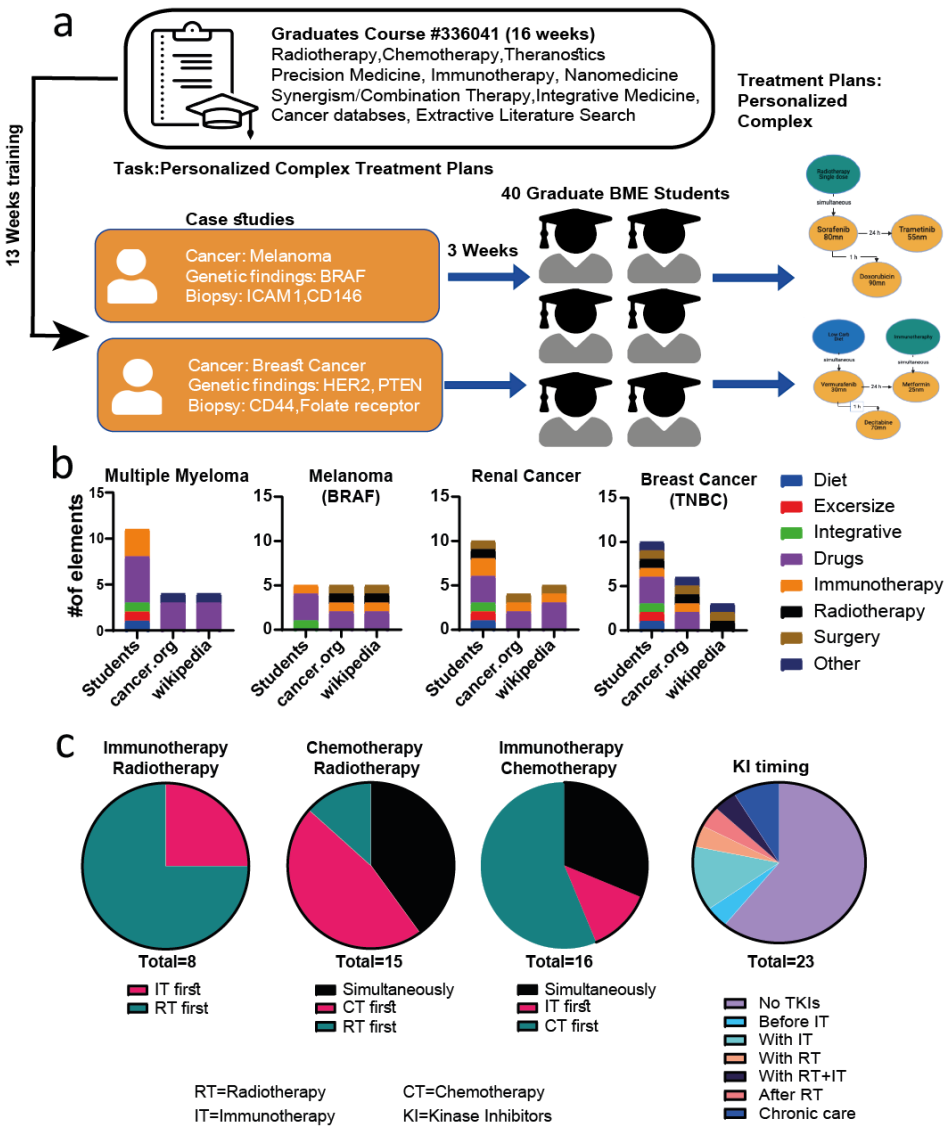


Figure 4. Crowdsourcing for high complex treatment plans. a) course workflow. **b)** number of treating elements in crowdsourced plans and the standards of care according to cancer.org and Wikipedia.com, divided by categories. **c)** order of treatments and timing of kinase inhibitors according to the crowdsourced plans.

The next step included a direct comparison between eight student generated plans and ones generated with the CPB app (**Fig 5a** and **5b)**. With CPB, each plan took less than 60 minutes to generate. Then, we compared them to plans generated by the students in the course and the approved treatment plans from Wikipedia and cancer.org, we saw a significant increase in the number of elements by a factor of 2.5 (**Fig 5a**). The CPB plans were versatile and most (62.5%) consist of three types of drugs- tyrosine kinase inhibitors, chemotherapies, and monoclonal antibodies- while the crowdsources plans consist mostly of two elements out of the three (**Fig 5b**).

In order to further compare the students' task with the app, we manually imported 3 student plans to the app for the purpose of scoring them with our scoring system. We formed 4 additional plans for the

same cancers, one human annotated, one consists of automatic choice of top 6 elements shown in the app, two random plans (Rand30 and Rand2000) and the 5th plan has 6 elements from the students' papers (**Fig 5c, and supplementary S5**). Elements such as exercise and diet are not recognized by the app and were not included. Overall, plans that were designed and annotated in the app by a human had the highest scores followed by plans formed by automatically choosing the top 6 elements (**Fig 5d**).

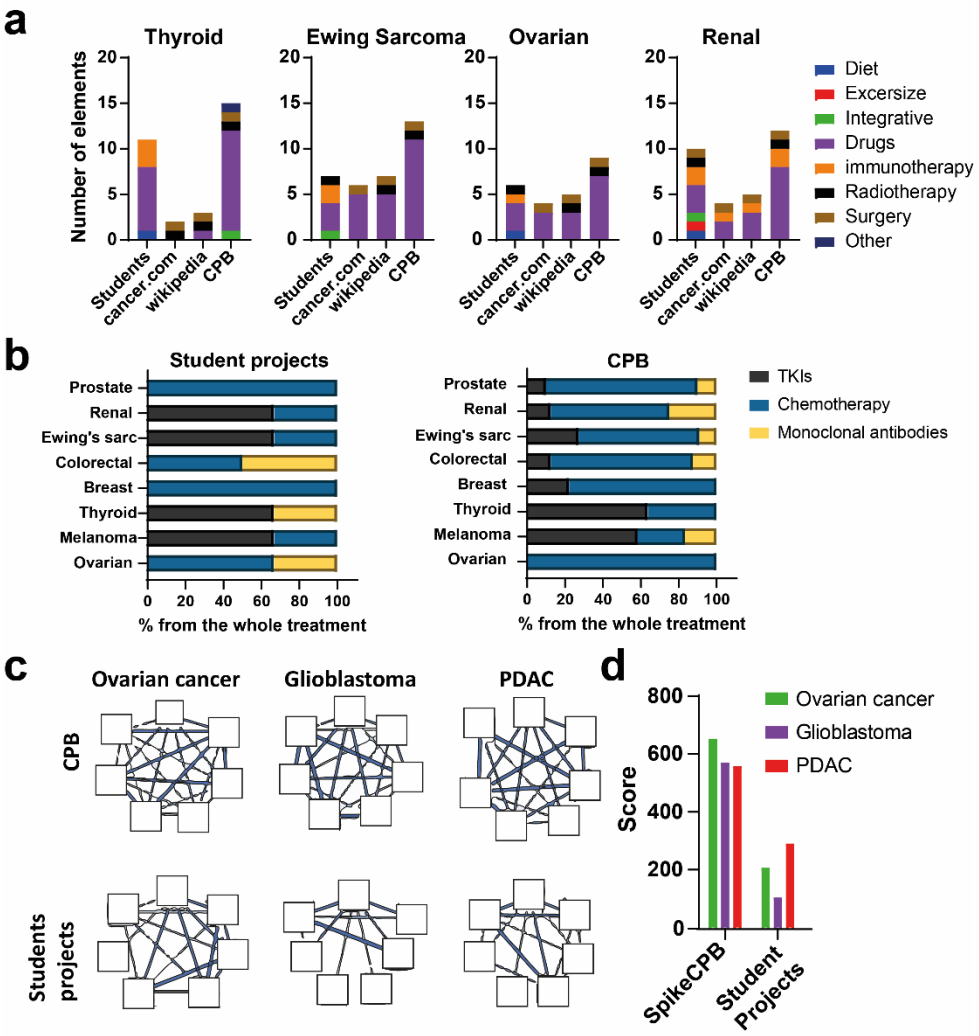


Figure 5. comparison between the crowdsourced plans and CPB app. a) number of treating elements in crowdsourced plans and the standards of care and CPB, divided by categories. b) distribution of drug types in the students' papers and plans formed by the app. c) 6-element treatment plans for four cancer types (DLBCL, ovarian cancer, glioblastoma and PDAC). All plans were generated with the app either by choosing from the top 6 ranked drugs (CPB) or from the students' projects. d) plans evaluation score.

Discussion

We were inspired by the historic development of multi component treatment regiments with high complexity, such as R-CHOP-I treatment for DLBCL[12]. Our interpretation of the process suggested that monotherapies which fail clinical trials but have significant outcomes can be added in a stepwise manner to the standard of care as a new combination therapy. In this example, the four drug combination CHOP (cyclophosphamide, doxorubicin, vincristine and prednisone) emerged as SOC in the late 1980s after decades and clinical trials reaching only 16% 10-year survival, until in 1997 the addition of rituximab was approved by the FDA and increased the overall survival to 43% by this five-drug treatment[15]. Recently, the addition of drugs such as ibrutinib to this combination was shown to increase survival to 83% in clinical trials[16]. In this case, the whole process for the development of new SOC for DLBCL took more than 3 decades. One of the reasons of the slow development of HCCT is the high burden of proof. The transition from a dual combination to an even slightly more complex combination is not a small nor simple. Conceptually, we need to acknowledge that with increasing complexity, the burden of proof increases dramatically which can be intimidating and highly limiting. For example, if we suggest a novel four-drug combo, we would be required to show it is better than any drug alone and pair and trio combo which yield 15 different control arms to test one hypothesis. It is therefore highly unfeasible to test all possible controls to prove that a complex combination is the optimal therapy. The correct combination integrated in a single treatment, could improve the outcomes significantly. It is suggested that many treatment elements would benefit from detailed methods for optimization and combinations of different elements, working together in harmony or synergy for a better therapy and outcome.

In this work, we exploit the scientific literature to build a framework for HCCT. We developed a dedicated tool, the CPB app together with a suggested workflow and metrics to evaluate the generated HCCT plans. In addition, we delegated the HCCT problem to the collective intelligence of graduate students (crowdsourcing), which is a complimentary approach to support and accelerate biomedical discoveries. By providing a platform upon which students can collaborate and communicate, aligning biomedical research with clinical needs to revolutionize healthcare worldwide.

Crowdsourcing problems can be broadly classified as either multiple microtasks or few megatasks. Microtasks are individually small tasks that typically come in big numbers where a large number of people contribute in parallel to save time and generate an aggregated solution. Megatasks are highly complex tasks which are often approached through open innovation challenges where the key

challenge is to find a few talented individuals from a large pool. An example for megatasks can be found in the work of Lakhani et al.[17]

Crowdsourcing has been used increasingly in oncology since 2012, covering cancer types as varied as breast, skin, ovarian, cervical, colorectal, and prostate[18]. The large pool of participants and time-efficient data collection that crowdsourcing affords, combined with low costs and the potential for rapid innovation, make such initiatives valuable in solving many of the challenging questions in oncology[19]. Good et al. demonstrated crowdsourcing usage for prognosis modeling case[20]. Their study found that the game called 'The Cure' demonstrated a crowdsourcing system that succeeded in capturing expert knowledge using an online web game. Smittenaar et al. crowdsourced the analysis of bladder cancer Immunohistochemistry data from tissue microarrays through smartphone app to accelerate the discovery of biomarkers and disease specific survival in bladder cancer[21].

Our approach for crowd sourcing was based on biomedical engineering (BME) students participating in a course which describes multiple therapeutic modalities in cancer. Specifically, to deal with clinical cancer related problems, and at the end of the course students were asked to work on a final project. The syllabus of the course included: Introduction to cancer hallmarks and personalized cancer medicine, the different types of cancer treatments such as radiotherapy, chemotherapy, etc. and the learning outcomes were as such: the students will be able to navigate through the complex world of cancer medicine using innovative tools and public databases (such as Human Protein Atlas, Broad CCLE). The final projects addressed a range of cancers (e.g., breast, skin, gynecologic, colorectal, prostate), in each the students wrote and presented a proposal for novel personalized combination therapies for hypothetic case studies. Each group tackled one type of cancer and tried to identify creative, yet reasonable and evidence based, solutions to treat it. After the students submitted their projects, they were reviewed manually by experts in the cancer field.

The plans formed by the app as well as plans planned by student show innovative approaches to treating different types of cancer and prove that complex, high resolution treatment can be planned by crowdsourcing and NLP based search.

Conclusion

To conclude, in this paper we develop a path to extend the boundaries of complexity in combinatorial cancer treatments. We first exploited TDM to characterize the current complexity of combination therapies in cancer and found that the current limit is 6 drugs per plan while the possible combinatorial space is astronomical. We then developed and validated a novel computational tool to generate

reasonable plan of any size. We showed that it performs better in any metric than a manual task performed by graduate students working on the same task for about a month.

Methods:

Complexity characterization

Clinicaltrials.gov. We used clinicaltrials.gov database and downloaded the data of ~7,300 trials from the past 10 years to a CSV file (file downloaded in July 2021). The search was filtered for 'cancer' in the disease field and the phrases 'combination' or 'combined' in the intervention field. From the CSV file we analyzed the different interventional arms in each trial and extracted the number of elements used in each one. We also extracted the number of drugs from the titles of the trials.

SPIKE search engine. This search engine was used in two stages. First, we searched for different phrasings used in published papers to describe drug combinations or drug regimens (for example, the phrasings 'drug combination', 'drug Regiment' or 'combining five drugs', etc.), this step was done by using the structure search in SPIKE search engine. Then, using the list formed in the first step and basic search, we extracted the number of drugs combined in a single treatment in all diseases, in cancer in general and in specific types of cancer. Searches were done in July 2021.

Crowdsourcing case studies and workflow

In Course number 336014 titled "Advanced Biomedical Methods for Cancer Therapy and Diagnosis" 40 students were taught during a semester about current therapeutic modalities and different aspects of cancer. The syllabus included subjects such as the hallmarks of cancer, personalized medicine, chemotherapy, targeted therapy, integrative medicine, etc.

In addition, we trained the students in extractive literature search using SPIKE search engine and other cancer databases such as human protein atlas (HPA), the broad institute cell encyclopedias (CCLE), TCGA genetic database, etc. The final task in this course was to generate high complexity combination therapy for fictional case studies, we demanded the use of minimum five intervention elements and recommended to search for drug synergies in the literature, and personalized treatment modalities to the patient. Each fictional case study had a cancer type, genetic finding (different mutated genes), protein findings, and more tumor microenvironment characteristics.

Scoring methods

Scoring the plans. To score the plans formed by CPB, the raw data was exported to a CSV. file and analyzed. the total number of evidence is the sum of the connections, we also counted the number of connections and elements in the plan, and eventually came up with a scoring method that combines the three parameters in order to have a comparison between different plans. This scoring method can be used to compare between plans for the treatment of a certain cancer type and the comparison of different cancer types is not fair, since not all cancers are studied equally.

Monotherapy and combination therapy ratings. The raw data of a plan built in CPB was downloaded to a CSV. file and the treatments were firstly organized according to their level of connectivity to the cancer type- this resulted in the monotherapy rating (the most connected has the best rating and so on). By forming a pivot table from the raw data and combining the level of connectivity between each treatment to the other treatments, we were able to rate how much it is used in combination therapies.

CPB tool development

To identify, aggregate and rank treatments the treatment planning app employs the SPIKE extractive search engine[10,22]. SPIKE is a freely accessible search engine over PubMed which offers a number of relevant capabilities absent from PubMed Search (i) it allows to load very large lists of terms and use these in queries (compared to PubMed which limits searches to 2048 terms); (ii) it allows for fine grained control over the scope of the matches, e.g. requiring a treatment term and a cancer term to appear in the same sentence or even in a specific syntactic configuration; and (iii) it supports "extractive searches", which means it not only finds sentences which match a query, but also extracts spans of text which correspond to designated query terms.

Searching for First Line Treatments. To support the search and extraction of treatments, we pre-curated a comprehensive list of drugs, and non-drug cancer treatments. The drug list was derived from drugbank.com with manual deletion of supplementary food, minerals and vitamins. Based on the input cancer name, genetic findings and additional keywords (hereafter referred to as biomarkers), the app runs a number of SPIKE-PubMed queries using a backoff strategy (see full syntax in Supplementary Materials). Queries are run in succession until the 100 most related treatments are obtained.

- The first query is for sentences containing a treatment from the treatments list and a biomarker from the patients' biomarkers. The sentences are restricted to originate from papers whose **title** includes the input cancer name

- If the result set does not include at least 100 unique treatments, we run a similar query, this time restricting the result sentences to originate papers whose **title or abstract** contain the input cancer name. We add the additional results into the result set.
- If the result set still does not include 100 unique treatments, we run a similar query, this time allowing for biomarkers to appear in the same paragraph as the treatment (rather than the same sentence).

We extract the treatment from each of the result sentences and score a treatment by the number of matches which referenced it. To reflect the fact that the queries are ordered by the likelihood of retrieving relevant treatments, a treatment gets 3 points for every match due to query 1, 2 points for matches due to query 2 and 1 point for matches due to query 3. The treatments are then ranked by their aggregated score and displayed to the user. The evidence for each treatment is a list of abstracts where matching sentences are highlighted as well as the relevant treatment and matched biomarkers in them. The evidence abstracts are ranked by (i) the unique number of matched biomarkers in them, and (ii) in case of a tie, the total number of matched biomarkers in them. Note that the ranking system is a critical component of the solution and massively reduces planning time. Instead of reviewing hundreds of papers in order to identify treatments and sort which ones are more frequent and more related to the input conditions, the app immediately surfaces the specific treatments which are most frequently discussed in the literature. Also, the ranking within the evidence snippets of a treatment leads to planners typically needing to review only 1-3 specific sentences to validate the treatment relevance

Searching for Synergistic Treatments. When the user clicks an existing treatment in the plan, the app searches PubMed for treatments which are potentially synergistic to it (with or without reference to the given cancer name). To support this search, we curated a list of terms whose presence in a sentence along with a pair of treatments typically indicates a synergistic combination (see Supplementary Materials).

As before, we run a sequence of SPIKE-PubMed queries using a backoff strategy:

- The first query is for sentences containing the input treatment, another treatment from our treatments list and a term indicating a synergistic combination. The sentences are restricted to originate from papers whose **title** contains the input cancer name.
- If the result set does not include 100 unique treatments, we rerun the query, this time restricting the result sentences to papers whose **title or abstract** containing the input cancer name
- If the result set still does not include 100 unique treatments, we rerun the query, this time **not restricting** the sentences to appear together with the input cancer name.

The ranking of treatments and of the evidence for each treatment is proceeds as described for first line treatments.

Integrating Supervised Machine Learning (ML) Models.

Supervised ML models typically have higher coverage then pattern-based methods, but they also introduce some additional noise. Integrating such models into our app is straightforward. Given a model which identifies drug combinations in text and labels the combination as synergistic or antagonistic, extract all such combinations from PubMed into a database, along with confidence scores and frequency data. This database can then be queried by the planning app to suggest treatments to the user. Based on Tiktinski[23] , we also integrated the output of ML models for drug combinations to the app, in an optional mode that allows the planner to use the more precise and more replicable non-ML results, or to extend the results with the higher coverage but noisier ML output. The experiments in this work are conducted on the more precise mode.

Acknowledgements

We would like to acknowledge the following funding sources: ISF grant #901/91, MOH grant # 3-14989.

References

1. Thanopoulou, E. & Judson, I. The safety proWle of imatinib in CML and GIST: long-term considerations. *Arch Toxicol* **86**, 1–12 (2012).

2. Goy, A. Succeeding in breaking the R-CHOP Ceiling in DLBCL: Learning from negative trials. *J. Clin. Oncol.* **35**, 3519–3522 (2017).
3. Guan, X. W. *et al.* Novel HDAC inhibitor Chidamide synergizes with Rituximab to inhibit diffuse large B-cell lymphoma tumour growth by upregulating CD20. *Cell Death Dis.* **11**, (2020).
4. Younes, A. *et al.* Randomized Phase III Trial of Ibrutinib and Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone in Non–Germinal Center B-Cell Diffuse Large B-Cell Lymphoma. *J. Clin. Oncol.* **37**, 1285 (2019).
5. Lauriola, M. *et al.* Diurnal suppression of EGFR signalling by glucocorticoids and implications for tumour progression and treatment. *Nat. Commun.* **5**, (2014).
6. van Leeuwen, R. W. F. *et al.* Tyrosine Kinase Inhibitors and Proton Pump Inhibitors: An Evaluation of Treatment Options. *Clin. Pharmacokinet.* **56**, 683 (2017).
7. Jacobi, J. *et al.* Targeting acid sphingomyelinase with anti-angiogenic chemotherapy. *Cell. Signal.* **29**, 52 (2017).
8. Hersh, D. S. *et al.* MR-guided transcranial focused ultrasound safely enhances interstitial dispersion of large polymeric nanoparticles in the living brain. *PLoS One* **13**, (2018).
9. Mead, B. P. *et al.* Focused Ultrasound Pre-Conditioning for Augmented Nanoparticle Penetration and Efficacy in the Central Nervous System. *Small* **15**, e1903460 (2019).
10. Shlain, M., Taub-Tabib, H., Sadde, S. & Goldberg, Y. Syntactic Search by Example. 17–23 (2020) doi:10.18653/V1/2020.ACL-DEMOS.3.
11. Luo, X.-Y., Wu, K.-M. & He, X.-X. Advances in drug development for hepatocellular carcinoma: clinical trials and potential therapeutic targets. doi:10.1186/s13046-021-01968-w.
12. Wilson, W. H. *et al.* Effect of ibrutinib with R-CHOP chemotherapy in genetic subtypes of DLBCL. *Cancer Cell* **39**, 1643-1653.e3 (2021).
13. Lake, R. A. & Robinson, B. W. S. Immunotherapy and chemotherapy — a practical partnership. *Nat. Rev. Cancer 2005 55* **5**, 397–405 (2005).
14. Wang, Y. *et al.* Combining immunotherapy and radiotherapy for cancer treatment: Current challenges and future directions. *Front. Pharmacol.* **9**, 185 (2018).
15. Sehn, L. H. A decade of R-CHOP. *Blood* **116**, 2000–2001 (2010).
16. Pasvolsky, O., Rozental, A., Raanani, P., Gafter-Gvili, A. & Gurion, R. R-CHOP compared to R-CHOP + X for newly diagnosed diffuse large B-cell lymphoma: a systematic review and meta-analysis. *Acta Oncol.* **60**, 744–749 (2021).
17. Lakhani, K. R. *et al.* Prize-Based Contests Can Provide Solutions to Computational Biology Problems. *Nat. Biotechnol.* **31**, 108 (2013).

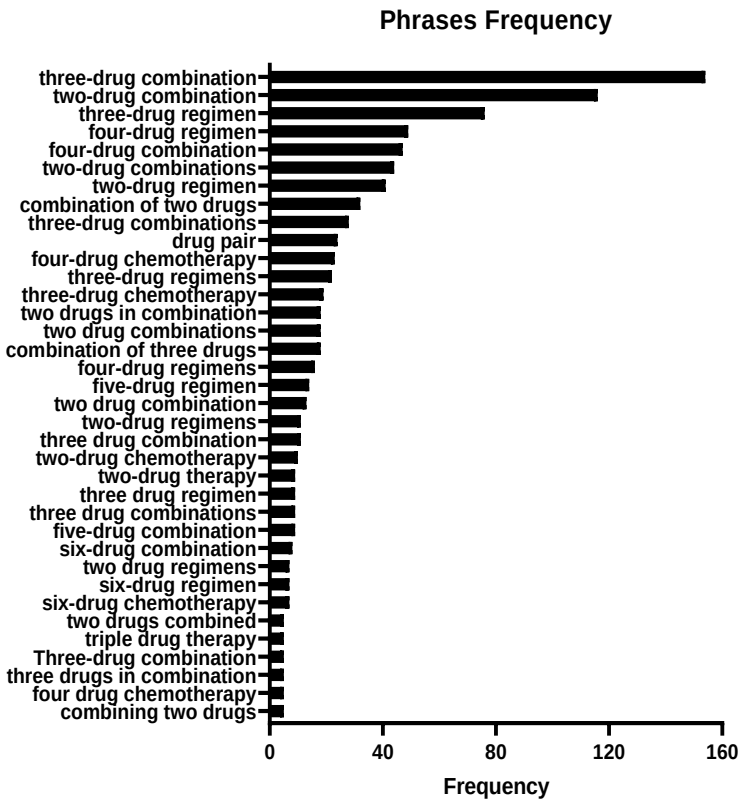
18. Lee, Y. J., Arida, J. A. & Donovan, H. S. The application of crowdsourcing approaches to cancer research: a systematic review. *Cancer Med.* **6**, 2595 (2017).
19. Desai, A. *et al.* Crowdsourcing a crisis response for COVID-19 in oncology. *Nat. cancer* **1**, 473–476 (2020).
20. Good, B. M. *et al.* The Cure: Design and Evaluation of a Crowdsourcing Game for Gene Selection for Breast Cancer Survival Prediction. *JMIR Serious Games* **2**, (2014).
21. Smittenaar, P. *et al.* Harnessing citizen science through mobile phone technology to screen for immunohistochemical biomarkers in bladder cancer. *Br. J. Cancer* **119**, 220–229 (2018).
22. Taub-Tabib, H. *et al.* Interactive Extractive Search over Biomedical Corpora. 28–37 (2020) doi:10.18653/V1/2020.BIONLP-1.3.
23. Tiktinsky Aryeh, Viswanathan Vijay, Niezni Danna, Meron Azagury Dana, Shamay Yosi, Taub-Tabib Hillel, H. T. and G. Y. A Dataset for N-ary Relation Extraction of Drug Combinations. in *North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2022).

Tables:

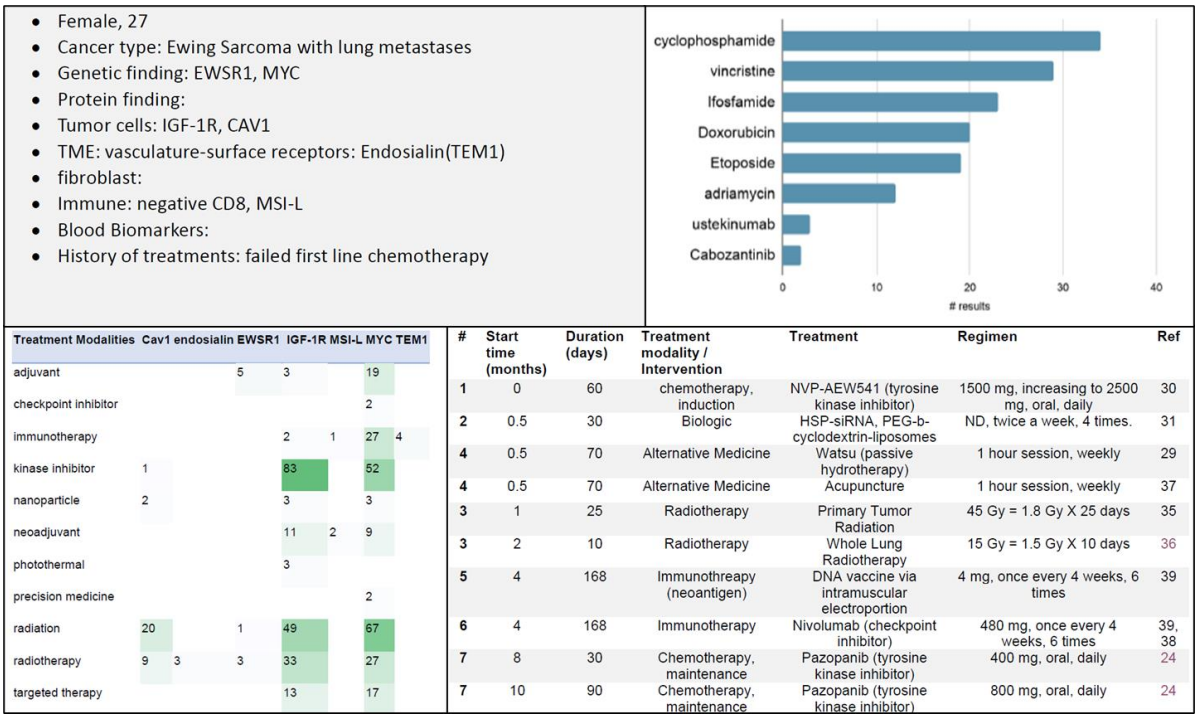
Table 1

symbol	**represents**	**unction**
P	sum evidence	mount of information
R	sum relations	Connectivity variable
N	sum nodes	Connectivity variable
P * (R/N)	multiplication	Final Score

Supplementary Information:



Supplementary Fig S1. Frequency of phrasing combinatorial drug treatments.



- Female, 27
- Cancer type: Ewing Sarcoma with lung metastases
- Genetic finding: EWSR1, MYC
- Protein finding:
- Tumor cells: IGF-1R, CAV1
- TME: vasculature-surface receptors: Endosialin(TEM1)
- fibroblast:
- Immune: negative CD8, MSI-L
- Blood Biomarkers:
- History of treatments: failed first line chemotherapy

Treatment Modalities	Cav1	endosialin	EWSR1	IGF-1R	MSI-L	MYC	TEM1
adjuvant			5	3		19	
checkpoint inhibitor						2	
immunotherapy				2	1	27	4
kinase inhibitor	1			83		52	
nanoparticle	2			3		3	
neoadjuvant				11	2	9	
photothermal				3			
precision medicine						2	
radiation	20		1	49		67	
radiotherapy	9	3	3	33		27	
targeted therapy				13		17	

#	Start time (months)	Duration (days)	Treatment modality / Intervention	Treatment	Regimen	Ref
1	0	60	chemotherapy, induction	NVP-AEW541 (tyrosine kinase inhibitor)	1500 mg, increasing to 2500 mg, oral, daily	30
2	0.5	30	Biologic	HSP-siRNA, PEG-b-cyclodextrin-liposomes	ND, twice a week, 4 times.	31
4	0.5	70	Alternative Medicine	Watsu (passive hydrotherapy)	1 hour session, weekly	29
4	0.5	70	Alternative Medicine	Acupuncture	1 hour session, weekly	37
3	1	25	Radiotherapy	Primary Tumor Radiation	45 Gy = 1.8 Gy X 25 days	35
3	2	10	Radiotherapy	Whole Lung Radiotherapy	15 Gy = 1.5 Gy X 10 days	36
5	4	168	Immunothreapy (neoantigen)	DNA vaccine via intramuscular electroportion	4 mg, once every 4 weeks, 6 times	39
6	4	168	Immunotherapy	Nivolumab (checkpoint inhibitor)	480 mg, once every 4 weeks, 6 times	39, 38
7	8	30	Chemotherapy, maintenance	Pazopanib (tyrosine kinase inhibitor)	400 mg, oral, daily	24
7	10	90	Chemotherapy, maintenance	Pazopanib (tyrosine kinase inhibitor)	800 mg, oral, daily	24

Supplementary Fig S2. elements from an example stunent's paper, decriptions from top left to bottom right. (A) the case study and patients characteristics, (B) relevant drugs for Ewing sarcoma (extracted from SPIKE), (C) the number of articles connecting between treatment modalities and the protein findings, (D) the final, detailed treatment plan.