PaccMann^{RL}: Designing anticancer drugs from transcriptomic data via reinforcement learning

Jannis Born[*] **Matteo Manica**[*] **Ali Oskooei**[*] **Joris Cadow** **María Rodríguez Martínez**
{jab,tte,osk,dow,mrm}@zurich.ibm.com
IBM Research, 4 Säumerstrasse, Rüschlikon, Switzerland 8803
[*] Equal contributions

Abstract

With the advent of deep generative models in computational chemistry, in silico anticancer drug design has undergone an unprecedented transformation. While state-of-the-art deep learning approaches have shown potential in generating compounds with desired chemical properties, they disregard the genetic profile and properties of the target disease. Here, we introduce the first generative model capable of tailoring anticancer compounds for a specific biomolecular profile. Using a RL framework, the transcriptomic profiles of cancer cells are used as a context for the generation of candidate molecules which is optimized through PaccMann (a previously developed drug sensitivity prediction model) to obtain effective anticancer compounds for the given context (i.e., transcriptomic profile). We verify our approach by investigating candidate drugs generated against specific cancer types and find the highest structural similarity to existing compounds with known efficacy against these cancer types. We envision our approach to transform in silico anticancer drug design by increasing success rates in lead compound discovery by leveraging the biomolecular characteristics of the disease.

1 Introduction

The last two decades have seen a decline in the productivity of the drug discovery pipeline while the investment into drug discovery has risen significantly [1]. Indeed, only a minimal portion of drug candidates obtain market approval (less than 0.01%), with an estimated 10-15 years until market release and costs that range between one [1] to three billion dollars per drug [2]. This low efficiency has been attributed to the high dropout rate of candidate molecules in the early stages of the pipeline, highlighting the need for more accurate in silico and in vitro models that produce more successful candidate drugs. Most recently, deep learning methods have gained popularity within the computational chemistry community [3] and a number of works have demonstrated the feasibility of in silico design of novel candidate compounds with desired chemical properties [4–6]. While very effective in generating compounds with desired chemical properties, these methods do not integrate information about the cellular environment in which the drug is intended to act. However, the two main causes of the increasing attrition rate in drug design are lacking efficacy against the specific disease of interest and off-target cytotoxicity [7], calling to bridge systems biology closer with drug discovery. In addition to the initial wet-lab validations, the discovery pipeline involves a sequential process that builds upon high-throughput screenings, ADMET-assessments (absorption, distribution, metabolism, excretion and toxicity, i.e., criteria for the pharmacological activity of a compound) and a lengthy phase of clinical trials. The costs of the experimental and clinical phase can be prohibitive and any solution that helps to reduce the number of required experimental assays can provide a competitive advantage and reduce time to market. To this end, we present a deep RL based model for anticancer molecule generation that builds on top of the previous approaches and, for the first time, enables generation of novel anticancer compounds while taking into account the disease

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

context encoded in the form of gene expression profile (GEP) of the disease e.g., tumor. The training procedure for the generator and the critic models consists of two stages. In the first stage, each model is trained independently. In the second stage, the generative model is retrained and optimized via a RL approach with a reward coming from the critic module. The goal of the optimization is to tune the generative model such that it generates (novel) compounds that have maximal efficacy against a given biomolecular profile; be it the characteristic for a cancer site, a patient subgroup or even an individual. By *efficacy*, we refer to cellular IC50 (i.e. the micromolar concentration necessary to inhibit 50% of the cells in a sample) as opposed to e.g. enzymatic IC50. It is important to note that this efficacy is a joint property of a drug-cell-pair and empirically it is well-known that the efficacy of a compound heavily varies for different types of cells. In this work, we emphasize profile-specific compound generation and optimize the generator using IC50 as the sole critic.

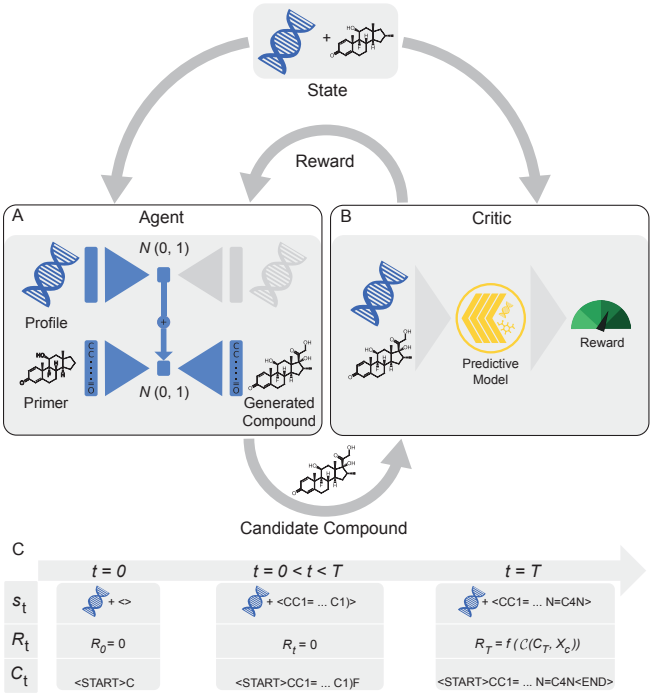

Figure 1: **The proposed framework for anticancer compound design against specific cancer profiles.** A biomolecular profile VAE (PVAE, top row of (A)) and a sequential compound generator VAE (SVAE, bottom row of (A)) are combined to obtain a conditional molecule generator. Each of the PVAE, SVAE and the predictive critic are pretrained independently. Thereafter, the conditional generation process starts with a biomolecular profile of interest e.g., transcriptomic profile from an individual patient. The given profile is encoded into the latent space of gene expression profiles and is then decoded through the molecular decoder to produce a candidate compound. The generative process can be "primed" through encoding a known, effective compound or a functional group with the molecular encoder. The proposed compound is then evaluated through the critic, a multimodal drug sensitivity prediction model that ingests the compound and the target profile of interest (B). The RL based optimization is conducted by maximizing the reward given by the critic. Over the course of training, the generator will thus learn to produce candidate compounds with higher and higher efficacy. The RL training evolution including the state (s_t), the reward (R_t) and the candidate compound (C_t) during a complete training cycle are shown in (C).

2 Methods

Figure 1 shows the two main components of our proposed end-to-end architecture: the conditional generator (Figure 1A) and the critic (Figure 1B).

Conditional generator ($\mathcal{G}$). This is a molecule generator that produces a candidate drug structure using its SMILES string representation [8]. In our use case, the generative process needs to be conditioned on a target biomolecular profile, e.g., from a patient or a disease. Inspired by [5], we concluded that VAEs are the ideal model for our task since by design they bring about a structurally ordered latent space that simplifies the combination of different information sources (in our case

transcriptomic profiles and chemicals). Our conditional generator combines two VAEs that are trained independently prior to being fused together: 1) a denoising VAE (with for cancer profile encoding/generation (PVAE) and 2) a sequential VAE (SVAE) for SMILES sequence generation. The detailed derivation and equations for the VAE can be found in [9, 10]. PVAE and SVAE are initially pretrained in isolation. PVAE is trained on gene expression profiles (GEP) to learn a consistent latent representation for biomolecular signatures. SVAE is trained on bioactive drug-like molecules to learn the syntax of valid SMILES and general molecular semantics. The fact that models that process SMILES sequences must have the ability to *count* the ring opening and closing symbols in a molecule necessitates the use of stack memory [11], in our case stack-augmented GRUs as proposed by [12]. Thereafter, the encoder of the PVAE is fused with the decoder of the SVAE via their latent space. The combination of the two models enables to learn a latent space that links biomolecular profiles and chemical structures providing an effective way to sample novel compounds given a specific GEP. In the final training phase, the weights of the fused model are fine-tuned using a reward from the critic in a RL framework.

Critic ($\mathcal{C}$). The critic is a multimodal drug sensitivity prediction model which evaluates the efficacy of any given candidate compound against a biomolecular profile of interest (e.g., gene expression of a cancer cell line). It outputs a non-negative reward, a function of the predicted IC50 of the candidate compound for the target profile, which is used in a RL framework to update the conditional generator. Following the most recent advances for multimodal drug sensitivity prediction we herein utilize PaccMann as a critic; specifically multiscale convolutional encoders as proposed in [13].

The RL framework. The conditional generator is retrained in combination with the critic in a RL-based optimization process to tailor molecules towards the given GEP. First, the GEP is encoded into its latent space, Z_c. This embedding is then added to the latent encoding of a primer compound or substructure (Z_p). The advantage of using a primer is that it enables injection of prior knowledge into the model by starting the generative process from an existing and proven effective compound or functional group – instead of designing a compound from scratch. Formally, the molecule generation is conditioned on a context $\mathcal{Z}$, where in this work $\mathcal{Z} = \{Z_c, Z_p\}$. Since Z_c and Z_p reflect embeddings learned from semantically different data sources (gene expression and molecules) it is non-trivial to combine them meaningfully. We use a summation because it is a permutation invariant operation and has been proposed in the deep sets architecture [14] to combine a variable set of unstructured latent encodings. Our additive latent representation is similar in concept to the conditional VAE with additive Gaussian encoding space [15]. Intuitively, this fusion most likely warps the latent space from encoding structural similarity (of molecules or GEP) into functional similarity [5] so as to aggregate molecules with similar predicted efficacy for a given cell line. Note that using a primer compound or substructure is optional and if no priming compound is used, simply the latent space representation of the `<START>` token is added to the latent encoding of the target GEP.
Next, the conditional generator decodes the latent encoding, Z_c + Z_p, and generates a molecular structure that, in combination with the GEP, is fed to the critic to produce a certain reward for the generated compound, as illustrated in Figure 1A and B. Following the notation of [4], the conditional generator, $\mathcal{G}$, acts as the *agent* and the multimodal IC50 prediction model, $\mathcal{C}$, represents the *critic*. We aim to optimize Θ, the parameters of $\mathcal{G}$, to produce candidate compounds, C_T, that target a specific GEP, X_c. In contrast to [4], we define the set of states $\mathcal{S}$ as all possible SMILES strings (with length $\leq T$) paired with the target GEP. As depicted in Figure 1C, molecules are generated by $\mathcal{G}$ by sampling an action a_t at each step($0 < t < T$) from $p(a_t|s_{t-1})$, where $s_{t-1} = (C_{t-1}, X_c)$. Terminal states $S^* \subset S$ are reached when either $t = T$ or when the terminal action $a_T =$ `<END>` has been sampled. $\mathcal{G}$ is trained to learn a policy, $\Pi(\Theta)$, by maximizing: $\Pi(\Theta) = \sum_{s_T \in S^*} p_\Theta(s_T) R(s_T)$. The reward $R_T = R(s_T) = f(\mathcal{C}(C_T, X_c))$ is the output of the critic scaled by a reward function f. In our experiments, all intermediate rewards $R(s_t) = 0$ where $t < T$, the sum is approximated using the REINFORCE algorithm [16] and the reward function f for determining the reward from the IC50 prediction, $\mathcal{C}(C_T, X_c)$, is computed by $f(IC50) = \exp\left(\frac{-IC50}{5.0}\right)$.

Data. For the PVAE, we employed a training dataset of 11,592 (normalized) RNA-Seq GEPs from healthy and cancerous human tissue from the TCGA database and validated it on 1,289 samples from the same database [17]. The number of genes was reduced to the same 2,128 genes as used in [18, 13], following the network propagation procedure described in [19]. The SVAE was pretrained on the SMILES representation of 1,576,904 compounds (10% were held out for performance validation) from the ChEMBL database [20]. For RL optimization of $\mathcal{G}$, we used GEPs publicly available from GDSC [21] and CCLE [22] databases. Drug sensitivity data (i.e., IC50) from the GDSC [21] and

CCLE [22] databases were used to train the IC50 prediction model, the critic ($\mathcal{C}$). We used the screening data for targeted small molecules to train the IC50 prediction model.

3 Results

Four different models were trained, one for each of the cancer sites: breast, lung, prostate and neuroblastoma. For the evaluation of the four models, all generated compounds with a predicted

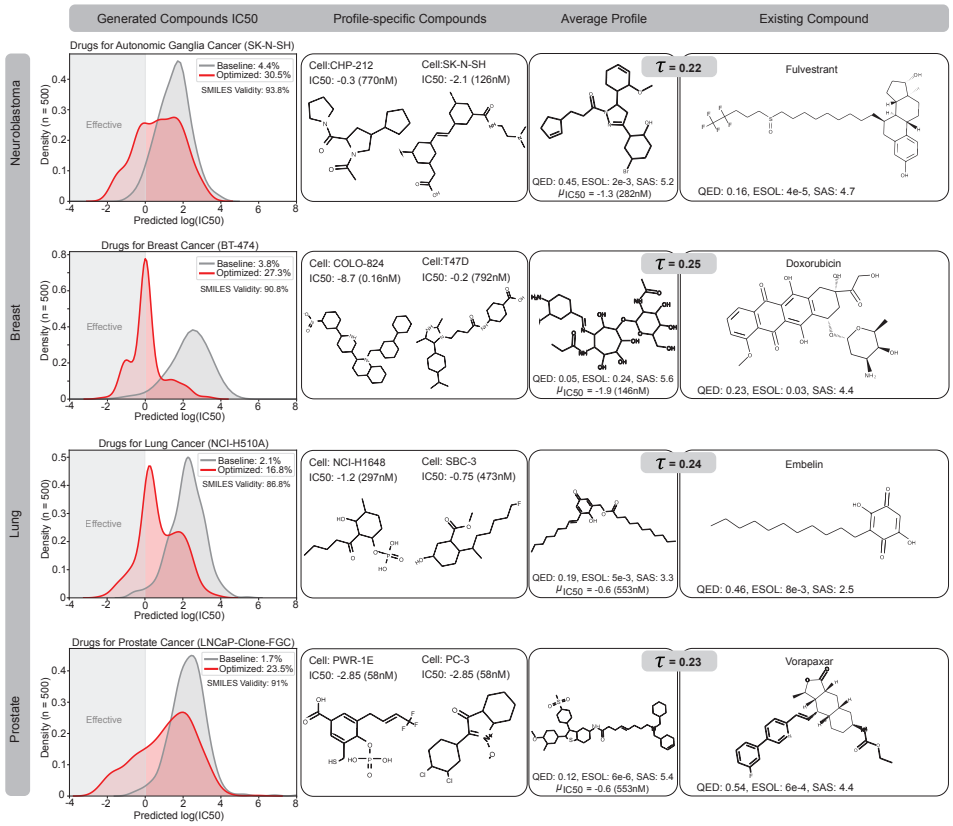

Figure 2: **Sample results for profile-driven model optimization and anticancer compound generation.** Each row illustrates the results of training the RL pipeline on cell lines from a specific site: neuroblastoma, breast, lung and prostate cancer. The first column compares the distributions of IC50 predictions given by the critic model for a set of n=500 drug candidates generated with RL optimization and without RL optimization. The second column presents candidate compounds with a high predicted efficacy (low IC50) against a particular cell line that was not seen during training. The third column showcases generated compounds that were optimized to be effective against the average cell-line profiles of the given cancer type in each row. In the fourth column, we present an *existing* anticancer compound (approved against at least one type of cancer), that was in the top-3 neighborhood of the generated compound in the third column. The existing and generated compounds are compared in terms of Tanimoto structural similarity as well as three chemical scores crucial in drug design namely, druglikeness (QED, 0 worst, 1 best), synthesizability (SAS, 1 best, 10 worst) and solubility (ESOL, given in M/L).

IC50 value below 1μM were considered as *effective*. Moreover, within each cancer site 80% of the cell lines were considered as training cell lines and used to optimize the parameters Θ of the conditional generator whereas the remaining 20% were set aside for testing. Our model learned to produce compounds with lower IC50 values, for unseen cell lines from the given cancer site. The first column of Figure 2 shows that the IC50 distribution of candidate compounds proposed by the generative model were successfully shifted towards higher efficacy. The baseline model corresponds to the pretrained SVAE from which $n = 500$ molecules were randomly sampled. In all four cases, a significant portion (between 17% and 30%) of generated molecules were assigned a IC50 value below 1μM, whereas only 1-4% of the candidates generated by the baseline model (i.e., the SVAE, a generative model of drug-like molecules trained with data from ChEMBL that was not yet optimized via RL retraining) were classified as effective. The second column of Figure 2 shows generated

molecules that are predicted as being effective against an unseen cell line from the respective cancer site. In the third column of Figure 2, we showcase novel molecules that were designed specifically for each cancer site, as opposed to a specific cancer cell profile. All compounds exhibited high predicted efficacy against the average cellular profile of the target site while maintaining efficacy against the majority of individual cell lines for that site. In the last column of Figure 2, we compare the four site-specific candidate compounds with one of their top-3 neighbors (Tanimoto similarity, τ) from several hundreds of existing anticancer compounds. The third closest neighbor of the generated compound against neuroblastoma (Figure 2 first row, third column) is Fulvestrant, an antagonist/modulator of ERα which has recently been proposed as a novel anticancer agent for neuroblastoma [23]. The candidate compound proposed against breast cancer (Figure 2 second row, third column), has Doxorubicin as one of the top-3 nearest neighbors. Doxorubicin is a commonly used chemotherapeutical against breast cancer [24]. The generated compound against lung cancer (Figure 2, third row, third column) results close to Embelin, an existing anticancer compound from the GDSC database. Embelin is known to be a promising anticancer compound as it is the only known non-peptide inhibitor of the XIAP protein [25], a protein that plays an important role in lung cancer development [26]. Lastly, the closest neighbor of the prostate-specific generated compound (Figure 2 fourth row, third column) is Vorapaxar. Its efficacy is highest against a prostate cancer cell line (DU_145) according to GDSC/CCLE. Vorapaxar is an antagonist of a protease-activated receptor (PAR-1) that is known to be overexpressed in various types of cancer, including prostate [27].

4 Discussion

We herein presented the first deep-learning based anti-cancer compound generator that enables us to condition the molecular generation on the biomolecular profile (transcriptomic profile in this work) of the target cell, tumor or cancer site. We demonstrated, using a RL optimization framework, that our proposed generative model could be optimized to produce candidate compounds with high predicted efficacy (IC50) against a given target profile. We showcased in a post-hoc analysis that each of the four site-specific generated compounds had structural similarities to known anticancer compounds commonly used to treat cancer of the same type as the generated compound was optimized for. Oftentimes however, medical chemists do not start the drug design from scratch, but from an approved drug and with the goal to find a drug with similar effects (e.g. increased efficacy or reduced side effects). Our framework neatly grants the option to incorporate this prior knowledge into the design process already. Another endeavour is to develop a drug that specifically targets a protein (e.g. one that has been implicated in tumor proliferation or treatment response according to a gene-knockout study). Whilst [28] very recently presented a model that proposed potent DDR1 kinase inhibitors, we are working towards a generic framework where the molecule generation can be conditioned on possibly multimodal context information such as a target protein, a primed drug, a tumor profile and notably also a combination thereof by utilising permutation invariant operations in the multimodal latent space [14]. We have, however, not yet fully explored the full potential of the framework, e.g. by conditioning the design on a known drug in conjunction with a cell profile. While we believe our results to be a promising stepping stone for profile-specific anticancer compound generation, we are aware further optimization must be done before it can be used a reliable tool for drug discovery. For instance, there are various other properties of a candidate drug other than its efficacy that determine its potential for becoming a successful anticancer compound. Our RL optimization framework can easily be extended with further critics that reward or punish the conditional generator for other critical properties, in addition to efficacy against a given cancer profile. In the future, we aim to amend the reward function to compute rewards not only for efficacy but also based on predicted cytoxicity, solubility, drug-likeness and synthesizability.

5 Availability of software and materials

The omics data used to pretrain the PVAE, the molecular data for the SVAE and the cell profiles used in the RL regime as well as the pretrained models can be found on https://ibm.box.com/v/paccmann-pytoda-data. To assess the critic, please see [13]. All code to reproduce the experiments is publicly available on https://github.com/PaccMann/.

Acknowledgements

The project leading to this publication has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826121.

References

[1] Jack W Scannell, Alex Blanckley, Helen Boldon, and Brian Warrington. Diagnosing the decline in pharmaceutical r&d efficiency. *Nature reviews Drug discovery*, 11(3):191, 2012.

[2] Gisbert Schneider. Mind and machine in drug design. *Nature Machine Intelligence*, page 1, 2019.

[3] Hongming Chen, Ola Engkvist, Yinhai Wang, Marcus Olivecrona, and Thomas Blaschke. The rise of deep learning in drug discovery. *Drug discovery today*, 2018.

[4] Mariya Popova, Olexandr Isayev, and Alexander Tropsha. Deep reinforcement learning for de novo drug design. *Science advances*, 4(7):eaap7885, 2018.

[5] Rafael Gomez-Bombarelli, Jennifer N Wei, David Duvenaud, Jose Miguel Hernandez-Lobato, et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.

[6] Jiaxuan You, Bowen Liu, Zhitao Ying, Vijay Pande, and Jure Leskovec. Graph convolutional policy network for goal-directed molecular graph generation. pages 6412–6422, 2018.

[7] Martin Wehling. Assessing the translatability of drug projects: What needs to be scored to predict success? *Nature reviews. Drug discovery*, 8:541–6, 07 2009. doi: 10.1038/nrd2898.

[8] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28 (1):31–36, 1988.

[9] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[10] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *Advances in neural information processing systems*, pages 3483–3491, 2015.

[11] John E. Hopcroft and Jeffrey D. Ullman. *Formal Languages and Their Relation to Automata*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1969.

[12] Armand Joulin and Tomas Mikolov. Inferring algorithmic patterns with stack-augmented recurrent nets. In *Advances in neural information processing systems*, pages 190–198, 2015.

[13] Matteo Manica, Ali Oskooei, Jannis Born, Vigneshwari Subramanian, Julio Saez-Rodriguez, and María Rodríguez Martínez. Toward explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders. *Molecular Pharmaceutics*, 2019. doi: 10.1021/acs.molpharmaceut.9b00520. PMID: 31618586.

[14] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in neural information processing systems*, pages 3391–3401, 2017.

[15] Liwei Wang, Alexander Schwing, and Svetlana Lazebnik. Diverse and accurate image description using a variational auto-encoder with an additive gaussian encoding space. In *Advances in Neural Information Processing Systems*, pages 5756–5766, 2017.

[16] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992.

[17] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, Joshua M Stuart, Cancer Genome Atlas Research Network, et al. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10): 1113, 2013.

[18] Ali Oskooei, Jannis Born, Matteo Manica, Vigneshwari Subramanian, et al. Paccmann: Prediction of anticancer compound sensitivity with multi-modal attention-based neural networks. *arXiv preprint arXiv:1811.06802 (presented at NeurIPS 2018 workshop: "Workshop for Machine Learning for Molecules and Materials")*, 2018.

[19] Ali Oskooei, Matteo Manica, Roland Mathis, and María Rodríguez Martínez. Network-based biased tree ensembles (NetBiTE) for drug sensitivity prediction and drug sensitivity biomarker identification in cancer. *arXiv preprint arXiv:1808.06603 (presented at ICML 2019 workshop: "Workshop on Computational Biology")*, 2018.

[20] Anna Gaulton, Anne Hersey, Michał Nowotka, A Patrícia Bento, Jon Chambers, David Mendez, Prudence Mutowo, Francis Atkinson, Louisa J Bellis, Elena Cibrián-Uhalte, et al. The chembl database in 2017. *Nucleic acids research*, 45(D1):D945–D954, 2016.

[21] Wanjuan Yang, Jorge Soares, Patricia Greninger, Edelman, et al. Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research*, 41(D1):D955–D961, 2012.

[22] Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483 (7391):603, 2012.

[23] Magdalena Gorska, Alicja Kuban-Jankowska, Ryszard Milczarek, and Michal Wozniak. Nitro-oxidative stress is involved in anticancer activity of 17β-estradiol derivative in neuroblastoma cells. *Anticancer research*, 36:1693–8, 04 2016.

[24] Juan Lao, Julia Madani, Teresa Puértolas, María Álvarez, Alba Hernández, Roberto Pazo-Cid, Ángel Artal, and Antonio Antón Torres. Liposomal doxorubicin in the treatment of breast cancer patients: a review. *Journal of drug delivery*, 2013, 2013.

[25] Radhika Poojari. Embelin–a drug of antiquity: shifting the paradigm towards modern medicine. *Expert opinion on investigational drugs*, 23(3):427–444, 2014.

[26] Yow-Jyun Cheng, Hang-Shiang Jiang, Shih-Lan Hsu, Li-Chiung Lin, Chieh-Liang Wu, Vithal K Ghanta, and Chi-Mei Hsueh. Xiap-mediated protection of h460 lung cancer cells against cisplatin. *European journal of pharmacology*, 627(1-3):75–84, 2010.

[27] Xiaotun Zhang, Wenbin Wang, Lawrence D. True, Robert L. Vessella, and Thomas K. Takayama. Protease-activated receptor-1 is upregulated in reactive stroma of primary prostate cancer and bone metastasis. *The Prostate*, 69(7):727–736, 2009. doi: 10.1002/pros.20920.

[28] Alex Zhavoronkov, Yan A Ivanenkov, Alex Aliper, Mark S Veselov, Vladimir A Aladinskiy, Anastasiya V Aladinskaya, Victor A Terentiev, Daniil A Polykovskiy, Maksim D Kuznetsov, Arip Asadulaev, et al. Deep learning enables rapid identification of potent ddr1 kinase inhibitors. *Nature biotechnology*, 37(9):1038–1040, 2019.