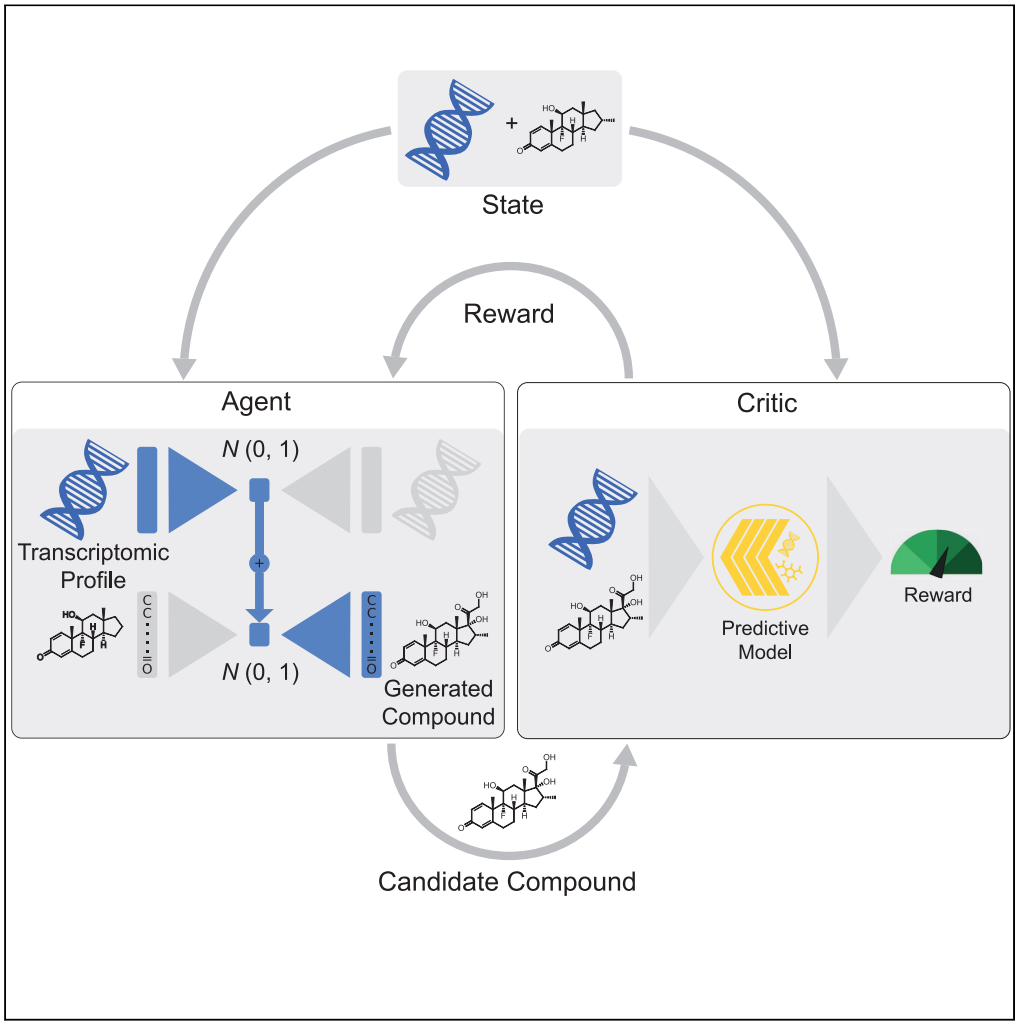
iScience

CellPress
OPEN ACCESS

Article

PaccMann^{RL}: De novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning



Jannis Born, Matteo Manica, Ali Oskooei, Joris Cadow, Greta Markert, María Rodríguez Martínez

jab@zurich.ibm.com (J.B.)
tte@zurich.ibm.com (M.M.)
mrm@zurich.ibm.com (M.R.M.)

Highlights

A conditional generative model for de novo design of anticancer hit molecules is devised

Drug sensitivity and toxicity models steer the molecule design via reinforcement learning

Molecules are designed to target individual transcriptomic profiles of cell lines

Targeted, hit-like molecules are generated more frequently, even for unseen cell lines

In silico, the molecules exhibit similar physicochemical properties to real cancer drugs

Born et al., iScience 24, 102269
April 23, 2021
https://doi.org/10.1016/j.isci.2021.102269

iScience

CellPress
OPEN ACCESS

Article

PaccMann^{RL}: De novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning

Jannis Born,[1,2,4,5,*] Matteo Manica,[1,4,*] Ali Oskooei,[1,4] Joris Cadow,[1] Greta Markert,[1,3] and María Rodríguez Martínez[1,*]

SUMMARY

With the advent of deep generative models in computational chemistry, *in-silico* drug design is undergoing an unprecedented transformation. Although deep learning approaches have shown potential in generating compounds with desired chemical properties, they disregard the cellular environment of target diseases. Bridging systems biology and drug design, we present a reinforcement learning method for de novo molecular design from gene expression profiles. We construct a hybrid Variational Autoencoder that tailors molecules to target-specific transcriptomic profiles, using an anticancer drug sensitivity prediction model (PaccMann) as reward function. Without incorporating information about anticancer drugs, the molecule generation is biased toward compounds with high predicted efficacy against cell lines or cancer types. The generation can be further refined by subsidiary constraints such as toxicity. Our cancer-type-specific candidate drugs are similar to cancer drugs in drug-likeness, synthesizability, and solubility and frequently exhibit the highest structural similarity to compounds with known efficacy against these cancer types.

INTRODUCTION

Drug discovery

Eroom's law describes the observation that the productivity of the drug discovery pipeline, as measured by the number of FDA-approved drugs per billion US dollar invested, has been halved every 9 years since the 1950s (Scannell et al., 2012). Indeed, only a minimal portion of all synthesized drug candidates obtain market approval (less than 0.01%), with an estimated 10–15 years until market release and costs that range between one (Scannell et al., 2012) and three billion dollars per drug (DiMasi et al., 2016). 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 potent candidate drugs. In addition to the initial wet-lab validations, the discovery pipeline involves a sequential process that builds upon high-throughput screenings, ADMET-assessments, 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. The problem's linchpin is on how to improve the exploration and navigation through the chemical space that has been estimated to contain $\sim 10^{30}$–10^{60} drug-like molecules with bioactive properties (Polishchuk et al., 2013).

Related work

Deep learning methods have recently gained popularity to aid drug discovery (Chen et al., 2018) and many have demonstrated the feasibility of in silico design of novel candidate compounds with desired chemical properties (Popova et al., 2018; Gomez-Bombarelli et al., 2018; You et al., 2018). In all of these models, the generative process is controlled by a structurally driven evaluator (or critic) that biases the generation of a chemical to satisfy the required chemical structural properties. Although very effective in generating compounds with desired chemical properties, these methods disregard system-level information, e.g. 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 a lack in efficacy against the specific disease of interest and off-target cytotoxicity (Wehling, 2009), calling to bridge systems biology closer with drug discovery. Related methodology has been used for protein-targeting

[1]IBM Research Europe, 8803 Rüschlikon, Switzerland

[2]Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland

[3]Department of Chemistry and Applied Biosciences, ETH Zurich, 8093 Zürich, Switzerland

[4]These authors contributed equally

[5]Lead contact

*Correspondence: jab@zurich.ibm.com (J.B.), tte@zurich.ibm.com (M.M.), mrm@zurich.ibm.com (M.R.M.)

https://doi.org/10.1016/j.isci.2021.102269

de novo generation (Zhavoronkov et al., 2019; Aumentado-Armstrong, 2018; Grechishnikova, 2021; Chenthamarakshan et al., 2020; Skalic et al., 2019; Krishnan et al., 2021). These contributions attempt to utilize deep learning methods for de novo design of compounds to specifically target a protein that has been implicated in tumor proliferation or treatment response (e.g. gene-knockout study). For example, the study by Zhavoronkov et al. (2019) curated and utilized, among others, patent data and several datasets about molecules (unspecific bioactive compounds, kinase inhibitors, DDR1 kinase inhibitors, molecules targeting non-kinase targets) specifically to develop DDR1 inhibitors. They synthesized and tested six drug candidates in cell assays. Two of them were found to be active, and one was even successfully validated in animal models. Envisioning a precision or even personalized medicine perspective, identifying protein targets is challenging, whereas sequencing and omics data are straightforward to gather.

Very recently, Méndez-Lucio et al. (2020) proposed a method for the de novo design of molecules against desired targets, represented by the gene expression signatures of knocked-out (suspected) targets. Notably, 97% of all anticancer candidate drugs fail in clinical trials and never receive FDA approval, questioning the current approaches of target identification for the discovery of pharmaceuticals (Wong et al., 2019). Taking ten drug-indication pairs from ongoing clinical trials, Lin et al. (2019) found that the proposed mechanism of action (MOA) for *all* of them were incorrect; knocking out the target genes did not ever hamper cancer fitness. Although the wrong target genes were identified through RNA interference with siRNA, seemingly silencing essential off-target genes, all drugs retained their anticancer effect through target-independent mechanisms. The fact that off-target cytotoxicity is a common MOA of anticancer drugs in clinical trials corroborates the need to to scrutinize current lead compound discovery strategies and calls to develop novel methodology with unconventional approaches. It is for this reason that we herein propose a novel framework to generate lead compound candidates solely based on a tumor's metabolic signature, as opposed to attempting to target a specific protein or incorporating information about potential targets directly into the design process. Here, we guide the learning process solely by transcriptomic profiles of cancer cells, which thus act as metabolic signatures. Transcriptomic data have been successfully used for de novo drug identification (Verbist et al., 2015; De Wolf et al., 2018) and has been advocated for a pivotal role in the future of drug discovery (Dopazo, 2014). Related work has addressed the generation of anticancer candidate drugs by conditionally sampling from an IC50 vector (Joo et al., 2020).

Our contribution

We present a novel framework for molecule generation based on deep generative models and reinforcement learning that, for the first time, enables the generation of molecules while taking into account the disease context encoded in the form of gene expression profile (GEP) data (for a graphical illustration see Figure 1A). Our framework is depicted in Figure 1B and consists of a conditional molecule generator (embodied by two separate Variational Autoencoders) and a critic module that evaluates the efficacy of proposed compounds on the target profile (see Figure 1D). The training procedure is split into two stages. In the first stage, the models are trained independently; one VAE is trained on gene expression data (in the following called *profile VAE* or just PVAE) from TCGA (Weinstein et al., 2013), and another VAE (in the following called *SMILES VAE* or just SVAE) is trained on bioactive small molecules from ChEMBL (Bento et al., 2013) (see Figure 1C). As a critic, we use `PaccMann`, a multimodal drug sensitivity prediction model developed and validated in our previous work (Manica et al., 2019; Cadow et al., 2020). In the second stage, the encoder of the profile VAE is combined with the decoder of the molecule VAE and exposed to a joint retraining that is optimized using a policy gradient regime 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 that is characteristic for a cancer site, a patient subgroup, or even an individual. By *efficacy*, we refer to predict cellular IC50 (i.e. the micromolar concentration necessary to inhibit 50% of the cells) as opposed to e.g. enzymatic IC50. This efficacy is a joint property of a drug-cell-pair, as treatment response to a compound heavily varies depending on the tumor's genomic and transcriptomic makeup (Geeleher et al., 2016). In this work, we focus on profile-specific compound generation and optimize the generator with IC50 as critic, but we would like to note that our framework can be extended to more complex reward functions and include a case study on the concurrent optimization of (predicted) drug efficacy and toxicity.

RESULTS

Pretraining profile VAE and SMILES VAE

In the first phase of training, the two components depicted in Figure 1C were trained independently. The profile VAE consisted of a set of stacked dense layers and was trained as a denoizing VAE to enhance

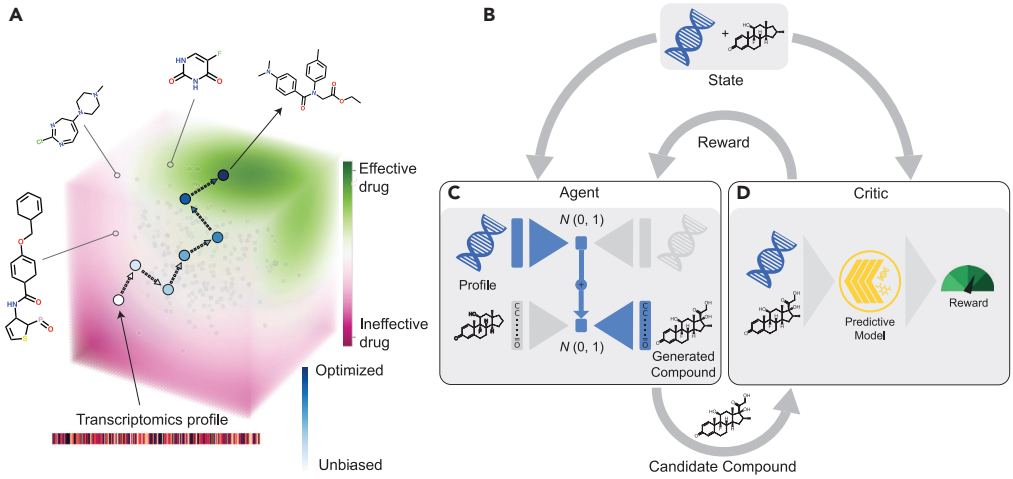


Figure 1. The proposed framework for anticancer compound design against specific cancer profiles

(A) Conceptually, the model performs a guided walk through the chemical space in order to find effective compounds. Starting from an unbiased molecule generator (trained only on a dataset of bioactive compounds without any information about cancer), compounds are sampled and screened in silico against the transcriptomic profiles of interest. The outcome of the screening guides the generator toward sampling from manifolds with more effective compounds.

(B–D) (B) The training process depicted in more detail. The conditional compound generator (called "agent") is embodied through two initially separate VAEs. The compound generation starts with a biomolecular profile of interest e.g. a transcriptomic profile. Through a pretrained omics VAE, the profile is encoded into the latent space of gene expression profiles. The latent representation of the profile is decoded through the molecular decoder of a separately pretrained molecule VAE to produce a candidate compound (see C). This generative process can optionally be "primed" through encoding a known, effective compound or a functional group with the molecular encoder. The proposed compound is then evaluated by a critic. Our critic is represented by a multimodal drug sensitivity prediction model that ingests the compound and the target profile of interest (see D). The IC50 efficacy, as predicted by the critic, is interpreted as reward and is subject to maximization during the RL-based optimization. Over the course of training, the generator will thus learn to produce candidate compounds with higher and higher efficacy. <START> is the start and <END> is the end token.

generalization abilities. The purpose of the PVAE was to find a lower dimensional representation of the cell profiles that maintains structural similarity and later allows a fusion with the latent representation of molecules. The encoder of the PVAE learned to meaningfully embed gene expression profiles (bulk RNA-Seq from TCGA (Weinstein et al., 2013)) into a latent space, such that the decoder could reconstruct the profiles, but also generate novel, seemingly realistic gene expression profiles (GEP). In alignment with the reported consensus between transcriptomic data in TCGA and cancer cell line databases (Ghandi et al., 2019), we found the distributions of GEPs in GDSC (Yang et al., 2012) and TCGA to be sufficiently similar to justify our decision to perform the PVAE pretraining on ~10k TCGA samples, whereas the reinforcement learning (RL) optimization was performed on GDSC cell lines (compare Figure S2 in supplementary material S2).

The SMILES VAE was pretrained for 10 epochs on ~1.4 million structures from ChEMBL (Bento et al., 2013) (for details see transparent methods). Both encoder and decoder consisted of stack-augmented gated-recurrent units as used in Popova et al. (2018). The purpose of the SVAE was to learn the syntax of SMILES and general semantics about bioactive compounds. The novelty and diversity of the generated molecules were validated by sampling 10,000 molecules through decoding random points from the latent space. About 96.2% of the 10,000 generated molecules were valid molecular structures (surpassing the results of Popova et al. (2018) who used the same stack-augmented GRUs and reported 95% SMILES validity). Moreover, 99.72% of the valid molecules were unique across the 10,000 generations, and the novelty was 1, i.e. none of the generated compounds was present in the training dataset. Comparing the Tanimoto similarity (Tanimoto, 1958) of the ECFP molecular fingerprints (Rogers and Hahn, 2010) of 1,000 generated molecules with the training and test data from ChEMBL, we found that the vast majority had a Tanimoto similarity (τ), between 0.2 and 0.6 (on average 0.41 $\pm$ 0.1 for training and 0.38 $\pm$ 0.08 for testing molecules), suggesting that our model learned to propose novel molecular structures from the chemical space. In addition, an interactive visualization of the chemical space with Faerun (Probst and Reymond, 2018) shows the

TMAP (Probst and Reymond, 2020) (i.e., an algorithm to represent high-dimensional data as minimum spanning trees) of ChEMBL and generated compounds through the TMAP algorithm. The generated molecules mix well with the training molecules into the chemical space. For detailed results of both the PVAE and the SVAE, see the supplementary material S3.

Omic-specific compound generation (RL training)

Here, we present the results of our molecule generator conditioned on gene expression profiles of cancer subtypes.

As a proof of concept, we show results for cancer in four different sites: breast (carcinoma), lung (carcinoma), prostate (carcinoma), and autonomic ganglia (neuroblastoma). The conditional generator was initialized as the SVAE, i.e. sampling from the unbiased generator yielded random molecules from the chemical space as learned from the ChEMBL data. For the evaluation, all generated compounds with a predicted IC50 value below 1 μM (i.e. pIC50 > 6) were considered to be *effective*. Moreover, within each cancer type (or site), 80% of the cell lines (breast: 50, lung: 169, prostate: 7, autonomic ganglia: 56) were considered as training cell lines and used to optimize the parameters of the conditional generator. We observed that over time the generator learned to produce more molecules with high predicted efficacy according to the critic. To test both the generalization abilities and whether the generator actually utilized the omics-profile for the generation, we used the remaining 20% of cell lines to verify whether conditioning the generator on unseen cell lines of the same site also leads to compounds with low IC50. As presented in Figure 2 (left column), our model learned to produce compounds with lower IC50 values, for unseen cell lines from the given cancer site. In other words, the IC50 distribution of candidate compounds proposed by the generative model were successfully shifted toward higher efficacy (lower IC50). The baseline model corresponds to the pretrained SVAE from which $n = 500$ molecules were randomly sampled. In all four cases, a significant portion (between 16% and 57%) of molecules generated from the optimized model were assigned an IC50 value below 1μM, whereas only 2%–5% of the candidates generated by the baseline model (i.e. the SVAE) were classified as effective. Moreover, in all cases the generator maintained an almost equal SMILES validity (84%–93%) compared with the baseline, much higher than what Méndez-Lucio et al. (2020) reported based on gene expression (8%–9%). The second column of Figure 2 shows generated molecules that are predicted to be effective against unseen cell lines from the respective cancer site. As opposed to the personalized regime in the second column, the third column of Figure 2 showcases a precision medicine regime. Here, novel molecules were designed specifically for each cancer site, i.e. a single, characteristic GEP. In all cases, the model generated compounds that exhibited high efficacy against the average cellular profile of the target site while maintaining efficacy against the majority of individual cell lines for that site. We have thus formulated a novel problem, namely how to drive a molecular generative model to produce molecules with low predicted efficacy against an omics profile of interest. To the best of our knowledge, this problem has not been tackled before, exacerbating a comparison to other works. In the most similar work, Méndez-Lucio et al. (2020) proposed a model that produces hit-like molecules to *induce* a desired gene expression profile.

Investigation of nearest neighbors

For a more quantitative assessment, the last column of Figure 2 compares the four cancer-type-specific candidate compounds with one of their top-3 neighbors using the Tanimoto similarity score, τ, from several hundreds of existing anticancer compounds. It is well known that the Tanimoto similarity across compounds is highly correlated with their induced sensitivity patterns on cancer cell lines (Shivakumar and Krauthammer, 2009). The candidate compound proposed against breast cancer (Figure 2 second row, third column) resembles a collection of fused sugarlike moieties and has doxorubicin, a commonly used chemotherapeutical against breast cancer (Lao et al., 2013), as one of the top-3 nearest neighbors. The generated compound against lung cancer (Figure 2, third row, third column) presents similarities to embelin, an existing anticancer compound from the GDSC database. Comparing the two structures, it is evident that the generated compound and embelin share a long carbon chain and a single six-membered fully carbonic ring. Embelin was tested against 965 cell lines from GDSC/CCLE from which the highest reported efficacy is against a lung cell line (`NT2-D1`). Embelin is also known to be the only known non-peptide inhibitor of XIAP (Poojari, 2014), a protein that plays an important role in lung cancer development (Cheng et al., 2010). 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

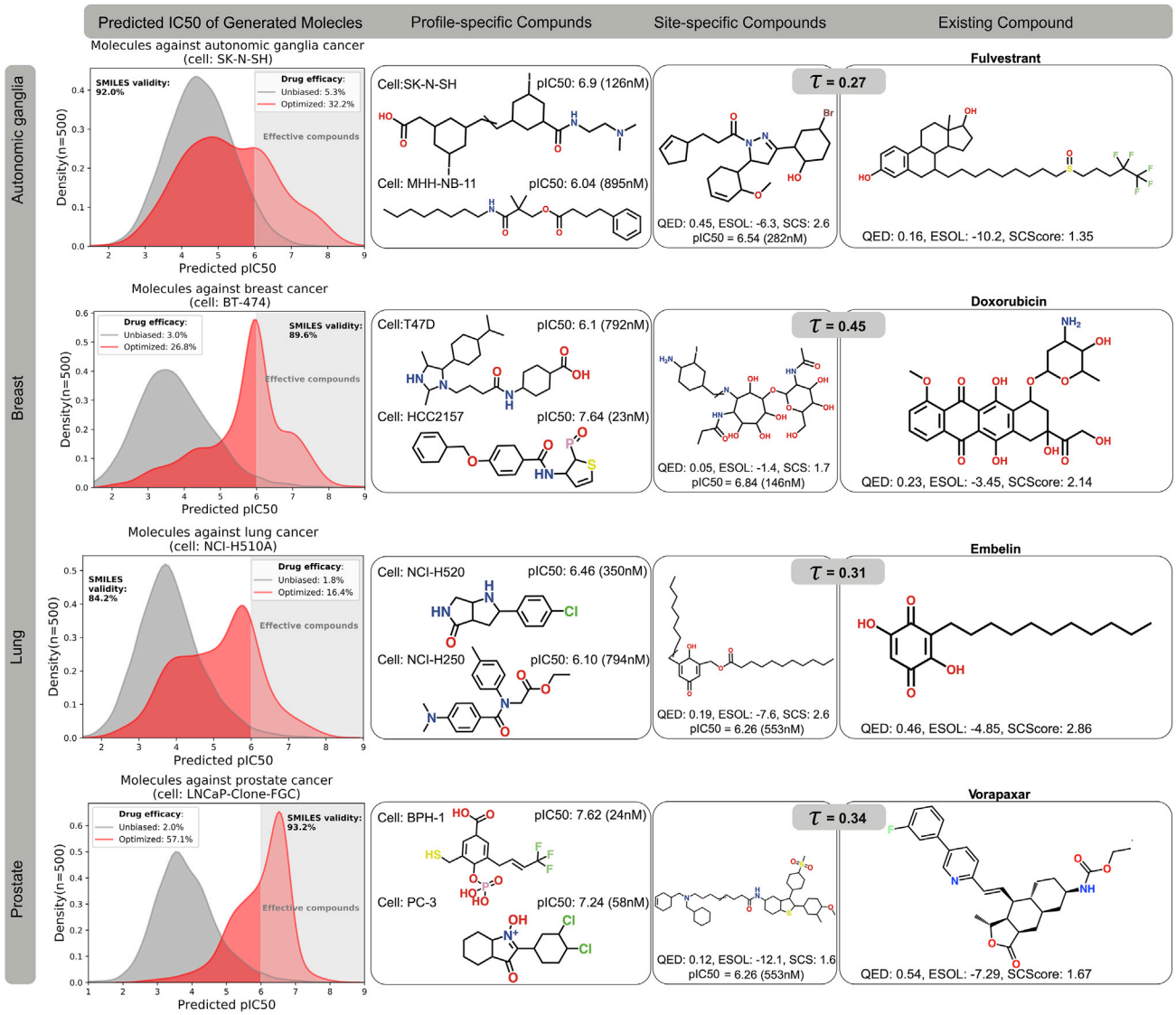


Figure 2. Sample results for profile-driven model optimization and anticancer compound generation

Each row illustrates the results of training the RL pipeline on cell lines from a specific cancer type: autonomic ganglia, breast, lung, and prostate cancer. The first column compares the distributions of pIC50 predictions given by the critic model for a set of n = 500 drug candidates generated with RL optimization and a set of 9,000 molecules without RL optimization. As demonstrated by the density plots, the RL optimization process leads to candidate compounds with a higher mean pIC50 for the target cancer, highlighting the successful optimization of the generative model toward the design of more effective compounds. The second column presents candidate compounds with a high predicted efficacy against a particular cell line that was not seen during training. The third column showcases generated compounds that were optimized to be effective against all 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 between `RDKit` fingerprints as well as three chemical scores crucial in drug design, namely druglikeness (QED, 0 worst, 1 best), synthetic complexity (SCS or SCScore, 1 best, 5 worst), and solubility (ESOL, given in *M*/*L*). The crossed carbon double bond in some molecule means that the E-Z configuration is undefined.

various types of cancer, including prostate (Zhang et al., 2009). Lastly, the third closest neighbor of the generated compound against neuroblastoma (Figure 2 first row, third column) is fulvestrant, an antagonist/modulator of ERα that has recently been proposed as a novel anticancer agent for neuroblastoma (Gorska et al., 2016). The predicted pIC50 activity profile of fulvestrant and our candidate drug are highly correlated across all cell lines (ρ = 0.88), indicating that they may exhibit similar pharmacological properties. Similar observations are made for the lung and prostate cancer candidates and their neighbors embelin and vorapaxar (ρ = 0.55 and ρ = 0.69).

To summarize, for all four investigated cancer types, the proposed compounds showed the highest structural similarity to anticancer drugs that are, for each specific cancer type, either (1) already FDA approved (breast), (2) known inhibitors of relevant targets (lung, prostate), or (3) have been advocated for (neuroblastoma). This result is remarkable, especially as the generator was never exposed to any anticancer compounds. Indeed, only the critic had seen two out of the four compounds during training, highlighting that the generator has learned some structural characteristics that make a compound efficacious against a particular cancer type, according to our critique.

In the above comparisons, the search space was restricted to compounds with known anticancer properties. To investigate whether the proposed compounds generally had a higher similarity to drugs associated with cancer, we carried out a comparison with compounds from a broader pool of chemicals, namely ChEMBL (Bento et al., 2013), a database of >1.5 million bioactive molecules with drug-like properties. The nearest neighbor (τ = 0.54) of our breast cancer compound in the ChEMBL database is CHEMBL1093122, a conjugate of plumbagin and phenyl-2-amino-1-thioglucoside that inhibits the synthesis of mycothiol (Gammon et al., 2010). Plumbagin itself and many of its derivatives are widely studied anti-breast cancer compounds (Zhang et al., 2016; Kawiak et al., 2017; Dandawate et al., 2014). For our lung cancer compound, the nearest neighbor (τ = 0.48) is polyoxyethylene dioleate, a surfactant that has been patented for the treatment of eight types of cancer, including three types of lung cancer (small lung cell cancer, lung adenocarcinoma, and metastatic lung cancer, Girsh (2007)). It is also utilized in targeted drug delivery systems for drug-resistant lung cancer (Kaur et al., 2016). The nearest neighbor (τ = 0.31) of the prostate cancer compound is Clinolamide (or Linolexamide), which is included in a patent of diagnostic and/or therapeutically active compounds for several types of cancer, including prostate cancer (Klaveness et al., 2004). The nearest neighbor (τ = 0.35) of the proposed neuroblastoma compound is NSC-715466. `NSC-715466` has been evaluated for anticancer effects in a recent release of the NCI-60 database (Shoemaker, 2006) and inhibits cancer cell growth by 65% $\pm$ 15% across all tested cell lines, with a below-average inhibition for cancer in the central nervous system (57% $\pm$ 9%). Regarding its efficacy, it only falls in the 51st percentile of all 53,217 compounds tested in NCI-60, which presumably prevented further investigations. The four discussed ChEMBL compounds as well as the analysis of `NSC-715466` can be found in Figure S4 in supplementary material S4.

It is promising to observe that the molecules with the highest Tanimoto similarity to our compounds are associated with cancer (some even to the specific types of cancer our compounds were optimized for) even when using a larger database of bioactive compounds. However, it is worth keeping in mind that a high Tanimoto score to a known cancer drug is not necessary for anticancer drug efficacy. Oftentimes, cancer drugs used for the same cancer type or even drugs sharing the same mechanism of action exhibit low Tanimoto similarity, e.g. TKIs used for NSCLC such as crizotinib and erlotinib, τ = 0.11. Across all anticancer compounds in GDSC and CCLE databases, we note that the average Tanimoto similarity (τ = 0.149 $\pm$ 0.05) is not much lower than the average similarity of two compounds of a given site (τ = 0.154 $\pm$ 0.06). For the following results and discussion, the anticancer compounds from GDSC/CCLE were associated with the site where they had the highest average IC50 efficacy.

To better understand whether the generated drugs mimic the space of cancer-specific anticancer drugs, Figure 3 shows visualizations of real and generated cancer drugs for one specific cancer type, using kernel PCA based on Tanimoto similarity (Schölkopf et al., 1998). In addition to the class belongings, the plots also depict the QED score (Bickerton et al., 2012), a quantitative estimate of drug-likeness (0 worst, 1 best), and SCScore (Coley et al., 2018), an estimated score of synthetic complexity (1 best, 5 worst). The fact that the generated molecules are well intermingled with the real drugs suggests that the generator proposes diversified structures that mimic some properties of anticancer compounds. It is also curious to see that several real drugs have low QED and/or high synthetic complexity scores (the same holds true for the generated molecules). Moreover, we provide interactive TMAP visualizations of the site-specific generated compounds (links in the availability section).

Chemical properties of generated molecules

In this work, the conditional generator is trained using PaccMann as sole critic. However, besides inhibitory efficacy, there is a myriad of properties of a candidate drug that crucially influence its potential for becoming an anticancer compound.

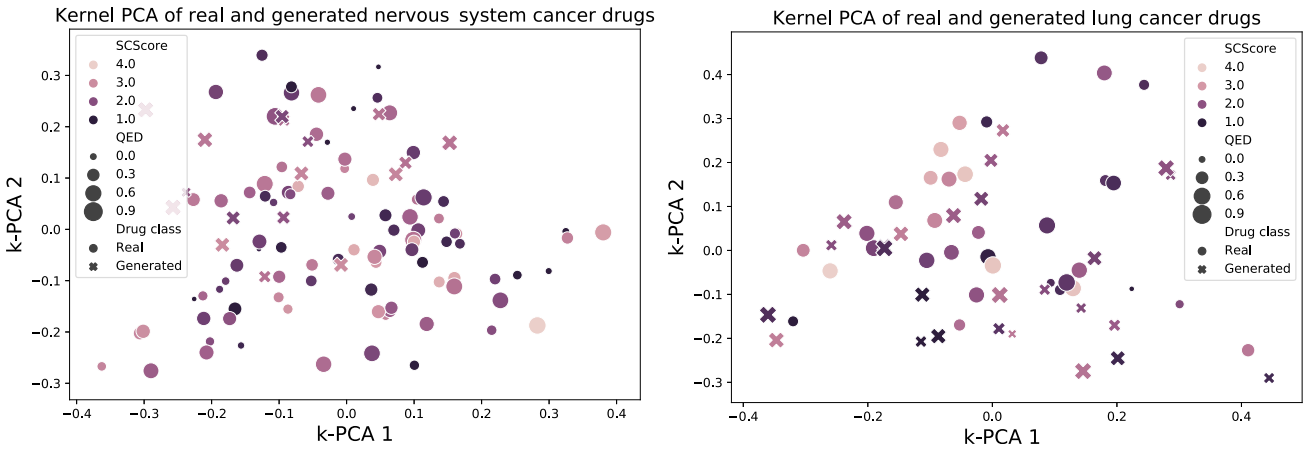


Figure 3. Visualization of generated molecules and real anticancer drugs

A kernel PCA of real and generated molecules based on Tanimoto similarity. The size of the points denotes the QED score, whereas the coloring represents the synthetic complexity score (SCScore). Overall, both generated and existing molecules are heterogeneously distributed in the 2D projection and do not form clear clusters.

Some of these can be approximated in silico, e.g. water solubility (ESOL, Delaney (2004)), drug-likeness (QED, Bickerton et al. (2012)), and synthesizability (SCScore, Coley et al. (2018)). Figure 4 gives an overview of the distribution of QED, ESOL, and SCScore for sets of (1) known anticancer compounds (blue), (2) molecules from ChEMBL (orange), (3) compounds generated by the SVAE (red), and (4) compounds proposed by the conditional generator (green). Despite the fact that none of these properties were explicitly optimized, comparing the distributions reveals a good agreement overall. Interestingly, anticancer drugs exhibit, compared with the ChEMBL compounds, much less drug-like properties (lower QED) on average and seem to be more synthesizable (lower SCScore). This tendency of anticancer drugs toward synthetically less complex structures is likely to result from the high attrition rate in clinical trials and the corresponding cost reduction policies. It is also encouraging to see, on average, that the unbiased generator (SVAE) generates molecules with more desired properties compared with the data used for training (ChEMBL compounds have lower QED and higher SCScore). Moreover, the cancer drugs have, on average, a significantly lower QED than the other three sets in Figure 4 (0.45 ± 0.2 with >10% of GDSC/CCLE drugs even having a QED <0.2, whereas it is 0.55 for the other three sets). Indeed the QED scores of the other three sets were so similar that we failed to reject the null hypothesis that the QED scores of these three sets are from different distributions (Kruskal-Wallis test, $\alpha = 0.05$). Regarding synthetic complexity, both the unbiased and the biased generator fail to produce molecules with SCScores as low as the anticancer drugs (MWU, $p < 0.01$), but they produce structures that are estimated to be less complex than the ChEMBL molecules (MWU, $p < 0.01$). Overall, we observe that the biased generator produces molecules with less desired properties than the unbiased generator (SVAE). This is to be expected because the unbiased generator was pretrained to mimic the data from ChEMBL, whereas no explicit optimization of chemical scores was performed during the RL optimization. For one generated compound, we exemplarily show a possible synthesis route that was assigned a high confidence score by the retrosynthesis model (Schwaller et al., 2020) and consists of four reactions and ten commercially available reactants (see Figure S6 in supplementary material S6 appendix (S8)).

Savjani et al. (2012) reported that 40% of novel chemicals cause practical problems due to insolubility. Water solubility remains challenging to approximate in silico (Sorkun et al., 2019) and thus we treat the good agreement in the ESOL scores (see Figure 4 middle panel) with caution.

Finally, we would like to point out that utilizing an IC50 drug sensitivity prediction model as sole critic limits the performance of the framework, as the expressive power of the conditional generator is inherently "upper bounded" by the predictive power of the critic. Crucially, due to a lack of available data, the critic was only trained on the cancer candidate compounds in GDSC (Yang et al., 2012) but not on compounds without any inhibitory efficacy against cell lines. Notably, however, GDSC contains many samples with IC50 in millimolar scale, leading the critic to cover vast ranges of irrelevant molecules with $3 < pIC50 < 6$. We additionally verified the generalization capabilities of the critic by comparing the predicted efficacy

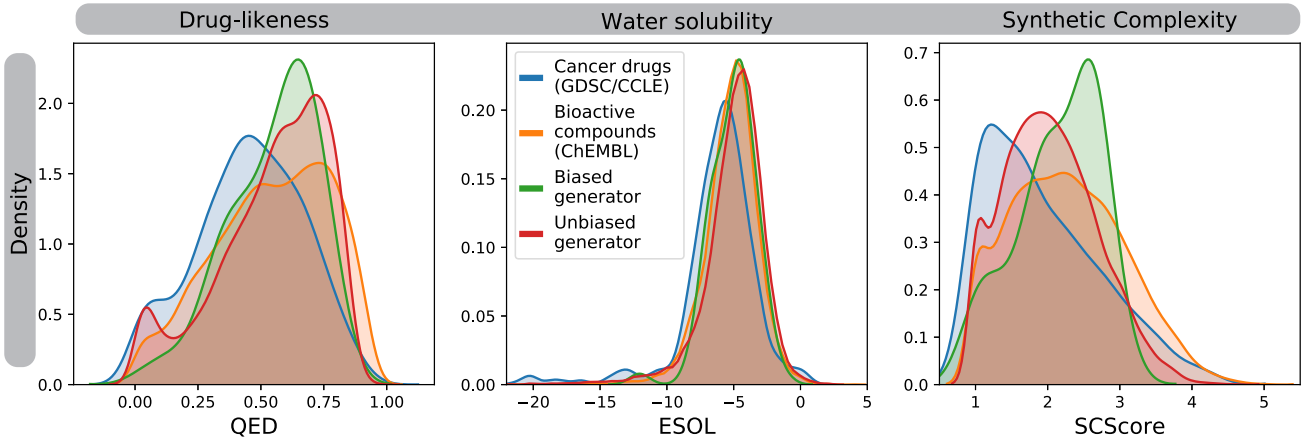


Figure 4. Comparison of chemical scores for real drugs in the GDSC and CCLE databases versus our generated compounds

We compared three chemical scores for druglikeness as assessed by QED score (0 worst, 1 best), for solubility as assessed via ESOL, given in log(*M*/*L*) (most drugs have a solubility between −8 and −2), and for synthetic accessibility as assessed by SAS (1 best, 10 worst). These three scores are computed for the panel of known anticancer drugs, bioactive molecules from ChEMBL, and molecules generated before (red) and after (green) RL optimization.

of cancer drugs (most of them were seen during training) and a "negative" set of molecules from ChEMBL across all 965 cell lines from GDSC. The results can be found in Figure S5 in supplementary material S5 and show that although 15.2% of the virtual drug screenings with anticancer compounds were positive (IC50<1μ mol), only 7.7% of all the ChEMBL molecules showed potential anticancer effects. Moreover, the generated anticancer compounds were found to have a significantly higher Tanimoto similarity to anticancer drugs than to either ChEMBL molecules ($p < 0.01$, one-sided MWU) or molecules generated without the RL optimization, i.e. from the SVAE ($p < 0.01$, one-sided MWU). These two results are encouraging and suggest that PaccMann can seemingly drive the molecule generation away from ordinary bioactive compounds such as in ChEMBL, more toward mimicking the properties of anticancer drugs.

Toxicity: Case study on multi-objective optimization

In the previous sections, the RL optimization of the molecular generation was guided using the predicted IC50 on a given cell profile as sole reward. To investigate whether PaccMann^{RL} is able to generate desired molecules in a multi-objective optimization setting, we performed RL optimization using an adjusted reward function that incorporates (1) low IC50 against a given omic profile (as in all other experiments), (2) low environmental toxicity, and (3) low adverse drug reactions, a measure of high-level phenotypic side-effect observations (for details see supplementary material S1.4). Similarly to IC50, toxicity and adverse drug reactions were also evaluated with predictive neural networks. The toxicity model was pretrained on Tox21 (Huang et al., 2016) (~8k molecules measured across 12 toxicity endpoints) and the adverse reaction model was pretrained on SIDER (~1.5k marketed drugs with 27 classes of reported adverse drug reactions). We note that the toxicity predictor achieved an ROC-AUC of 0.877 on Tox21 (surpassing the previous state-of-the-art) and an ROC-AUC of 0.835 on SIDER (Markert et al., 2020). For details on these predictive models see Markert et al. (2020).

Figure 5 displays a collection of molecules that were generated against the NCI-H520 cell line (a human lung squamous cancer cell line) during an RL optimization that was restricted to lung cancer cell lines (similar to Figure 2, third row). Notably, the `NCI-H520` cell line was in the test set for the RL optimization. The adjusted reward function (Equation 5) used low IC50 as primary and low toxicity scores as secondary objective. All displayed molecules fulfilled the multi-objective, i.e., they were predicted to be active on `NCI-H520` (IC50<1μM) while not being predicted toxic in any of the 12 Tox21 toxicity assays. Furthermore, we computed Tanimoto similarities of the molecules shown in Figure 5 with approved cancer drugs, according to the National Cancer Institute. For molecules **1** and **2** we obtained Tanimoto coefficients >0.45 for several drugs that are FDA approved for the use against lung cancer, such as vinorelbine, vincristine, vinblastine, irinotecan, topotecan, and alectinib (Wishart et al., 2018). Notably, across all investigated FDA-approved cancer drugs, molecule **1** (which was generated against `NCI-H520`, an NSCLC cell line) exhibited the highest Tanimoto similarity with vinorelbine, a targeted drug for NSCLC.

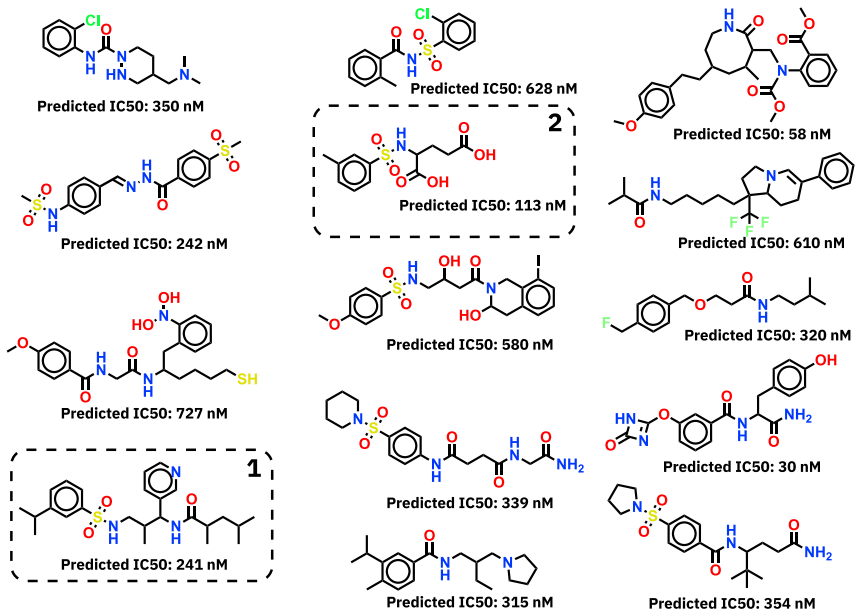


Figure 5. Case study on multi-objective optimization including toxicity

Molecules generated against the NCI-H520 cell-line with a multi-objective reward function based on a combination of low IC50 and low toxicity in Tox21 and SIDER tasks. For all fourteen molecules depicted, all Tox21 assay predictions are negative. Molecules 1 and 2 are further discussed in the text.

DISCUSSION

Summary

In this work we presented PaccMann^{RL}, a novel framework for molecular generation that enables us to condition on the transcriptomic profile of the target cell or cancer site. By using RL optimization we demonstrated that our proposed generative model is able to produce candidate compounds with high predicted efficacy (low IC50) against a given target profile, even if this profile was never seen during training. Notably, this was achieved despite the fact that the generator was never exposed to anticancer drugs explicitly but only pretrained on bioactive compounds from ChEMBL (the only component that was trained on drugs with known anticancer effects is the critic). Moreover, an analysis of the generated compounds for four different cancer types revealed that the predicted compounds share many structural similarities with known anticancer compounds for the same cancer types that the generation was optimized for. In a case study, we briefly investigated the ability of PaccMann^{RL} to generate molecules that fulfill multiple objectives (specifically high activity against a lung cancer cell line and low toxicity) and found many candidates with desired predicted properties.

Interpretation and future work

Although our results are a modest step toward disease-specific molecular generation, a lot of further optimization is needed. For instance, subsidiary properties of a candidate drug that determine its potential for becoming a successful anticancer compound (e.g., water solubility, drug-likeness, synthesizability or off-target cytotoxicity) are not directly optimized. However, despite not explicitly incorporating them into the reward function, we find that the produced molecules exhibit desired properties in terms of drug-likeness, water solubility, and ease of synthesis. One possible future endeavor is to incorporate information in the reward function not only about drug efficacy but also about other drug-relevant chemical properties; one such example was shown in the concurrent optimization of molecules with a low IC50 and low toxicity. An additional investigation toward that end might be a substructure search for toxicophores within the generated molecules. This could help to winnow the promising molecules from those that exhibit low IC50 due to cytotoxicity caused by a toxicophore. Another observation we made is that the stability of the conditional generation can diminish over RL training time and result in overly simplistic molecules, especially if the learning rate is set too high. However, after pretraining of the SVAE (and before starting the RL training), the generated molecules were found to be reasonably similar to the training molecules (bioactive compounds from ChEMBL). We, therefore, conceive that for more targeted applications, an improved chemical space could be obtained by fine-tuning the SVAE on, e.g., anticancer drugs or even bioactive molecules with a shared mechanism of action, like JAK inhibitors. If the set of possible target

molecules is small, SMILES augmentation can further be used to improve the quality of molecular generators (Arús-Pous et al., 2019). We also emphasize that the chemical space covered by the SVAE encompasses 86 distinct SMILES tokens, significantly more than in most related work and more than in landmark publications that produced de novo molecules with high efficacy *in vitro* and *in vivo* (Zhavoronkov et al., 2019).

The resulting multimodal objectives may be difficult to optimize due to possibly counteracting/interfering gradients. A possible approach to circumvent this challenge is potentially to use explicit compensation techniques (Yu et al., 2020) or defining gradient-free global objectives (Häse et al., 2018). Another challenge that needs to be overcome to improve the reliability and accuracy of the critic is the expected distributional differences between the data used for training (cancer cell lines) and the targeted data (human data from clinical trials). A possible approach that can be explored is the exploitation of transfer learning techniques, as suggested in Sharifi-Noghabi et al. (2020). Oftentimes, medical chemists do not start the drug design from scratch but from the scaffold of an approved drug. The goal of the scaffold hopping is to find a drug with similar effects (e.g. increased efficacy or reduced side effects). Although our framework enables users to incorporate prior knowledge into the design process by priming the latent code, we have not yet explored the full potential of this idea here. As the decoded molecule is not guaranteed to maintain similarity to the primer, the recently proposed "deep generative scaffold decorator" could be integrated into our framework to facilitate a more systematic exploration and the possibility of adding fragments to established drug scaffolds (Arús-Pous et al., 2020). An alternative future endeavor is to explore graph-based, instead of sequential, representations of molecules to directly generate a molecular graph from the context set using a conditional structure generation framework (Yang et al., 2019). For future work, one may conceive a generic framework where the molecule generation can be conditioned on possibly multimodal context information such as a (multi)omics profiles, a target protein, a primed drug, a drug scaffold, or a combination thereof. The resulting latent spaces, consisting of semantically distinct modalities, could then be jointly explored by machine learning techniques designed to operate on sets instead of fixed-length vectors, such as permutation-invariant operations (Zaheer et al., 2017).

Conclusion

Due to growing evidence of our insufficient understanding of the mechanisms of action of drug candidates even in clinical trials (Lin et al., 2019), the design of generative approaches that can, as shown here, bypass the need for a detailed characterization of drug targets and cytotoxicity mechanisms may be a novel, valuable approach within generative chemistry.

Limitations of the study

Without IC50 assays on the investigated cell lines, the *in vitro* efficacy of the proposed molecules remains unclear. Even in case of positive cell line drug sensitivity assays, these results may not translate to animal models due to numerous reasons (e.g., adverse reactions or drug delivery). Therefore, the potential for the presented molecules to become anticancer drugs cannot be conclusively assessed in this study. Furthermore, the quality of the molecular generator is inherently limited by the predictive power of the in silico IC50 evaluator and critically depends on how well this predictor generalizes across the chemical space. In addition, the proposed molecules are represented by SMILES sequences that are generated auto-regressively by using a stochastic sampling process, the common practice in SMILES generation. When the seed is not fixed, a point in the latent space does not necessarily map to a unique molecule.

Resource availability

Lead contact

Jannis Born, jab@zurich.ibm.com.

Materials availability

This study did not generate any new materials.

Data and code availability

Code: the source code is publicly available on: https://github.com/PaccMann/. For a detailed example see: https://github.com/PaccMann/paccmann_rl. To assess the critic, please see Manica et al. (2019).

Data: this study did not generate any new data and exclusively used publicly available databases. The transcriptomic data used to pretrain the Profile VAE and the molecular data used to train the SMILES VAE are

processed variants of TCGA (Weinstein et al., 2013) and CheMBL (Bento et al., 2013) and available via: https://ibm.box.com/v/paccmann-pytoda-data. The GDSC (Yang et al., 2012) cell profiles as well as the pretrained models can be found under the same link.

The interactive TMAP visualizations of the molecules generated by the (unbiased) SVAE can be found on https://paccmann.github.io/rl/unbiased.html. The TMAPs of the cancer-site-specific candidate compounds are accessible on https://paccmann.github.io/.

METHODS

All methods can be found in the accompanying transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102269.

ACKNOWLEDGMENTS

The authors thank Prof. Dr. Karsten Borgwardt and Dr. Maria Gabrani for continuous support and useful discussions on the manuscript. The project leading to this publication has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 826121. J.B. acknowledges financial support from the German Academic Exchange Service (DAAD).

AUTHOR CONTRIBUTIONS

J.B, M.M., and A.O. contributed equally to this work. Conceptualization: M.M, J.B., and A.O.; Methodology: J.B., M.M., and A.O.; Software: J.B., M.M., A.O., J.C., and G.M.; Investigation: J.B., A.O., M.M., and G.M.; Data curation: M.M.; Writing—Original Draft: A.O., J.B., and M.M; Writing—Review & Editing: J.B.; Visualization: J.B., M.M., and A.O.; Project Administration: M.R.M.; Funding Acquisition: M.R.M.;

DECLARATION OF INTERESTS

The authors declare that a patent associated to the methodology presented in this manuscript was granted (US20200365238A1).

INCLUSION AND DIVERSITY

While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: October 5, 2020
Revised: January 11, 2021
Accepted: March 1, 2021
Published: April 23, 2021

REFERENCES

Arús-Pous, J., Johansson, S.V., Prykhodko, O., Bjerrum, E.J., Tyrchan, C., Reymond, J.-L., Chen, H., and Engkvist, O. (2019). Randomized smiles strings improve the quality of molecular generative models. J. Cheminform. *11*, 1–13.

Arús-Pous, J., Patronov, A., Bjerrum, E.J., Tyrchan, C., Reymond, J.-L., Chen, H., and Engkvist, O. (2020). Smiles-based deep generative scaffold decorator for de-novo drug design. J. Cheminform. *12*, 1–18.

Aumentado-Armstrong, T. (2018). Latent Molecular Optimization for Targeted Therapeutic Design (arXiv), p. 1809.02032.

Bento, A.P., Gaulton, A., Hersey, A., Bellis, L.J., Chambers, J., Davies, M., Krüger, F.A., Light, Y., Mak, L., McGlinchey, S., et al. (2013). The ChEMBL bioactivity database: an update. Nucleic Acids Res. *42*, D1083–D1090.

Bickerton, G.R., Paolini, G.V., Besnard, J., Muresan, S., and Hopkins, A.L. (2012). Quantifying the chemical beauty of drugs. Nat. Chem. *4*, 90.

Cadow, J., Born, J., Manica, M., Oskooei, A., and Rodríguez Martínez, M. (2020). Paccmann: a web service for interpretable anticancer compound sensitivity prediction. Nucleic Acids Res. *48*, W502–W508.

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., and Blaschke, T. (2018). The rise of deep learning in drug discovery. Drug Discov. Today *23*, 1241–1250.

Cheng, Y.-J., Jiang, H.-S., Hsu, S.-L., Lin, L.-C., Wu, C.-L., Ghanta, V.K., and Hsueh, C.-M. (2010). Xiap-mediated protection of h460 lung cancer cells against cisplatin. Eur. J. Pharmacol. *627*, 75–84.

Chenthamarakshan, V., Das, P., Hoffman, S.C., Strobelt, H., Padhi, I., Lim, K.W., Hoover, B., Manica, M., Born, J., Laino, T., and Mojsilovic, A. (2020). Cogmol: target-specific and selective drug design for COVID-19 using deep generative models. Adv. Neural Inf. Process. Syst. *33*.

Coley, C.W., Rogers, L., Green, W.H., and Jensen, K.F. (2018). Scscore: synthetic complexity learned from a reaction corpus. J. Chem. Inf. Model. *58*, 252–261.

Dandawate, P., Ahmad, A., Deshpande, J., Swamy, K.V., Khan, E.M., Khetmalas, M., Padhye, S., and Sarkar, F. (2014). Anticancer phytochemical analogs 37: synthesis, characterization, molecular docking and cytotoxicity of novel plumbagin hydrazones against breast cancer cells. Bioorg. Med. Chem. Lett. *24*, 2900–2904.

De Wolf, H., Cougnaud, L., Van Hoorde, K., De Bondt, A., Wegner, J.K., Ceulemans, H., and Göhlmann, H. (2018). High-throughput gene expression profiles to define drug similarity and predict compound activity. Assay Drug Dev. Tech. *16*, 162–176.

Delaney, J.S. (2004). Esol: estimating aqueous solubility directly from molecular structure. J. Chem. Inf. Comput. Sci. *44*, 1000–1005.

DiMasi, J.A., Grabowski, H.G., and Hansen, R.W. (2016). Innovation in the pharmaceutical industry: new estimates of r&d costs. J. Health Econ. *47*, 20–33.

Dopazo, J. (2014). Genomics and transcriptomics in drug discovery. Drug Discov. Today *19*, 126–132.

Gammon, D.W., Steenkamp, D.J., Mavumengwana, V., Marakalala, M.J., Mudzunga, T.T., Hunter, R., and Munyololo, M. (2010). Conjugates of plumbagin and phenyl-2-amino-1-thioglucoside inhibit mshb, a deacetylase involved in the biosynthesis of mycothiol. Bioorg. Med. Chem. *18*, 2501–2514.

Geeleher, P., Cox, N.J., and Huang, R.S. (2016). Cancer biomarker discovery is improved by accounting for variability in general levels of drug sensitivity in pre-clinical models. Genome Biol. *17*, 190.

Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the cancer cell line encyclopedia. Nature *569*, 503.

Girsh, L. (2007). Lipid-containing compositions and methods of using them. US Patent App. 11/501,380.

Gomez-Bombarelli, R., Wei, J.N., Duvenaud, D., Hernandez-Lobato, J.M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T.D., Adams, R.P., Aspuru-Guzik, A., et al. (2018). Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent. Sci. *4*, 268–276.

Gorska, M., Kuban-Jankowska, A., Milczarek, R., and Wozniak, M. (2016). Nitro-oxidative stress is involved in anticancer activity of 17β-estradiol derivative in neuroblastoma cells. Anticancer Res. *36*, 1693–1698.

Grechishnikova, D. (2021). Transformer neural network for protein-specific de novo drug generation as a machine translation problem. Sci. Rep. *11*, 1–13.

Häse, F., Roch, L.M., and Aspuru-Guzik, A. (2018). Chimera: enabling hierarchy based multi-objective optimization for self-driving laboratories. Chem. Sci. *9*, 7642–7655.

Huang, R., Xia, M., Nguyen, D.-T., Zhao, T., Sakamuru, S., Zhao, J., Shahane, S.A., Rossoshek, A., and Simeonov, A. (2016). Tox21challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. Front. Environ. Sci. *3*, 85.

Joo, S., Kim, M.S., Yang, J., and Park, J. (2020). Generative model for proposing drug candidates satisfying anticancer properties using a conditional variational autoencoder. ACS Omega *5*, 18642–18650.

Kaur, P., Garg, T., Rath, G., Murthy, R., and Goyal, A.K. (2016). Surfactant-based drug delivery systems for treating drug-resistant lung cancer. Drug Deliv. *23*, 717–728.

Kawiak, A., Domachowska, A., Jaworska, A., and Lojkowska, E. (2017). Plumbagin sensitizes breast cancer cells to tamoxifen-induced cell death through grp78 inhibition and bik upregulation. Sci. Rep. *7*, 43781.

Klaveness, J., Rongved, P., Høgset, A., Tolleshaug, H., Cuthbertson, A., Godal, A., Hoff, L., Gogstad, G., Bryn, K., Naevestad, A., et al. (2004). Diagnostic/therapeutic Agents (US Patent), p. 6,680,047.

Krishnan, S.R., Bung, N., Bulusu, G., and Roy, A. (2021). Accelerating de novo drug design against novel proteins using deep learning. J. Chem. Inf. Model. *61*, 621–630.

Lao, J., Madani, J., Puértolas, T., Álvarez, M., Hernández, A., Pazo-Cid, R., Artal, Á., and Antón Torres, A. (2013). Liposomal doxorubicin in the treatment of breast cancer patients: a review. J. Drug Deliv. *2013*, 456409.

Lin, A., Giuliano, C.J., Palladino, A., John, K.M., Abramowicz, C., Yuan, M.L., Sausville, E.L., Lukow, D.A., Liu, L., Chait, A.R., et al. (2019). Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. Sci. Transl. Med. *11*, eaaw8412.

Manica, M., Oskooei, A., Born, J., Subramanian, V., Saez-Rodriguez, J., and Rodriguez Martinez, M. (2019). Toward explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders. Mol. Pharm. *16*, 4797–4806.

Markert, G., Born, J., Manica, M., Schneider, G., and Rodriguez Martinez, M. (2020). Chemical representation learning for toxicity prediction. PharML Workshop at ECML-PKDD (European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases).

Méndez-Lucio, O., Baillif, B., Clevert, D.-A., Rouquié, D., and Wichard, J. (2020). De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. Nat. Commun. *11*, 1–10.

Polishchuk, P.G., Madzhidov, T.I., and Varnek, A. (2013). Estimation of the size of drug-like chemical space based on gdb-17 data. J. Comput. Aided Mol. Des. *27*, 675–679.

Poojari, R. (2014). Embelin–a drug of antiquity: shifting the paradigm towards modern medicine. Expert Opin. Investig. Drugs *23*, 427–444.

Popova, M., Isayev, O., and Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. Sci. Adv. *4*, eaap7885.

Probst, D., and Reymond, J.-L. (2018). Fun: a framework for interactive visualizations of large, high-dimensional datasets on the web. Bioinformatics *34*, 1433–1435.

Probst, D., and Reymond, J.-L. (2020). Visualization of very large high-dimensional data sets as minimum spanning trees. J. Cheminform. *12*, 1–13.

Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. J. Chem. Inf. Model. *50*, 742–754.

Savjani, K.T., Gajjar, A.K., and Savjani, J.K. (2012). Drug solubility: importance and enhancement techniques. ISRN Pharm. *2012*, 195727.

Scannell, J.W., Blanckley, A., Boldon, H., and Warrington, B. (2012). Diagnosing the decline in pharmaceutical r&d efficiency. Nat. Rev. Drug Discov. *11*, 191.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. *10*, 1299–1319.

Schwaller, P., Petraglia, R., Zullo, V., Nair, V.H., Haeuselmann, R.A., Pisoni, R., Bekas, C., Iuliano, A., and Laino, T. (2020). Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. Chem. Sci. *11*, 3316–3325.

Sharifi-Noghabi, H., Peng, S., Zolotareva, O., Collins, C.C., and Ester, M. (2020). Aitl: adversarial inductive transfer learning with input and output space adaptation for pharmacogenomics. Bioinformatics *36*, i380–i388.

Shivakumar, P., and Krauthammer, M. (2009). Structural similarity assessment for drug sensitivity prediction in cancer. In BMC Bioinformatics, *volume 10*BMC Bioinformatics (Springer), p. S17.

Shoemaker, R.H. (2006). The nci60 human tumour cell line anticancer drug screen. Nat. Rev. Cancer *6*, 813.

Skalic, M., Sabbadin, D., Sattarov, B., Sciabola, S., and De Fabritiis, G. (2019). From target to drug: generative modeling for the multimodal structure-based ligand design. Mol. Pharm. *16*, 4282–4291.

Sorkun, M.C., Khetan, A., and Er, S. (2019). Aqsoldb, a curated reference set of aqueous solubility and 2d descriptors for a diverse set of compounds. Sci. Data *6*, 1–8.

Tanimoto, T.T. (1958). Elementary Mathematical Theory of Classification and Prediction (IBM Internal Report).

Verbist, B., Klambauer, G., Vervoort, L., Talloen, W., Shkedy, Z., Thas, O., Bender, A., Göhlmann, H.W., Hochreiter, S., Consortium, Q., et al. (2015). Using transcriptomics to guide lead optimization in drug discovery projects: lessons learned from

the qstar project. Drug Discov. Today *20*, 505–513.

Wehling, M. (2009). Assessing the translatability of drug projects: what needs to be scored to predict success? *Nature reviews*. Drug Discov. *8*, 541–546.

Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M., Network, C.G.A.R., et al. (2013). The cancer genome atlas pan-cancer analysis project. Nat. Genet. *45*, 1113.

Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic Acids Res. *46*, D1074–D1082.

Wong, C.H., Siah, K.W., and Lo, A.W. (2019). Estimation of clinical trial success rates and related parameters. Biostatistics *20*, 273–286.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson IR, et al.. (2012). Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. *41*, D955–D961.

Yang, C., Zhuang, P., Shi, W., Luu, A., and Li, P. (2019). Conditional structure generation through graph variational generative adversarial nets. Adv. Neural. Inf. Process. Syst. *32*, 1338–1349.

You, J., Liu, B., Ying, Z., Pande, V., and Leskovec, J. (2018). Graph convolutional policy network for goal-directed molecular graph generation. Adv. Neural Inf. Process. Syst. *31*, 6410–6421.

Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K., and Finn, C. (2020). Gradient surgery for multi-task learning. Adv. Neural. Inf. Process. Syst. *33*.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., and Smola, A.J. (2017). Deep sets. Adv. Neural. Inf. Process. Syst. *30*, 3391–3401.

Zhang, X., Wang, W., True, L.D., Vessella, R.L., and Takayama, T.K. (2009). Protease-activated receptor-1 is upregulated in reactive stroma of primary prostate cancer and bone metastasis. Prostate *69*, 727–736.

Zhang, X., Yang, C., Rao, X., and Xiong, J. (2016). Plumbagin shows anti-cancer activity in human breast cancer cells by the upregulation of p53 and p21 and suppression of g1 cell cycle regulators. Eur. J. Gynaecol. Oncol. *37*, 30–35.

Zhavoronkov, A., Ivanenkov, Y.A., Aliper, A., Veselov, M.S., Aladinskiy, V.A., Aladinskaya, A.V., Terentiev, V.A., Polykovskiy, D.A., Kuznetsov, M.D., Asadulaev, A., et al. (2019). Deep learning enables rapid identification of potent ddr1 kinase inhibitors. Nat. Biotechnol. *37*, 1038–1040.

iScience, Volume *24*

Supplemental information

PaccMann^{RL}: De novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning

Jannis Born, Matteo Manica, Ali Oskooei, Joris Cadow, Greta Markert, and María Rodríguez Martínez

Supplementary Material

S1 Transparent Methods

S1.1 Neural network architectures

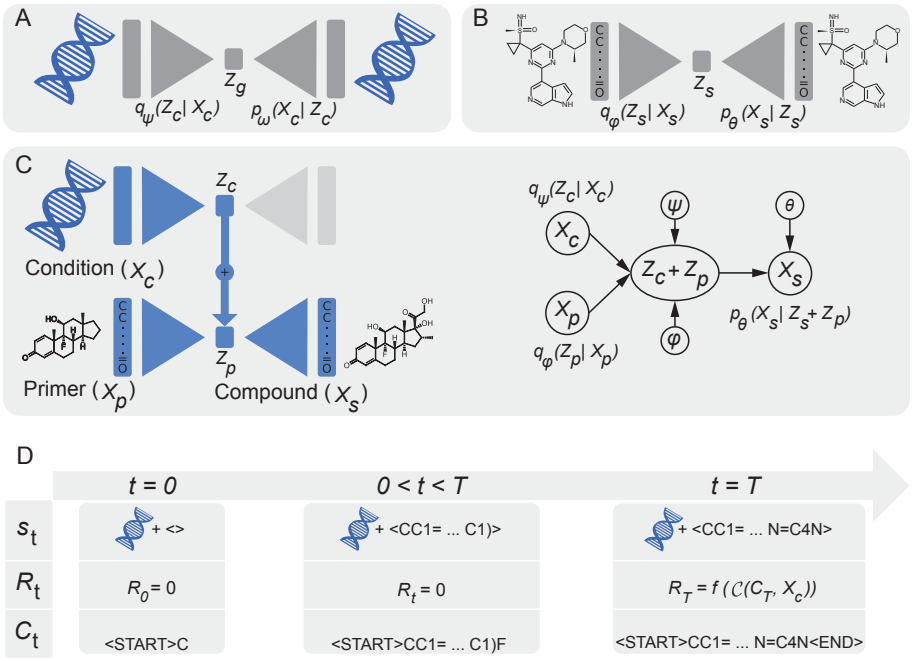


Figure S1: **Architectural details of the conditional molecular generator; related to Figure 1.** **A)** A biomolecular profile VAE (PVAE) was pretrained on RNA-Seq data from TCGA to encode a transcriptomic profile X_c into a latent code Z_c, before attempting to decode X_c from it. **B)** Similarily, a sequential compound generator VAE (SVAE) was trained to encode and decode SMILES representations X_s of molecules. **C)** PVAE and SVAE are combined to obtain a conditional molecule generator. As shown in the graphical model, the combination is achieved by using a permutation-invariant operation (e.g. addition) to fuse the latent spaces of omics profiles and molecules to a joint, multimodal representation. **D)** Molecules are generated directly as SMILES sequences and are assembled in a sequential process, one token at a time. A full cycle of this process includes a state (s_t, where $s_0 = X_c$, i.e. a TCGA RNA-Seq transcriptomic profile), a reward (R_t) and a generated candidate compound (C_t).

Conditional generator.

Our conditional generator is a molecule generator that produces a molecular structure represented as its SMILES sequences (Weininger, 1988). SMILES sequences are preferable over (functional) fingerprint-based representations of molecules (e.g. ECFP (Rogers and Hahn, 2010)) since they have shown to be superior in both predictive (Jastrzebski et al., 2016; Manica et al., 2019) and generative models for molecules (Bjerrum and Sattarov, 2018). In our use case, the generative process is conditioned on a target biomolecular profile, e.g. from a patient or a disease. Inspired by Gomez-Bombarelli et al. (2018), we concluded that variational autoencoders (VAE) are the ideal generative model for our task since by design they bring about a structurally ordered latent space that simplifies the combination of different information sources. Our conditional generator combines

two VAEs that are trained independently prior to being fused together: 1) a denoising VAE for cancer profile encoding/generation (called PVAE, Figure S1A) and 2) a sequential VAE (SVAE) for SMILES sequence generation (Figure S1B). The mathematical formulation of VAEs can be found in Kingma and Welling (2014); Sohn et al. (2015). PVAE is pretrained on gene expression profiles (GEP) to learn a consistent latent representation for biomolecular signatures. SVAE is pretrained on bioactive drug-like molecules to learn the syntax of valid SMILES and general molecular semantics. Generative models of SMILES sequences necessitate the ability to *count* Models that process SMILES sequences greatly benefit from the ability to *count* the ring opening and closing symbols in a molecule, as a single mistake in the sequential generation of a SMILES renders the entire string invalid. To circumvent that standard recurrent and convolutional networks lack the proficiency to count, we utilize a stack memory (Hopcroft and Ullman, 1969), in our case implemented through stack-augmented GRUs as proposed by Joulin and Mikolov (2015) (for equations and other details of the SVAE and the stack see subsubsection S1.3. Thereafter, the encoder of the PVAE is fused with the decoder of the SVAE via their latent space (Figure S1C). 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 RL optimization phase, the weights of the fused model (which were pretrained independently) are fine-tuned using a reward from the critic.

Critic ($\mathcal{C}$).
The critic is a multimodal drug sensitivity prediction model that evaluates the efficacy of any given candidate compound against a biomolecular profile of interest, e.g. gene expression of a cancer cell line. The critic outputs a non-negative reward that depends on the candidate compound predicted IC50 for the target profile, such that low IC50 values associated with higher compound efficacy receive higher rewards than high IC50 values. The reward is then 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 (Manica et al., 2019).

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 a latent space, Z_c (see Figure S1C). 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. However, this priming is optional and we do not only sample closely around existing compounds but we instead sample a larger fraction of the chemical space. As can be seen in the graphical model in Figure S1C the molecule generation is conditioned on a context $\mathcal{Z}$, where in this work $\mathcal{Z} = \{Z_c, Z_p\}$. Z_c and Z_p reflect embeddings learned from semantically different data modalities (gene expression and molecules). To combine these (latent) representations, we use summation because it is a permutation invariant operation and has been proposed to combine a variable set of unstructured latent encodings (Zaheer et al., 2017). Alternatives include mixup functions such as weighted sums or dimension-wise sampling from a categorical (Bernoulli) distribution (Beckham et al., 2019). Our additive latent representation is similar in concept to the conditional VAE with additive Gaussian encoding space (Wang et al., 2017). Intuitively, this fusion presumably warps the latent space from encoding structural similarity (of molecules or GEP) into functional similarity so as to aggregate molecules with similar predicted efficacy for a given cell line (Gomez-Bombarelli et al., 2018). 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 S1C. Following the notation of Popova et al. (2018), the conditional generator, $\mathcal{G}$, acts as the *agent* and PaccMann (the multimodal IC50 prediction model, $\mathcal{C}$) represents the *critic*. The weights of $\mathcal{C}$ are fixed. 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 Popova et al. (2018), we define the set of states $\mathcal{S}$ as all possible SMILES strings (with length $\leq T$) paired with the target GEP. The set of possible actions a that $\mathcal{G}$ can take is a set $\mathcal{A}$, which is a vocabulary of all characters and symbols of the canonical SMILES language. As depicted in Figure S1D, 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 $s_{t-1} = (C_{t-1}, X_c)$ and C_0 is simply the `<START>` token. 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:

$$\Pi(\Theta) = \sum_{s_T \in S^*} P_\Theta(s_T) R(s_T) \tag{1}$$

where $P_\Theta(s_T) := \prod_{t=0:T} p(a_t|s_{t-1})$ and the state $s_T = (C_T, X_c)$ is a tuple of the candidate compound C_T and the cell profile X_c and the reward $R(s_T) = f(\mathcal{C}(C_T, X_c))$ is the output of the critic C scaled by a reward function f. In our experiments, all intermediate rewards $R(s_t)$ are set to 0, since C_t (the intermediate SMILES string) will in almost all cases not resemble a valid molecule. The sum is approximated using policy gradients, specifically the REINFORCE algorithm (Williams, 1992) and the reward function f for determining the reward from the predicted log micromolar IC50 is computed by

$$R(S_T) = f(\mathcal{C}(C_T, X_c)) = \exp\left(-\frac{\mathcal{C}(C_T, X_c)}{\alpha}\right) \tag{2}$$

. C_T is the proposed compound, X_c the omic profile and $\alpha \in \mathbb{R}^+$ is a hyperparameter that determines how much the generator is rewarded for designing effective versus ineffective compounds see Figure S2). Smaller α leads to a greedier generator. For all simulations, we set $\alpha = 5$.

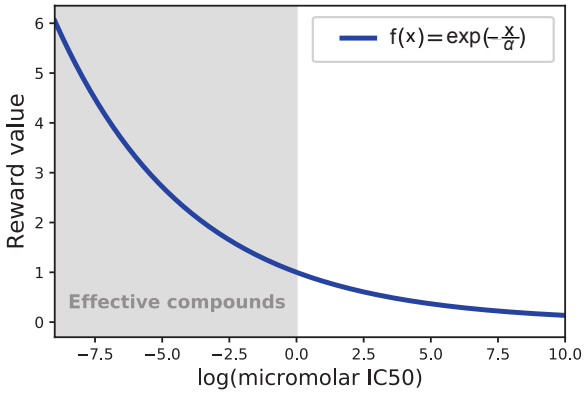


Figure S2: **Reward function of conditional generator; related to Figure 2.** Reward function to map the predicted IC50 of the critic to a reward being fed to the conditional generator. Note that the predictions are given in a log micromolar scale (and not as pIC50), so that a value of 0 corresponds to 1μmol, a commonly considered efficacy threshold. In all experiments and this plot, α was set to 5.

S1.2 Data.

PVAE. For the PVAE, we employed a training dataset of 11,592 (standardized) RNA-Seq GEPs from healthy and cancerous human tissue from the TCGA database and validated it on 1,289 samples from the same database (Weinstein et al., 2013). Since the dataset was too small to train on the full cohort of 20,000 genes and most genes are correlated to a subset of landmark genes, the number of genes was reduced to the same 2,128 genes as used in Manica et al. (2019), following the network propagation procedure described in Oskooei et al. (2019).
SVAE. For pretraining the SVAE, a dataset containing SMILES of 1,576,904 compounds was compiled from the ChEMBL database (Gaulton et al., 2016). 10% of the samples were held out for performance validation, the rest was used for training. Starting from the entire ChEMBL database, no specific handling was applied to parent molecules, salts, pepttides or chelators but SMILES that could not be parsed by `RDKit` were cut off. No sequence length or molecular weight cutoff was needed since our implementation uses an efficient representation of varying lenghts' inputs in `PyTorch`.

For RL optimization of $\mathcal{G}$, we used GEPs publicly available from GDSC (Yang et al., 2012) and CCLE (Barretina et al., 2012) databases. Since the RNA-Seq of these cancer cell line databases were passed through the PVAE (pretrained on human samples from TCGA (Weinstein et al., 2013)), we compared the standardized gene expression distributions for the selected genes across these databases and found good agreement (compare Figure S4 in Supplementary Material S2), in alignment with the reported consensus between transcriptomic data in CCLE and TCGA (Ghandi et al., 2019). To train the critic ($\mathcal{C}$), IC50 drug sensitivity data from GDSC and CCLE was utilized.

S1.3 Implementation and training details

All models were implemented in `PyTorch` 1.0 and trained on a cluster equipped with `POWER8` processors and a `NVIDIA Tesla P100`.

PVAE. The model consisted of four dense layers of [1024, 512, 256 and 200] units with ReLU activation function and dropout of $p = 0.2$ in both, the encoder and the decoder. The dimensionality of the latent space (n) was 128. We minimized the variational loss, consisting of the reconstruction loss and KL divergence, using Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1\text{e}{-8}$) and a decreasing learning rate starting at 0.001 (**?**). To further regularize the PVAE, denoising methods were employed by 1) applying a dropout of 0.1 on the input genes and 2) adding noise to gene expression values ($\varepsilon \sim \mathcal{N}(0, 0.1)$). The model was trained with a batch size of 64 for a maximum of 2000 epochs.

SVAE. The model was trained on molecules provided in SMILES notation, the longest molecules had 1423 tokens. Both encoder and decoder consisted of two layers of bidirectional GRU (hidden size of 128, dropout of 0.1 at the first layer), each complemented with 50 parallel memory stacks with the depth of 50. The latent space of SVAE had the same dimensionality as the PVAE (128) to enable the addition of encodings. Similar optimization parameters as PVAE were used. This model further utilized teacher forcing (Williams and Zipser, 1989), i.e., the model's output is conditioned on the previous ground truth sample as opposed to its generated output. Whilst this significantly simplifies learning, it may drive the generator to predominantly rely on the decoder (thus neglecting the latent encoding). This so called posterior collapse was resolved by applying a token dropout rate of 0.1 during teacher forcing as suggested by Bowman et al. (2016). In addition to token dropout, KL cost-annealing (Bowman et al., 2016) was employed during training, The model was trained with a batch size of 128 for a maximum (early stopping) of $\sim$ 110,000 steps (i.e., exactly 10 epochs) During

training, KL cost-annealing as described in (Bowman et al., 2016) was explored in order to trade-off reconstruction and KL loss.

SVAE – Details on StackGRU architecture. To enable neural networks to count, Joulin and Mikolov (2015) introduced stack-augmented RNN. Stack-RNNs complement RNNs with a differentiable push-down stack operated through learnable controllers, op_t at step t, that involve three operations: `PUSH`, `POP` and `NO-OP` (see Figure S3).

$$op_t = \mathbf{s}(W_{op} h_t), \tag{3}$$

where h_t is the hidden state, W_{op} is a $3 \times H$ matrix (H being the dimension of hidden state) and $\mathbf{s}$ is the softmax function. At each time step the controller probabilities are determined from Equation 3 and the stack memory is updated using the learned controller via a multiplicative gating mechanism:

$$\begin{cases} S_t[0] & = op_t[\texttt{PUSH}]\mathbf{s}(W_{so} h_t) + op_t[\texttt{POP}] S_{t-1}[1] + \\ & \quad op_t[\texttt{NO-OP}] S_{t-1}[0] \\ h_t & = \mathbf{s}(W_i X_t + W_R h_{t-1} + W_{si} S_{t-1}) \end{cases} \tag{4}$$

where S_t is the stack, W_{so} is a $1 \times H$ matrix and W_{si} is a $H \times N$ matrix (N being the stack height). W_i is the input matrix applied to the sequence and W_R is the recurrent matrix. It should be noted that for the sake of brevity, we only show the update equation for the topmost element of the stack in Equation 4.

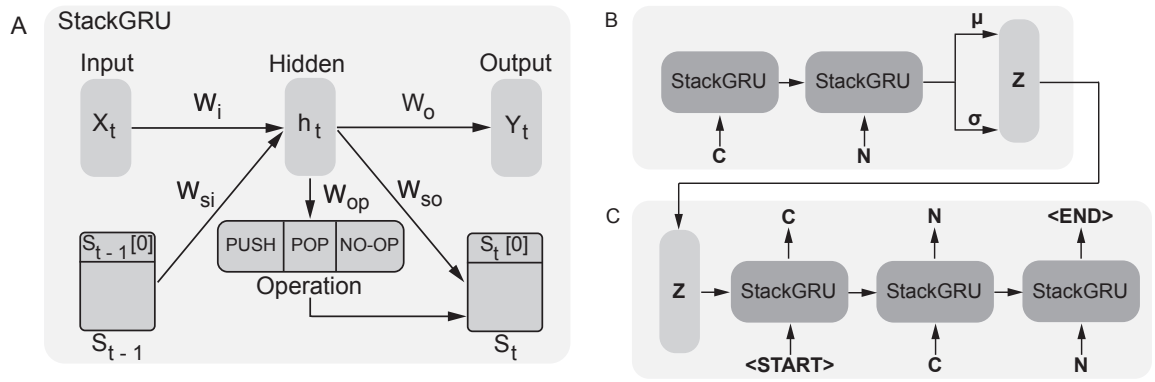


Figure S3: **Stack-GRU architecture employed in the SMILES VAE; related to Figure 1.** (A) The StackGRU architecture adopted in the SVAE. The stack-augmented GRU (StackGRU) architecture complements a regular GRU with a stack that allows one out of three possible operations at each time-step: `PUSH`, `POP` and `NO-OP`. The operation vector is determined through a softmax from the hidden state of each time step.
(B) and (C) are encoder and decoder of the SVAE architecture. (B) encodes the SMILES sequences into multivariate Gaussians with parameters μ and σ. (C) The decoder StackGRU units reconstruct the SMILES sequence from a latent representation (Z_p) sampled from the multivariate Gaussian.

Critic. The critic was trained using the parameters reported in Manica et al. (2019) and replicating the best performing architecture based on multiscale convolutional encoders.

RL training. In order to maximize Equation 1, we employed Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1e{-}4$, weight decay $1e{-}4$) and a decreasing learning rate starting at $1e{-}5$. The gradients were clipped to 2 to prevent $\mathcal{G}$ from destroying its chemical knowledge about SMILES syntax obtained through pretraining on ChEMBL. The reward function hyperparameter α was set to 5.

S1.4 Additional critics for toxicity prediction

For the case study presented in **??**, we utilized the output of two additional predictive neural networks to compute the reward for the generative model. For these experiments, the reward function was computed as:

$$R(S_T) = w_1 \cdot f(\mathcal{C}_{\text{IC50}}(C_T, X_c)) + w_2 \cdot \mathcal{C}_{\text{Tox21}}(C_T) + w_3 \cdot \mathcal{C}_{\text{SIDER}}(C_T) \tag{5}$$

We note that the first summand is identical to the reward function used in the remaining experiments (compare Equation 2). Let Θ_{Tox21} be the neural network that predicts the toxicity of the 12 Tox21 assays. Then $\mathcal{C}_{\text{Tox21}}(C_T) = 1$ iff the output of Θ_{Tox21} is < 0.5 for all 12 Tox21 assays. Otherwise the reward is 0 (as Θ_{Tox21} predicted that C_T is toxic in at least one assay with a probability > 0.5). Similarly, if Θ_{SIDER} is the network that predicts 27 types of adverse drug reactions, then the reward $\mathcal{C}_{\text{SIDER}}(C_T) = 1 - \bar{\mathbf{y}}$, i.e., the inverted mean of the adverse reaction types. Finally, $\vec{\mathbf{w}}$ holds the weights to compute the reward as the weighted sum of the three individual components. We set $w_1 = 1$, $w_2 = 0.2$ and $w_3 = 0.1$. Θ_{Tox21} and Θ_{SIDER} are parametrized using a multiscale convolutional attention (MCA) neural network, a simplification of the architecture developed in Manica et al. (2019) which is detailed in Markert et al. (2020).

S2 Gene expression in human samples and cancer cell lines

Comparing the standardized gene expression values of GDSC (Yang et al., 2012) and CCLE (Barretina et al., 2012) with the one from human samples from TCGA (Weinstein et al., 2013) reveals a similarity (Figure S4). This justifies our choice of utilizing the encoder of the PVAE for cell line data during the RL regime, although it was initially pretrained on human samples from TCGA. By pretraining on TCGA, it was possible to leverage more data ($\sim$10k samples compared to $<$1k in GDSC) which enabled to learn more generic encodings of GEP data.

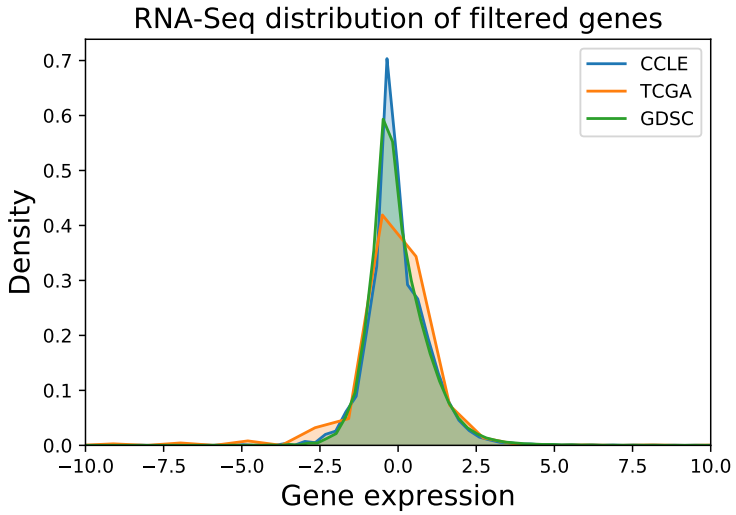


Figure S4: **Distribution of RNA-Seq data across databases; related to Figure 1.** Distribution of standardized gene expression values across the cancer cell line databases CCLE and GDSC as well as the human sample database TCGA.

S3 Results for gene expression profile VAE and SMILES VAE

Profile VAE (PVAE) The pretraining results of the PVAE are presented in Figure S5A, B and C. As shown in Figure S5B, the reconstructed gene expression profiles (GEP), shown in blue, as well as the generated GEPs (green) accurately mimic the distribution of the original GEPs (red). Furthermore, the sampled GEPs follow the same lognormal distribution as the original data. Figure S5C shows that the generated GEPs exhibit a higher similarity to the testing than to the training sample. Overall, these results suggest that the PVAE learns to embed GEPs meaningfully into a latent space that allows both reconstruction and generation of new realistic GEPs of human cells.

SMILES VAE (SVAE) Figure S5D, E and F give a quantitative analysis of the SVAE results following pretraining for 10 epochs with ∼1.4 million structures from ChEMBL. The kernel density estimate (KDE) of the dimensions of the latent space (Figure S5E) validates that the SVAE fulfills the variational constraint as imposed by the Kullback-Leibler divergence in its loss function. We then utilized Tanimoto similarity (Tanimoto, 1958) to compare the ECFP (Rogers and Hahn, 2010) of a subset of 1000 generated molecules with the training and test data from ChEMBL. Figure S5F presents the distributions of the highest Tanimoto similarity between each generated compound and all compounds in training and test dataset respectively. Only a negligible fraction of the generated molecules existed in either of the datasets, whereas the vast majority had a Tanimoto similarity (τ) between 0.2 and 0.6 suggesting that our model learned to propose novel molecular structures from the chemical space of about 10^{30} to 10^{60} molecules (Polishchuk et al., 2013). A snapshot of the interactive Faerun visualization (Probst and Reymond, 2018) of the TMAP Probst and Reymond (2020) of real and generated molecules is shown in Figure S7.

Figure S6A, showcases a panel of 12 generated molecules for qualitative assessment of the molecular structures. The generated molecules generally share drug-like structural features. To inspect the

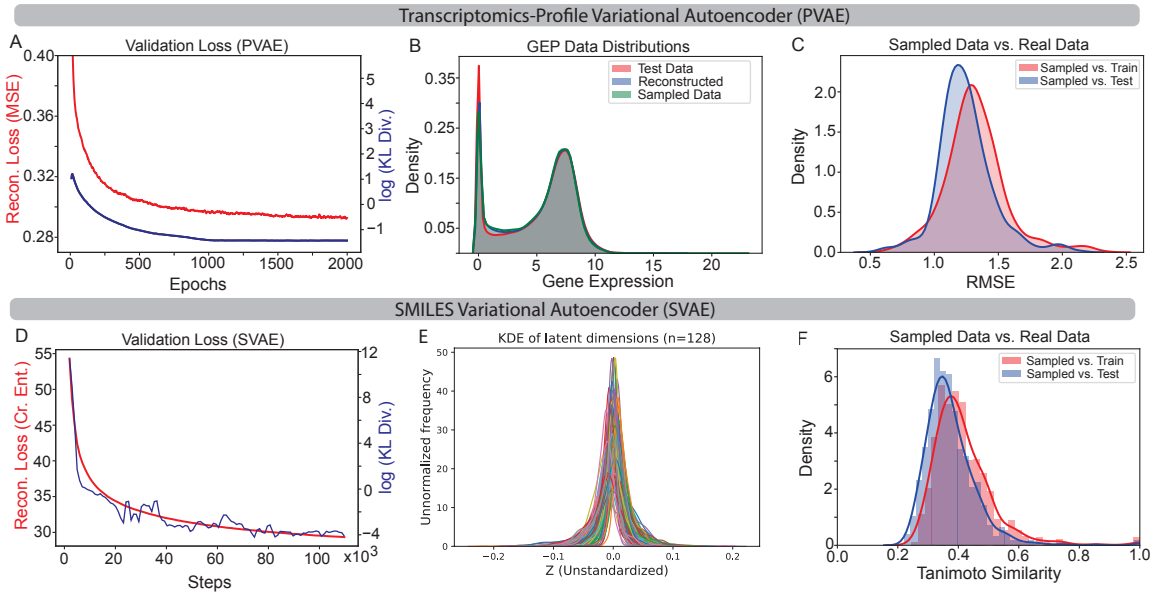


Figure S5: **Results of pretrained PVAE and SVAE model; related to Figure 2.** PVAE: (A) Development of validation error over the course of training. Reconstruction loss (MSE) and KL divergence are shown separately for comparison. (B) Distribution of gene expression values in real, reconstructed and generated samples. (C) Sampled (i.e. generated) data from the latent space of PVAE compared against training and test datasets from TCGA. SVAE: (D) Development of validation error over the course of training. Cross-entropy between target and generated SMILES is shown separately from the KL divergence (log scale for visual clarity). One epoch corresponds to ∼11 000 training steps. (E) Kernel density estimates of all 128 latent dimensions before decoding the test samples. As enforced by the variational constraint, the latent variables follow Gaussian distributions. (F) The Tanimoto similarity between the Morgan fingerprints (ECFP) of the generated molecules and the structures from ChEMBL train and test datasets is used to verify that the generated compounds are sufficiently different from the training data.

smoothness of the latent space of molecules, we encoded a reference molecule shown at the top of Figure S6B into the latent space and decoded four points in the vicinity of the reference molecule leading to the generation of structurally similar yet different compounds.

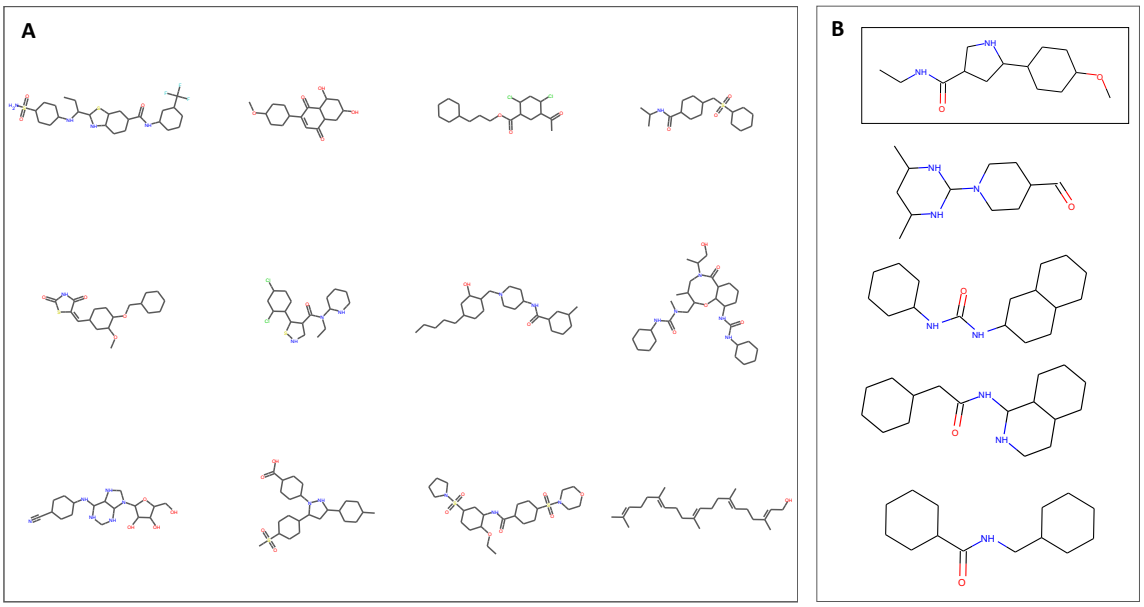


Figure S6: **Qualitative inspection of generated molecules; related to Figure 2.** (A) A sample of 12 molecular structures produced with the SVAE. (B) The molecule depicted at the top was encoded into the latent space. The four molecules below show different decodings from the latent space in the vicinity of the starting molecule.

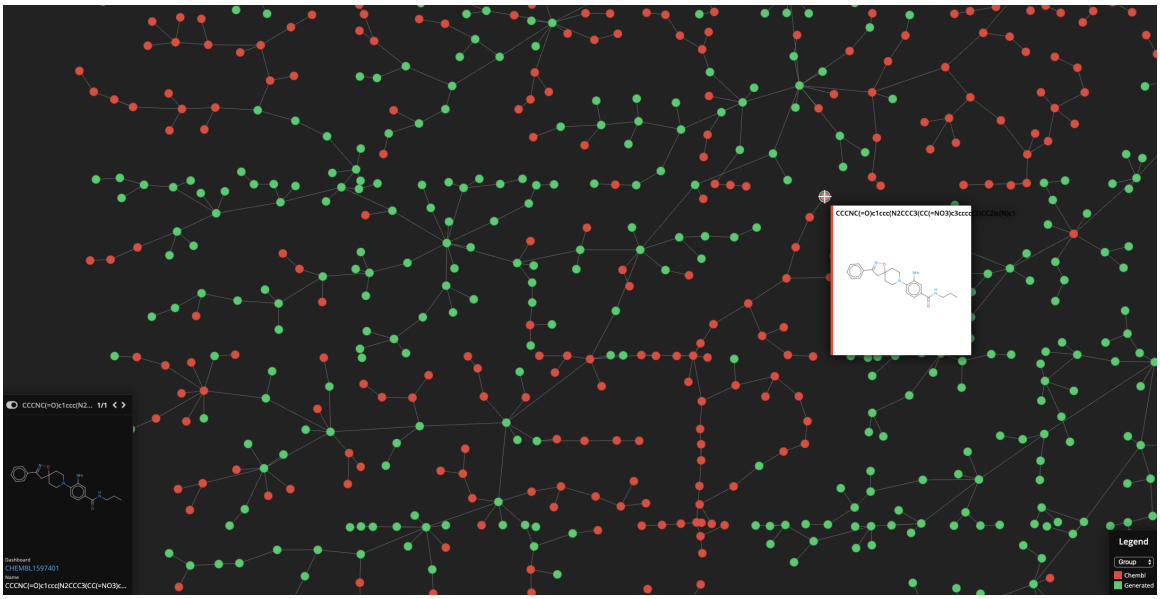


Figure S7: **Snapshot of the interactive TMAP visualization; related to Figure 3.** Generated molecules (green) and ChEMBL compounds (red) are shown through the TMAP algorithm which visualizes the chemical space by aggregating molecules with similar fingerprints (ECFP). The similarity in fingerprints is proportional to the distance of the repsective nodes on the spanning tree. To explore the visualization interactively, please visit https://paccmann.github.io/rl/unbiased.

S4 Nearest neighbors in ChEMBL

To further validate the four site-specific compounds as proposed by our model, Figure S8 depicts the respective nearest neighbors (measured by Tanimoto similarity) in the ChEMBL database of bioactive compounds.

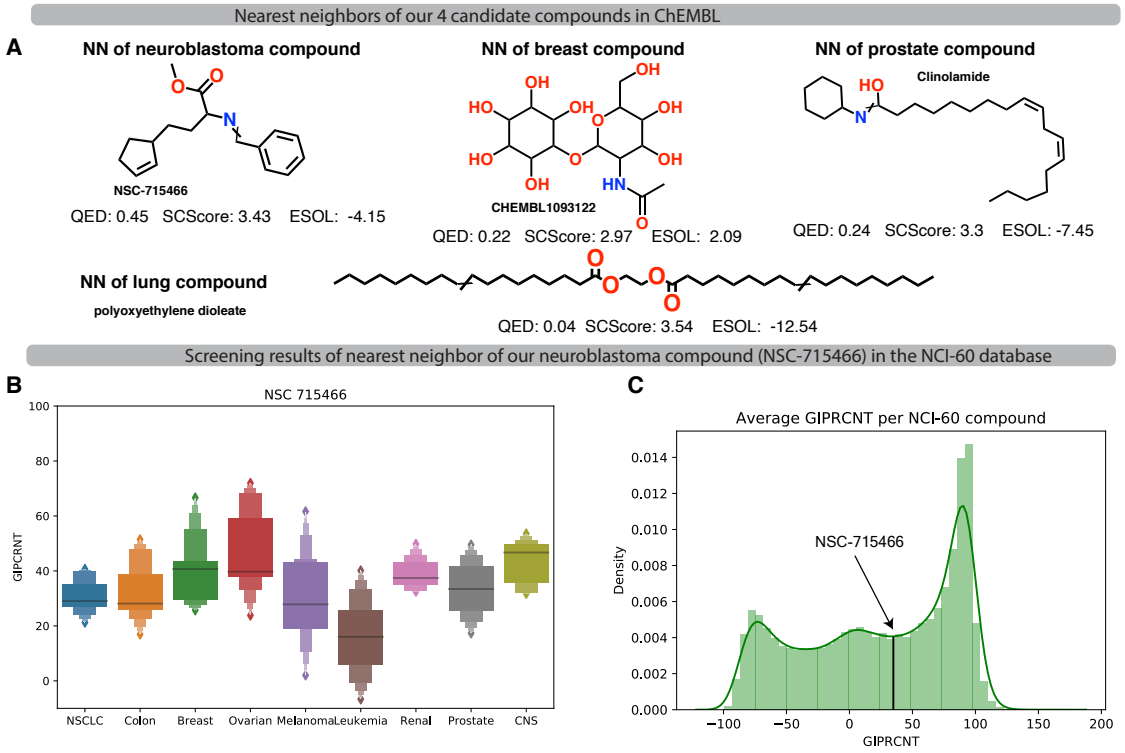


Figure S8: **Nearest neighbors of our site-specific compounds in the ChEMBL database; related to Figure 2. A)** For all four site-specific generated compounds, we performed a similarity search across all molecules in the ChEMBL database. The nearest neighbors are depicted together with relevant drug-like properties. **B)** NSC-715466 was tested in the NCI-60 database (Shoemaker, 2006), where it showed the strongest cell growth inhibition effect against leukemia cell lines. The GIPRCNT is a metric to measure cytotoxicity, where 100% refers to unchanged cell proliferation (identical to the control cells), 0% to complete stopping of cell proliferation and -100% to a full inhibition of all cells. **C)** NSC-715466 showed only moderate anticancer effects as reported in the NCI-60 database.

S5 Validation of critic (PaccMann) on ChEMBL data

Our critic, PaccMann, is an anticancer drug sensitivity prediction model that has been trained *only* on anticancer compounds from GDSC. Since IC50 cell screening data for compounds with knowingly no anticancer effects are notoriously unavailable, PaccMann lacks a *negative training set* which would help extending its generalization across the space of known anticancer compounds. One could thus suspect that PaccMann is generally a flawed evaluator of compounds falling outside the space of anticancer drugs and that it would be biased towards predicting high efficacy for compounds without anticancer effects.

For that reason, Figure S9 shows the predicted efficacy of all anticancer compounds from GDSC (both training and testing data) as well as the predicted efficacy of a representative set of 1000 molecules from ChEMBL. The predicted logarithmic IC50 across all cancer drugs was 2.2 ± 2.2 whereas it was only 3.2 ± 1.6 for ChEMBL molecules.

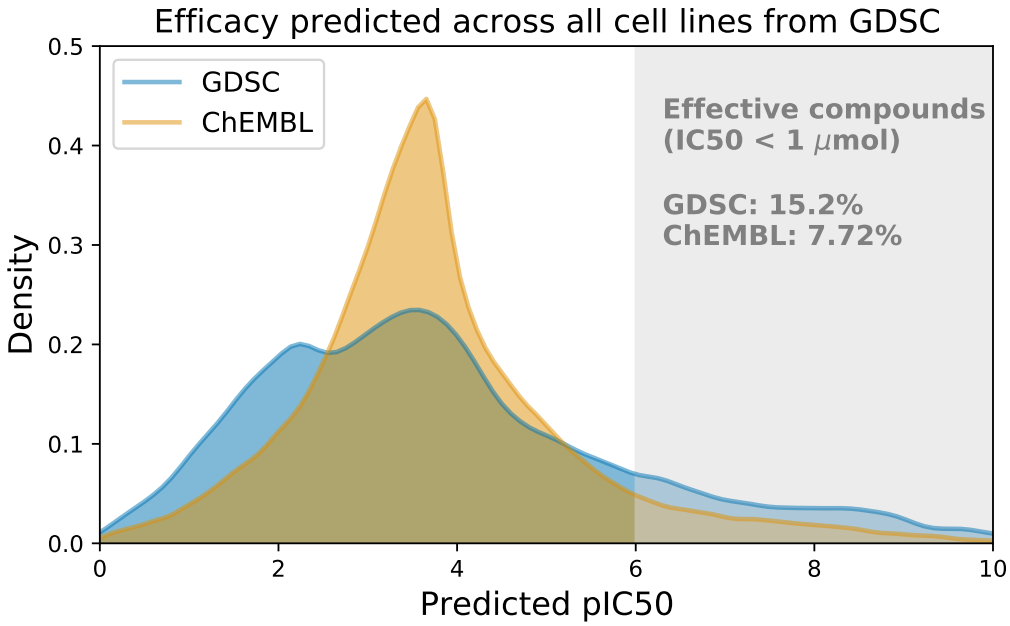


Figure S9: **Comparison of the predicted IC50 of GDSC drugs and bioactive, drug-like compounds from ChEMBL; related to Figure 2.** The density plot of IC50 values as predicted by PaccMann from 210 cancer drugs from GDSC (blue) and 1000 molecules from ChEMBL (orange) across all 965 cell lines from the GDSC panel shows that only a small portion (7.7%) of ChEMBL compounds are predicted as effective against a given cell line, whereas this holds for a significantly larger fraction of GDSC compounds.

S6 Possible synthesis routes for generated molecules

In order to assess the complexity of a potential synthesis of compound proposed by our model Figure S10 shows a predicted synthesis route for a compound gen rated against nervous system cancer.

While the model proposes a panel of possible synthesis routes, the one depicted in Figure S10 decomposes the synthesis into four reactions and a total of ten commercially available reactants. The simplicity of the proposed route as well as the high confidence scores (> 0.8) associated to each reaction seem to be promising indicators towards a possible synthesis. Details of the synthesis route are depicted at the end of document.

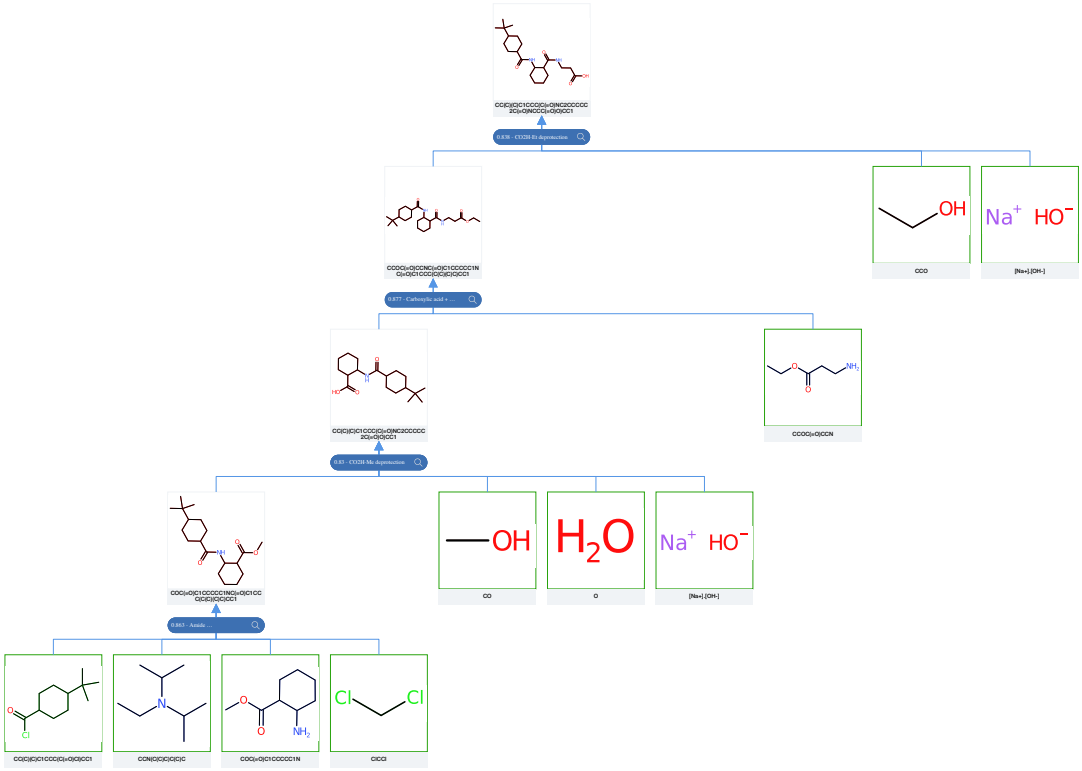


Figure S10: **A retrosynthesis route for a generated molecule; related to Figure 4.** A possible synthesis route, predicted by a molecular retrosynthesis model (Schwaller et al., 2020) is shown for a compound proposed against nervous system cancer (top middle). The predicted synthesis consists of four sequential reactions with a total 10 commercially available reactants (green).

References

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., et al. (2012). The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603.

Beckham, C., Honari, S., Verma, V., Lamb, A. M., Ghadiri, F., Hjelm, R. D., Bengio, Y., and Pal, C. (2019). On adversarial mixup resynthesis. In *Advances in Neural Information Processing Systems*, pages 4348–4359.

Bjerrum, E. and Sattarov, B. (2018). Improving chemical autoencoder latent space and molecular de novo generation diversity with heteroencoders. *Biomolecules*, 8(4):131.

Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A., Jozefowicz, R., and Bengio, S. (2016). Generating sentences from a continuous space. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 10–21. Association for Computational Linguistics.

Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrián-Uhalte, E., et al. (2016). The chembl database in 2017. *Nucleic acids research*, 45(D1):D945–D954.

Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., Barretina, J., Gelfand, E. T., Bielski, C. M., Li, H., et al. (2019). Next-generation characterization of the cancer cell line encyclopedia. *Nature*, 569(7757):503.

Gomez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernandez-Lobato, J. M., et al. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276.

Hopcroft, J. E. and Ullman, J. D. (1969). *Formal Languages and Their Relation to Automata*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Jastrzebski, S., Leśniak, D., and Czarnecki, W. M. (2016). Learning to smile (s). *arXiv preprint arXiv:1602.06289*. ICLR 2016 Workshop.

Joulin, A. and Mikolov, T. (2015). Inferring algorithmic patterns with stack-augmented recurrent nets. In *Advances in neural information processing systems*, pages 190–198.

Kingma, D. P. and Welling, M. (2014). Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR 2014*.

Manica, M., Oskooei, A., Born, J., Subramanian, V., Saez-Rodriguez, J., and Rodriguez Martinez, M. (2019). Toward explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders. *Molecular Pharmaceutics*. PMID: 31618586.

Markert, G., Born, J., Manica, M., Schneider, G., and Rodriguez Martinez, M. (2020). Chemical representation learning for toxicity prediction. *PharML Workshop at ECML-PKDD (European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases)*.

Oskooei, A., Manica, M., Mathis, R., and Martínez, M. R. (2019). Network-based biased tree ensembles (netbite) for drug sensitivity prediction and drug sensitivity biomarker identification in cancer. *Scientific reports*, 9(1):1–13.

Polishchuk, P. G., Madzhidov, T. I., and Varnek, A. (2013). Estimation of the size of drug-like chemical space based on gdb-17 data. *Journal of computer-aided molecular design*, 27(8):675–679.

Popova, M., Isayev, O., and Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. *Science advances*, 4(7):eaap7885.

Probst, D. and Reymond, J.-L. (2018). Fun: a framework for interactive visualizations of large, high-dimensional datasets on the web. *Bioinformatics*, 34(8):1433–1435.

Probst, D. and Reymond, J.-L. (2020). Visualization of very large high-dimensional data sets as minimum spanning trees. *Journal of Cheminformatics*, 12(1):1–13.

Rogers, D. and Hahn, M. (2010). Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754.

Schwaller, P., Petraglia, R., Zullo, V., Nair, V. H., Haeuselmann, R. A., Pisoni, R., Bekas, C., Laino, T., et al. (2020). Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chemical Science*.

Shoemaker, R. H. (2006). The nci60 human tumour cell line anticancer drug screen. *Nature Reviews Cancer*, 6(10):813.

Sohn, K., Lee, H., and Yan, X. (2015). Learning structured output representation using deep conditional generative models. In *Advances in neural information processing systems*, pages 3483–3491.

Tanimoto, T. T. (1958). Elementary mathematical theory of classification and prediction. *IBM Internal Report*.

Wang, L., Schwing, A., and Lazebnik, S. (2017). 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.

Weininger, D. (1988). 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.

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J. M., Network, C. G. A. R., et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113.

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

Williams, R. J. and Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280.

Yang, W., Soares, J., Greninger, P., Edelman, et al. (2012). Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research*, 41(D1):D955–D961.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. R., and Smola, A. J. (2017). Deep sets. In *Advances in neural information processing systems*, pages 3391–3401.