1. Austė Kanapeckaitė (A, C)
2. Sarper Okuyan (D)
3. David James Wagg (E, F)
4. Jan Koster (D)
5. Ligita Jančorienė (G)
6. Indrė Sakalauskaitė (C, H)
7. Birutė Brasiūnienė (C, I)
8. Andrea Townsend-Nicholson (B)

A. Columbia University Irving Medical Center, 630 W. 168th St. New York, NY 10032. United States
B. Research Department of Structural & Molecular Biology, Division of Biosciences, University College London, London, WC1E 6BT, United Kingdom
C. Faculty of Medicine, Vilnius University, M. K. Čiurlionio g. 21, LT-03101, Vilnius, Lithuania
D. Laboratory for Experimental Oncology and Radiobiology (LEXOR), Cancer Center Amsterdam, Amsterdam UMC and University of Amsterdam; Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands
E. Department of Mechanical Engineering, University of Sheffield, S1 3JD, United Kingdom
F. The Alan Turing Institute, London, NW1 2DB, United Kingdom
G. Vilnius University, Faculty of Medicine, Institute of Clinical Medicine, Department of Infectious Diseases and Dermatovenerology, M. K. Čiurlionio g. 21, LT-03101, Vilnius, Lithuania
H. Vilnius University Hospital Santaros Clinics, Santariškių g. 2, LT-08661 Vilnius, Lithuania
I. Center of Medical Oncology, National Cancer Center of Vilnius University Hospital Santaros Clinics, Vilnius, Lithuania; Santariškių g. 2, LT-08661 Vilnius, Lithuania

Contributions

AK, study design and analysis, writing the manuscript; ATN, senior author, review, edits, and suggestions; SO, R2 expression analysis, Figure 7 preparation; DW, review, edits, and suggestions; JK, R2 incorporation suggestions; LJ, clinical review; IS, clinical and surgery review; BB, clinical oncology review

Corresponding Authors

* Austė Kanapeckaitė, ak4094@columbia.edu; Andrea Townsend-Nicholson, a.townsend-nicholson@ucl.ac.uk

Fragment-Guided New Therapeutic Molecule Discovery and Mapping of Clinically Relevant Interactomes

Abstract Therapeutic intervention solutions for complex diseases depend on the targeted modulation of key pathways in pathology. While growing clinical needs continue to drive advancements in the drug discovery space, current strategies primarily rely on searching large volumes of chemical data without addressing the specific contributions of molecular features. Moreover, both clinicians and researchers recognize the need for improving drug discovery methods and characterization that could aid in clinical strategy selection. To address these challenges, we offer a new perspective on targeted therapy development as well as interactome mapping, utilizing molecular fragments. The present study focused on therapeutic areas that represent emerging targets, namely JAK2 and GLP-1R, both of which have broad clinical potential. We developed a new self-adjusting neural network to capture drug features, which helped discover novel therapeutic candidates for the selected targets with improved binding. In addition, fragment-guided chemical space exploration allowed us to identify new metabolic trajectories that could support drug repurposing efforts and improve the prediction of side effects. Importantly, our work revealed that even a small compound library can effectively generate lead candidates, expediting the search and exploration process. Furthermore, building a robust in silico pipeline with integrated screening data can significantly reduce costs and guide therapy adoption. Thus, our proposed strategy underscores promising avenues for the discovery of new therapeutics and the development of clinical interventions.

Introduction

The drug discovery process can have several trajectories, depending on the search strategy. Whether the choice is to start with a specific target or compound library, the goal is to identify therapeutic candidates with promising activity. This involves a gradual refinement of hits to lead compounds that exhibit favorable drug-like properties (Leelananda & Lindert, 2016; Lu et al., 2022). Moreover, traditional high-throughput methods have been augmented with computer-aided drug design (CADD) to expedite screening and detection. Nevertheless, the chemical space is estimated to be in the order of 10^{64}, rendering the exploration computationally intractable when using traditional methods (Lu et al., 2022). Consequently, new computational strategies are being developed to address the combinatorial complexity of iterating through candidate molecules. Despite the wealth of tools, limitations remain, as a more targeted drug search is needed to sift through chemical variation and to detect features of therapeutically relevant drug-like compounds (Schneider and Clark, 2019; Lu et al., 2022). To tackle these challenges, we need to leverage several observations that emerged in the clinical space.

Firstly, drugs and active compounds often exhibit polypharmacological effects (i.e., modulation of multiple targets) (Ryszkiewicz et al., 2023). This suggests that a compound of interest might have several functional groups that either separately or in-combination influence specific therapeutic outcomes across several targets. Secondly, drug-target stoichiometry and response variability indicate that both drug's efficacy and potency might be related to the functional group composition (Tonge, 2017). Given these insights, we propose a targeted exploration guided by fragmenting active compounds and creating a functional probabilistic space mapping that can be explored and juxtaposed to other readouts (e.g., structure-activity relationships, target networks, disease interactome, etc.). Using this method, instead of trying to group entire compounds - which could prevent us from capturing therapeutically meaningful subgroups - the fragmentation allows us to map the chemical interactome across different therapeutics (Fig.1). Specifically, we focus on active (or assumed to be active) subsets to create 'probability valleys' that can enable targeted feature search and the generation of novel drug candidates via the fragment assembly. Our work demonstrates that even a small library of around 10,000 molecules is sufficient to generate highly diverse compounds when compared to known drugs. This contrasts with the expanding library investments and scale (billions of molecules), which studies have shown can result in missing drug-like candidates (Lyu and Irwin, 2023). In addition, we overlaid fragments with known drug targets and captured additional therapeutic trajectories as well as flagged potential side effect contributors. Therefore, switching from brute-force approaches to rational mapping and design could not only help uncover previously overlooked chemical spaces but also improve cost-effectiveness and speed of the discovery process. To test these ideas, we focused on therapeutic areas with both unmet needs and broad clinical impact potential.

For a successful mapping campaign, we integrated several cutting-edge methods, namely positional encoding (PE) with Laplacian eigenvectors (Belin et al., 2003 and Dwivedi et., al 2021) and the graph isomorphism convolution layer (GIN) (Dwivedi et al., 2023), into an adaptive Gaussian mixture (GM) variational network. Permutation equivalence, uniqueness, and distance awareness with respect to the graph topology are some of the key features of Laplacian eigenvectors, providing a meaningful coordinate system and preserving both the local and global graph structure for compounds of interest. Thus, in our approach, Laplacian PE is used to update fragment features before passing them into a GIN layer for molecular feature embedding. The number of clusters in the variational space is adaptively learned with a Gumbel softmax layer (Jang et al., 2016). Consequently, the low-dimensional GM embedding of molecular fragments enables the learning of data point distributions and allows the network to self-adjust based on the inputs (Varolgunes et al., 2020) (Fig.1).

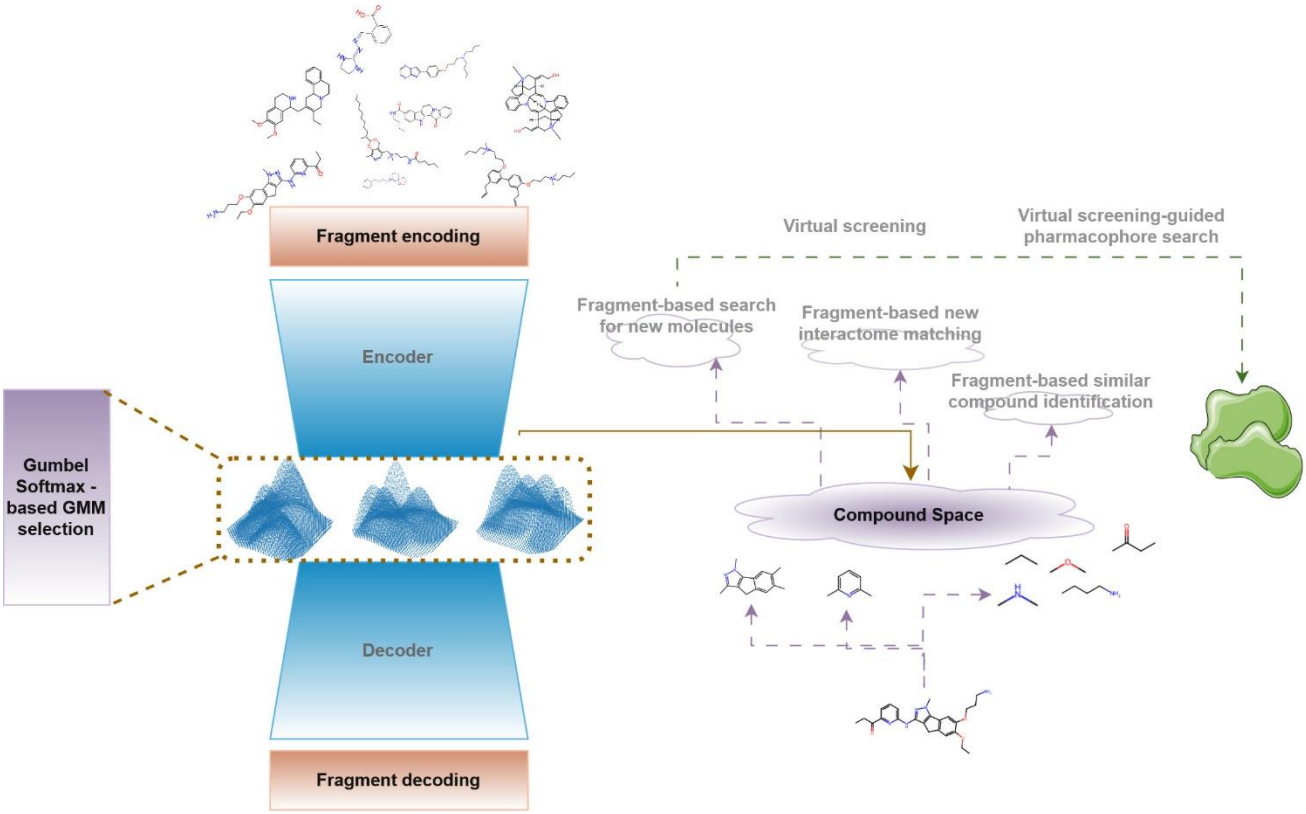


Figure 1. Model and workflow overview. The analysis begins with adaptive learning-based Gaussian mixture latent space mapping of compound fragments. The mapping is used to identify fragments for new molecule generation and relevant interactome identification.

Because 35% of approved drugs target G protein-coupled receptors (GPCRs) and another major class represents kinase inhibitors, we selected therapeutic scenarios focusing on these protein families and their associated drugs (Sriram and Insel, 2018; Anderson et al., 2023). The first candidate was Janus kinase (JAK), which, after phosphorylation, activates the signal transducer and activator of transcription (STAT), initiating a cascade of intracellular and extracellular signaling (Dodington et al., 2018). To investigate JAK modulation, ruxolitinib (INCB018424) was chosen since it is a selective type I kinase inhibitor of JAK1 and JAK2. The drug's action reduces cellular proliferation, lowers cytokine plasma levels, and induces apoptosis (Sapre et al., 2019). Due to intricate regulatory networks, modulating JAK1/JAK2 signaling shows potential in treating not only myeloproliferative neoplasms (MPNs), but also inflammation, as well as glucose and lipid dysregulation (Dodington et al., 2018; Sapre et al., 2019). Furthermore, recent studies have highlighted JAK/STAT involvement in obesity and type 2 diabetes mellitus. Thus, chronic inflammatory conditions could be responsive to JAK inhibitors (JAKi), offering broad applications in clinical care from immune regulation, oncology to surgery and tissue

healing (Walton et al., 2022; Collotta et al., 2023; Yin et al. 2023). The second target was the glucagon-like peptide-1 receptor (GLP-1R), paired with an experimental small-molecule agonist - danuglipron (PF-06882961) (Griffith et al., 2022). Activation of GLP-1Rs by glucagon-like peptide-1 (GLP-1) stimulates insulin secretion and inhibits glucagon release, while delaying gastric emptying. The positive neuroendocrine effects of GLP-1R stimulation result in reduced food intake and improved satiety (Drucker, 2018; Griffith et al., 2022). However, the main drawback of currently available injectable incretin therapies lies in patient adherence, their complex use, and side effects (Nauck, 2016; Griffith et al., 2022; Drucker, 2025). Consequently, orally available small molecule agonists with a good safety profile could offer greater therapeutic flexibility (Griffith et al., 2022).

We first performed a deep learning-driven fragmentation of ruxolitinib and danuglipron, alongside probabilistic embedding to explore the chemical vicinity of their fragments. Subsequently, fragments were selected for new molecule *in silico* retrosynthesis using the BRICS algorithm (Degen et al., 2008). This step was followed by virtual screening to explore binding energetics. The screening confirmed that most new candidate molecules had similar or improved binding compared to the parent molecule. Interestingly, new candidates demonstrated interactions that were not present in the original target molecule. In the case of danuglipron, our analysis hints at another key residue important for binding and molecule orientation that was not identified by previous studies (Griffith et al., 2022). Top candidates were also highly unique when searched against known drugs. In parallel, new molecules were used to optimize a pharmacophore-based search which helped identify two more molecules with markedly improved binding energetics. Thus, fragmentation could be part of the virtual screening pipeline, informing the search and early candidate selection as well as guiding a more focused optimization either through pharmacophore-based search or via candidate-driven library development.

We also aimed to bridge chemical and biological spaces by sampling the drug-target network induced by fragment mapping. Selecting fragment molecules and overlaying them with known biological targets of their parent compounds can be used to explore intricate interactomes. In particular, generated samples from ruxolitinib and danuglipron neighborhoods resulted in enriched clusters spanning multiple metabolic and immune nodes. These insights demonstrate that fragments in the latent space are organized by capturing features beyond the chemical space. Moreover, as our expressome analysis has shown, this strategy could be critical in establishing additional trajectories for drug repurposing and predicting side effects. For example, the ruxolitinib-induced network includes proteins responsible for sphingosine related signaling which could be exploited in certain indications (Mohammed et al., 2023). This subcluster could potentially be suppressed by the downregulation of close neighbors, dependent on JAK2 interactions.

Similarly, off-target effect prediction could offer better risk management for any drug discovery pipeline. The recent discontinuation of Pfizer's danuglipron phase 3 clinical trial underscores the importance of such new methodologies in R&D (Pfizer, 2025; Drucker, 2025). Notably, we also identified interactors that may have contributed to drug induced liver injury observed in Pfizer's trial.

Our findings established that fragmentation can aid with rational molecule design, using significantly smaller compound libraries compared to the current commercial practices. An in-depth look into the fragment-induced interactome and its integration with omics datasets can provide guidance for determining the therapeutic spectrum and off-target events. This is especially important for clinicians, who must continuously adapt and expand treatment protocols to improve responsiveness and limit adverse effects.

Methods

Data and data preparation. Training data consisted of 11,548 small molecules that are in clinical development or have already been approved (ChEMBL Database, 2024). This set was cleaned from duplicates and redundancies, resulting in 10,930 unique molecules. Testing and validation data (454 data points) was generated from the blood-brain barrier penetration (BBBP) dataset, after removing any overlaps with the training data (Sakiyama et al.). ZINC15 (>13 M compounds) (Sterling and Irwin, 2015) was employed for pharmacophore search. DrugBank (Knox et al., 2024) was used for similar compound search. Structures for the analysis were downloaded from the PDB database (Zardecki, 2022), namely GLP-1R (PDB ID: 7S15; cryogenic electron microscope resolved receptor complex with danuglipron analogue PF-06883365) and JAK2 (PDB ID: 6VGL; co-crystal structure with the type-I JAK inhibitor, ruxolitinib, bound to the JAK2 kinase domain in the active phosphorylated conformation). Drug-target search was performed with DGIdb (Cannon et al., 2024) and the interactome parsing relied on the STRING database (v12.0) (Szklarczyk et al., 2023). KEGG database was used for additional pathway exploration (Kanehisa et al., 2025). The expressome analysis included a bulk RNA sequencing study for ruxolitinib (GSE263709) (Ding et al., 2025) and an RNA sequencing of a preclinical mouse model fed high fat diet (HFD) to induce a metabolic syndrome for the treatment with semaglutide (GSE266899) (Cencioni et al., 2025).

Bioinformatics, cheminformatics and data analytics. Python (v 3.10) and R (v 4.4.1) were used as the programing environments. Python packages utilized included Numpy (v 2.2.4), pandas (v 2.2.3), seaborn (v 0.13.2), scikit-learn (v1.6), and matplotlib (v 3.10.1). R libraries employed for the analysis were STRINGdb (v2.16.4) and biomaRt (v2.60.1). RDKit (v 2.13) and Datamol (v0.12.5) facilitated compound analysis and preparation. Molecules were processed using SMILES encoding with implicit hydrogens. Compounds were fragmented

using the BRICS algorithm (Degen et al., 2008) in combination with Bemis-Murcko scaffold identification rules (Bemis and Murcko, 1996). Cutoffs for fragment size were set between 3 and 18 atoms, inclusively. The BRICS algorithm (Degen et al., 2008) was used for in silico retrosynthesis of new compounds in the Datamol (v0.12.5) environment. Docking was performed with SwissDock Suite (Bugnon et al., 2024) and AutoDock Vina (v1.2.0) (Eberhardt et al., 2021) with the following parameters for JAK2: box center: 28 - 10 – 52, box size: 22 - 22 – 22, sampling exhaustivity: 4, and GLP-1R: box center: 68 - 68 – 62, box size: 22 - 35 – 22, sampling exhaustivity: 4 . SwissSimilarity Suite was employed for a comprehensive compound search, comparing 2D fingerprints and electroshape between molecules of interest and DrugBank references (Bragina et al., 2022). Compound absorption, distribution, metabolism, and excretion (ADME) were profiled with SwissADME Suite (Daina et al., 2017). Pharmacophore search was performed with Pharmit against the ZINC15 database (Sunseri and Koes, 2016). PyMOL (Version 3.0 Schrödinger) and ChimeraX (v1.9; Meng et al., 2023) analyses and visualization tools were used for interaction and surface visualization. Adaptive Poisson-Boltzmann Solver (APBS) was utilized for electrostatic potential mapping (Jurrus et al., 2018), while lipophilicity was determined based on molecular lipophilicity potential (MLP) principles (Broto et al., 1984). Both properties were visualized using the ChimeraX suite. Network analysis for fragments involved searching the STRING database (v12.0) with a minimum required interaction score = 0.4. Markov chain clustering (MLC) with inflation set to 3 was applied for interactor group identification. Enrichment parameters were set to false discovery rate (FDR)<=0.05, signal>=0.5, strength>=0.5, min. network count n=2 and background= whole genome (STRING database, v12.0). R2 Genomics Analysis and Visualization Platform (https://r2.amc.nl/) was used for expression profiling and analysis, following recommended parameters, setting FDR<=0.05 and absolute log2 fold change (log2FC)>=0.37

Deep learning. Positional processing and graph embedding was done via Molgraph (v 0.8.0). Deep learning framework was built with Tensorflow (v 2.15), Tensorflow probability (v0.22), and Keras (v2.13). Computational environment employed a Tesla T4 core with 16GB high-bandwidth memory (GDDR6). The network consisted of the encoder, the decoder, the prior set for regularization (cluster size: 3,5,10,15,20,25,30), and the integrated Gumbel softmax learning layer (Tau=0.25) which guided the appropriate selection of a Gaussian mixture layer. Latent space embedding dimensions were fixed at 32. Kullback-Leibler regularization was integrated with probability layer with the following parameters: batch size=250, Monte Carlo sampling size=1000, test point reduction axis=0, weight=1/(batch size * sampling size)). Activation (where applicable) was 'selu' and the dropout rate was set to 0.2. Encoding and decoding loss was custom root mean squared

error (RMSE) which was also supplemented by Kullback-Leibler for detected clusters between the encoder and decoder weighted by 0.25. A detailed network schema is available in Supplementary Figure 2. The Adam optimizer with learning rate=1e-4 and weight_decay=True, and batch size set to 250 were used to train the model over 250 epochs with appropriate validation sets. Principal component analysis (PCA) was used for the latent space visualization and dimensionality reduction after the data points had been normalized using StandardScaler (scikit-learn, v1.6).

Results

Graph-based probabilistic modelling revealed an intricate organization of the chemical space

A ChEMBL library comprising 10,930 compounds was filtered for training, with almost 6,000 compounds belonging to phase 2 clinical trials, and approved drugs included 2,336 therapeutics (Suppl. Fig. 1). For modelling purposes, we focused on the molecular features derived from SMILES (with implicit hydrogens) to characterize molecules using Molgraph (Kensert et al., 2024). The number of atoms under these constraints ranged from 1 to 227. Similarly, molecules contained between 0 to 252 bonds. The average number of atoms per molecule was 26.23 (Suppl. Fig. 1). Training (n= 10,324) and validation (n=227) libraries were prepared by restricting fragment sizes between 3 and 18 atoms, inclusively. These readouts guided our fragmentation strategy to generate 23,314 molecules for training and 1,396 fragments for validation.

Instead of grouping entire compounds - which could limit the identification of therapeutically meaningful subgroups - the fragmentation approach enabled us to map the chemical space across different therapeutics (Fig.1). Detailed mathematical reasoning can be found in Supplementary document 1. By using Laplacian PE and GIN convolutional network integration, we were able to capture both local and global graph structures. The inputs are encoded with Laplacian PE which combines molecule featurization with topological graph information. The model is trained to decode this input as reconstructing graph structure directly has limitations in accuracy (Ochiai et al., 2023). GM embedding of molecular fragments provided a method to model complex spaces and to uncover probabilistic distributions of data points. Although determining cluster distributions might rely on different theoretical assumptions (i.e., Dirichlet processes, categorical selection, etc.), we note that our latent space modeling supports more flexible cluster assignments with different underlying meanings. For our study purposes, each GM was governed by a predefined cluster number, selected from a discrete set that did not exceed the number of latent space dimensions and did not include the edge case of an isotropic Gaussian. The training converged with a 25-cluster GM model, achieving the overall loss after the 250th

epoch: 0.0153 and KL penalty: 0.00011819 (validation loss: 0.0193, KL validation penalty: 0.0160) (Suppl. Fig. 2).

Figure 2 illustrates how the fragments of our targets, namely ruxolitinib and danuglipron, are distributed in the learned latent space. We can see that simpler fragments cluster towards the top-right corner. Furthermore, the fragments evolve in complexity outwardly from this point. It is important to note that the latent space mapping is primarily governed by the researchers' objectives and alternative approaches, such as kernel PCA, could be used to better visualize non-linear relationships. In specific cases - particularly for smaller size compounds - the entire molecular entity can be included in the mapping.

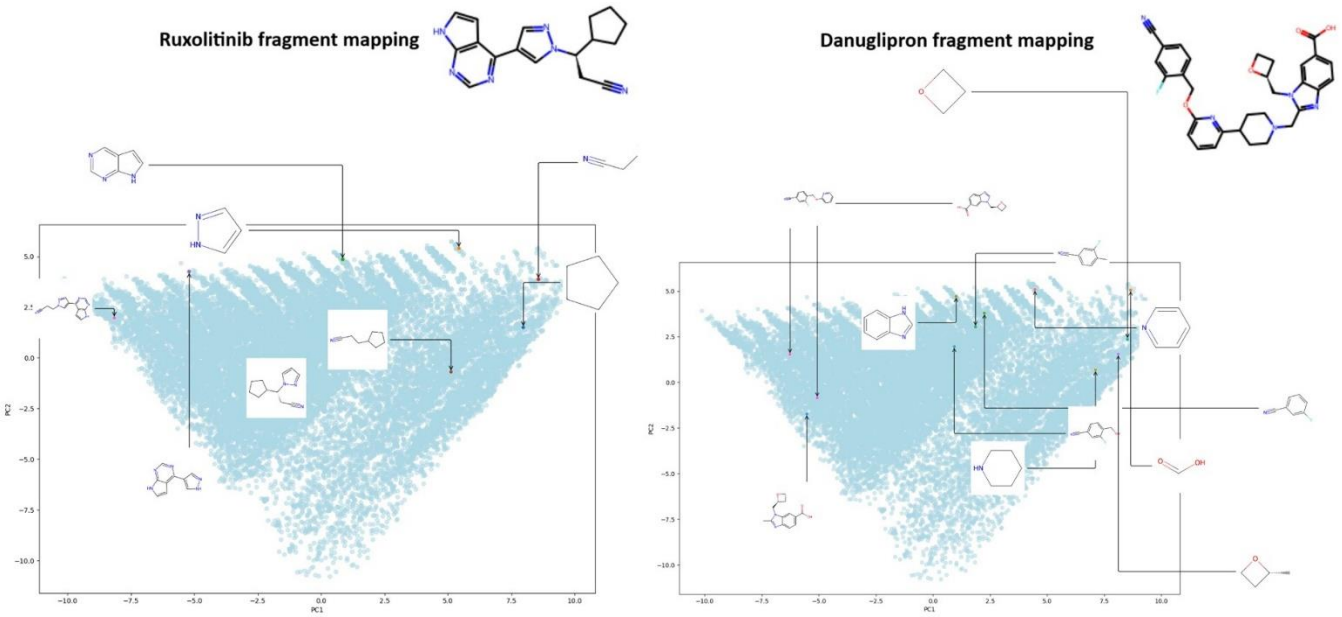


Figure 2. Ruxolitinib and danuglipron fragment mapping with the target compound displayed. PCA transformation is used for dimensionality reduction and better visualization of the latent space, showing the first two principal components.

Fragment-guided molecule generation led to new and improved binding candidate detection

We turned our attention to the 10 nearest neighbors of every mapped fragment of ruxolitinib and danuglipron (assessed via mean squared error for distance). This analysis returned a diverse set of chemical entities for both drugs (Suppl. Fig 3 and suppl. Fig 4). The

neighboring fragments informed the selection of specific candidates for new molecule generation (Suppl. document 2) that structurally and chemically shared at least several features with the parent molecule. Assembly was performed by iterating through all possible fragment binding arrangements under the BRICS algorithm (Degen et al., 2008). We then chose three candidates for JAK2 and GLP1-R based on the predefined criteria (Suppl. document 2). Newly generated molecules were docked using AutoDock Vina (Eberhardt et al., 2021) by setting the binding site, identified through crystallographic and cryogenic electron microscopy studies (JAK2 bound to ruxolitinib PDB ID: 6VGL; GLP-1R with danuglipron analogue PF-06883365 used for the binding site study PDB ID: 7S15). Thus, these experimental studies served as a reference point for the location, and the docking of ruxolitinb and danuglipron established a baseline for binding energetics (Table 1).

Table 1. Docking results for compounds

Ruxolitinib and new therapeutic candidate analysis					
	Ruxolitinib (kcal/mol)	Candidate 1 (kcal/mol)	Candidate 2 (kcal/mol)	Candidate 3 (kcal/mol)	**Pharmacophore candidate 1 based on the lead structure of candidate 3 (kcal/mol)**
Dock 1	-7.686	-7.891	-7.51	-8.757	-9.847
Dock 2	-7.666	-7.542	-7.385	-8.345	-9.752
Dock 3	-7.618	-7.369	-7.34	-8.158	-9.539
Dock 4	-7.584	-7.329	-7.241	-7.954	-9.516
Dock 5	-7.573	-7.221	-7.123	-7.852	-9.328
Mean	**-7.6254**	**-7.4704**	**-7.3198**	**-8.2132**	**-9.5964**
SD	**0.05**	**0.262**	**0.146**	**0.358**	**0.205**
Danuglipron and new therapeutic candidate analysis					
	Danuglipron (kcal/mol)	Candidate 1 (kcal/mol)	Candidate 2 (kcal/mol)	Candidate 3 (kcal/mol)	**Pharmacophore candidate 1 based on the lead structure of candidate 2 (kcal/mol)**

Dock 1	-10.778	-9.871	-10.234	-9.658	-11
Dock 2	-9.702	-9.839	-9.929	-9.651	-10.7
Dock 3	-9.589	-9.797	-9.822	-9.41	-10.554
Dock 4	-9.538	-9.632	-9.693	-9.37	-10.354
Dock 5	-9.281	-9.546	-9.67	-9.28	-10.327
Mean	**-9.7776**	**-9.737**	**-9.8696**	**-9.4738**	**-10.587**
SD	**0.58**	**0.141**	**0.229**	**0.172**	**0.277**
*SD – standard deviation					

The docking summary (Table 1) reveals that the five best binding poses for ruxolitinib were close in energetics (SD: 0.05), with the best binding pose scoring -7.686 kcal/mol. Crystallographically resolved structure of ruxolitinib was the closest to the fifth best pose (-7.573 kcal/mol). Candidate 1 and 2 exhibited more variation but had values comparable to the parent molecule. Interestingly, candidate 3 demonstrated the best overall binding (top pose: -8.757 kcal/mol; mean: -8.2132 kcal/mol), while also showing the most variability (SD: 0.358 kcal/mol). In contrast, danuglipron varied the most in binding poses when compared to other three candidates (SD: 0.58 kcal/mol), and it also had the best binding score of -10.778 kcal/mol. In addition, the resolved structure for GLP-1R with the danuglipron analogue mirrored the best binding pose model of the drug molecule. Other candidates selected for danuglipron optimization had lower variability, with candidate 2 achieving an overall better binding score (mean: -9.8696 kcal/mol; SD: 0.229 kcal/mol).

The observed dynamics in binding energetics prompted us to investigate interactions of the top poses for the parent and candidate molecules (Table 1 and Figure 3).

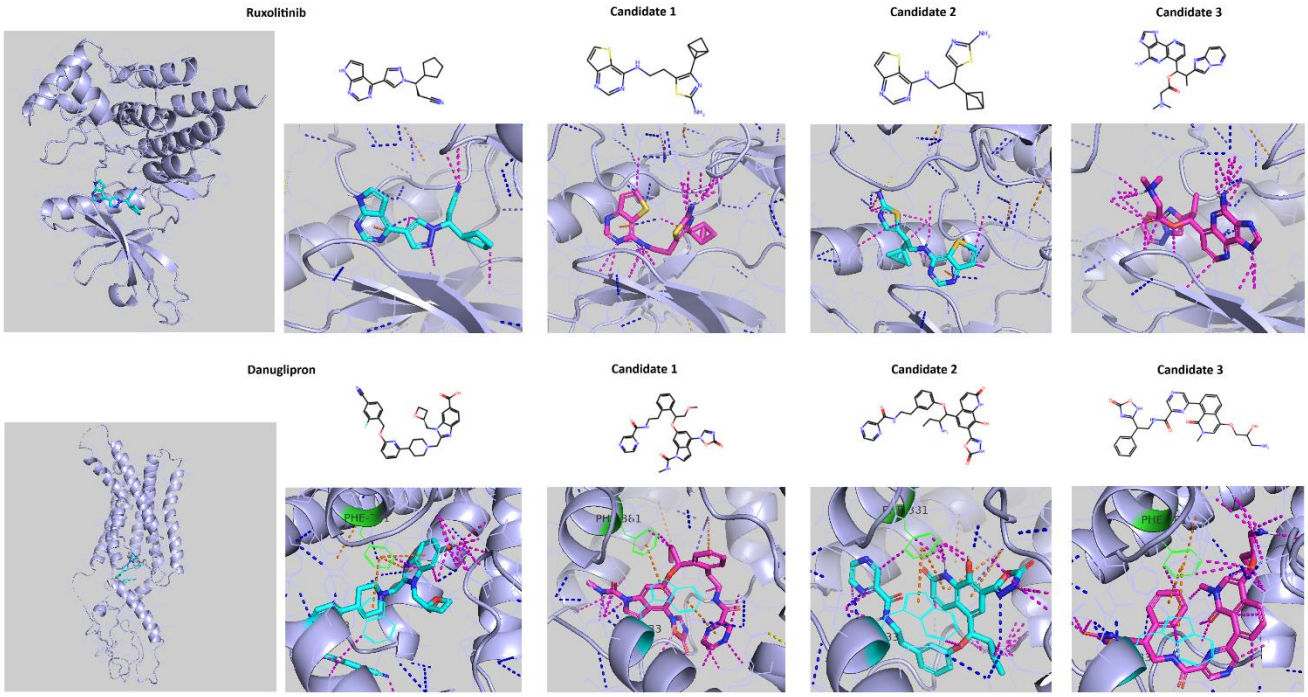


Figure 3. Docking results for the target and generated molecules, where the pose with the best docking score was analyzed. Protein structures include JAK2 (PDB ID: 6VGL) and GLP1-R (PDB ID: 7S15). Blue represents hydrogen bonds and polar bonds, yellow - ionic bonds, orange - all π interactions, magenta - all interchain interactions within 3.5 Å (PyMOL interaction graphics). Highlighted residues for danuglipron are PHE381 (green) and TRP33 (cyan).

As illustrated in Figure 3, ruxolitinib docks with interchain interactions dominating and forming a single π-coordination bond between lysine residue (LYS882) and the pyrrolo[2,3-d]pyrimidine core. Other candidates have noticeably more anchoring points, including both interchain and hydrogen bonds. Candidate 3 employs multiple contacts with side and main chains, which likely contribute to the observed binding strength. Danuglipron appears to have a complex coordination with phenylalanine (PHE381) and tryptophane (TRP33) via 1H-1,3-benzodiazole ring. The drug is further stabilized through polar and non-polar contacts with the side chains. For example, candidates 2 and 3 share similar diverse interactions within the binding pocket; however, candidate 1 no longer participates in π-coordination with PHE381(Fig. 3). Moreover, candidate's 1 pyrazine ring and TRP33 orient edge-to-face their π orbitals and PHE381 maintains a strained coordination with TRP33 via parallel-displaced interchain interactions.

Based on the overall binding performance, we selected candidate 3 for ruxolitinib and candidate 2 for danuglipron to search for similar drugs. We used a combination fingerprint

(electrosphere and 3D fingerprint) to search DrugBank for any similar chemical entities. However, none of the identified candidates exceed a similarity score of 0.4 (Suppl. Fig. 6). Consequently, we further investigated whether we could exploit potentially promising interactions and build a pharmacophore for a large library search to identify better compounds.

We began by anchoring various pharmacophore features to the ruxolitinib candidate 3 for the ZINC15 library search. This iterative process resulted in identifying an optimal pharmacophore structure that led to hit compound ZINC5203833 (Fig. 4). The subsequent docking returned a markedly improved binding (best result: -9.847 kcal/mol, mean: -9.5964 kcal/mol, SD: 0.205 kcal/mol; Table 1). Analysis of APBS electrostatic and MLP surface revealed that candidate 3 is seated deeply within the binding pocket, surrounded by the electron-rich surface with hydrophobic sites adjacent to methyl groups. The electrophilic and nucleophilic site distribution is mirrored by donor-acceptor sites on the pharmacophore. ZINC5203833 takes advantage of these contacts together with π-coordination (LYS882), resulting in a noticeably improved binding. We followed the same analysis steps for dangulipron candidate 2. The molecule is positioned deep inside the binding pocket of the receptor that has both hydrophilic and hydrophobic sites (Fig. 4). Using the mapped pharmacophore for compound search returned ZINC32985857 which demonstrated the best binding energetics when compared to both the generated molecules and the parent compound (best result: 11.000 kcal/mol, mean: -10.587 kcal/mol, SD: 0.277 kcal/mol) (Fig. 4, Table 1). Notably, engagement with phenylalanine (PHE381) and tryptophane (TRP33) residues is also favored when orienting the compound in the constricted binding site (Fig. 4). This was also observed in danuglipron binding (Fig. 3). All compounds were analyzed for their ADME profiles, which, combined with further lead studies and experimental testing, could aid with improved lead formulation (Supplementary Table 1).

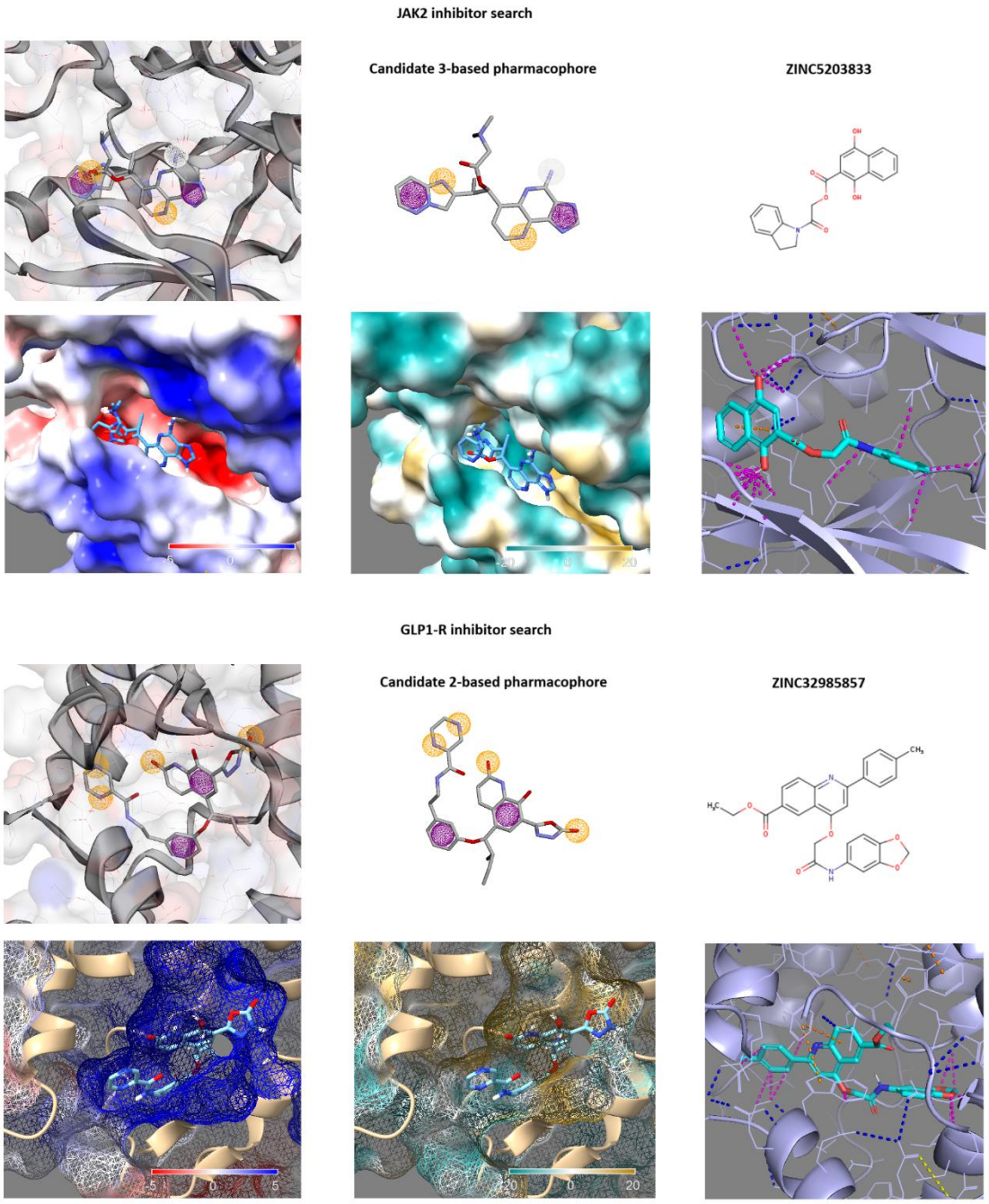
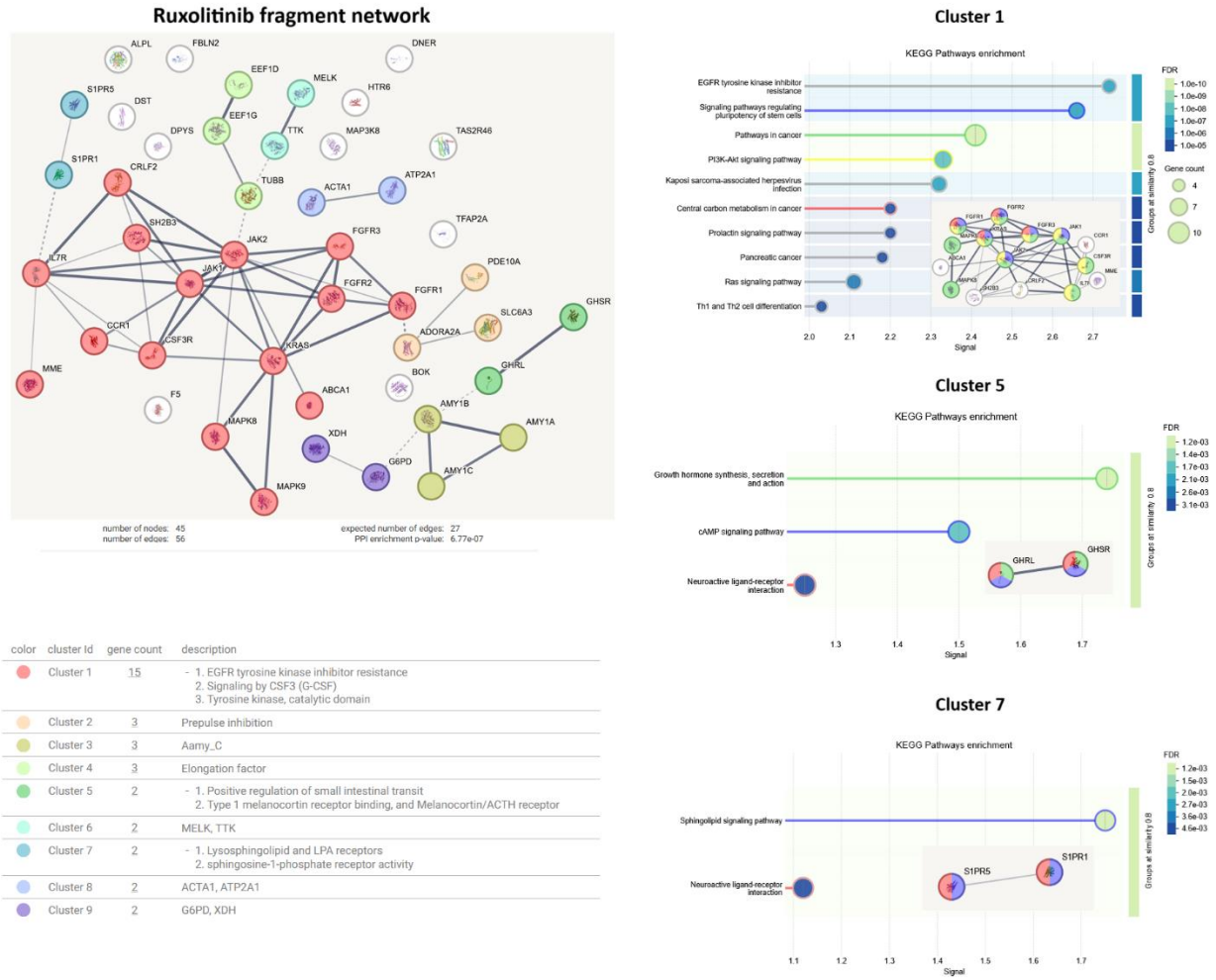


Figure 4. For each protein, the top three panels depict the pharmacophore model (with and without the binding site), where orange spheres (r=1 Å) depict hydrogen acceptors, white spheres (r=1 Å) are hydrogen donors, and magenta spheres (r=1 Å) represent aromatic groups (Pharmit visualisation). The matched structure is displayed in the top-right corner for a given panel set. Bottom panels for each protein, from left to right, visualize Adaptive Poisson-Boltzmann Solver potential (APBS), with a range from -5 kT/e

(red) to blue (+5 kT/e) and molecular lipophilicity potential (MLP), where dark cyan is the most hydrophilic site and dark goldenrod is the most hydrophobic area (ChimeraX visualization) for respective pharmacophore models. The last panel shows binding interactions for the identified molecule (blue represents hydrogen bonds and polar bonds, yellow - ionic bonds, orange - all π interactions, magenta - all interchain interactions within 3.5 Å) (PyMol visualization). GLP-1R surface is shown as a mesh for better visibility of the buried compound.

Fragment-guided interactome search identifies new interactors and networks

To gain insight into the interactome for ruxilitinib and danuglipron, we used a previously identified fragment neighborhood to map molecules to their parent compounds (PCA-based dimensionality reduction). Since fragments can have a varying number of associated parent molecules, it was important to normalize by randomly selecting only one compound per fragment. Alternatively, researchers can search for specific drugs that match fragments of interest for better lead optimization. These compounds created a search set for drug targets in the DGIdb database (Cannon et al., 2024). For ruxolitinib, we identified 47 drug neighbors and their targets above the set interaction threshold of 0.5 (Suppl. Table 2). Using the STRING database, we mapped interactions for these proteins (minimum required interaction score=0.4), which returned a network of 45 nodes (PPI enrichment p-value: 6.77e-07) (Fig. 5). As can be seen, there is one major cluster (n=15, cluster 1) with two prominent subclusters centered around KRAS (cancer metabolism and signaling) and JAK2 (immunological signaling). Noticeably, these proteins are involved in cross-cluster communication either directly or through close network members, emphasizing overlaps for different biological processes (dashed lines, Fig. 5). For example, ADORA2A or adenosine A2a receptor is linked to cluster 1 through FGFR1 (fibroblast growth factor receptor 1). The adenosine receptor is part of purinergic signaling pathways (Fig. 5, cluster 2 enrichment). IL7 receptor signaling is also in close proximity to JAK2, overlapping with multiple pathways and regulatory networks (Fig. 5, cluster 1 enrichment). Clusters 5 and 7 provide further examples of metabolic networks that might span a spectrum of physiological functions (Fig. 5). Ghrelin (GHLR) is the ligand of growth hormone secretagogue receptor type 1 (GHSR) and they both play a role in satiety, intestinal transit, growth functions, as well as neuroactive signaling (KEGG database ID: hsa04080). Cluster 7 consists of sphingosine signaling associated members S1PR1 and S1PR5 that play a role in multiple cellular processes. Despite the fact that the largest cluster 1 dominates the network and the other groups consist of only 2 or 3 members, the significant enrichment suggests the possibility of a more complex interactome.



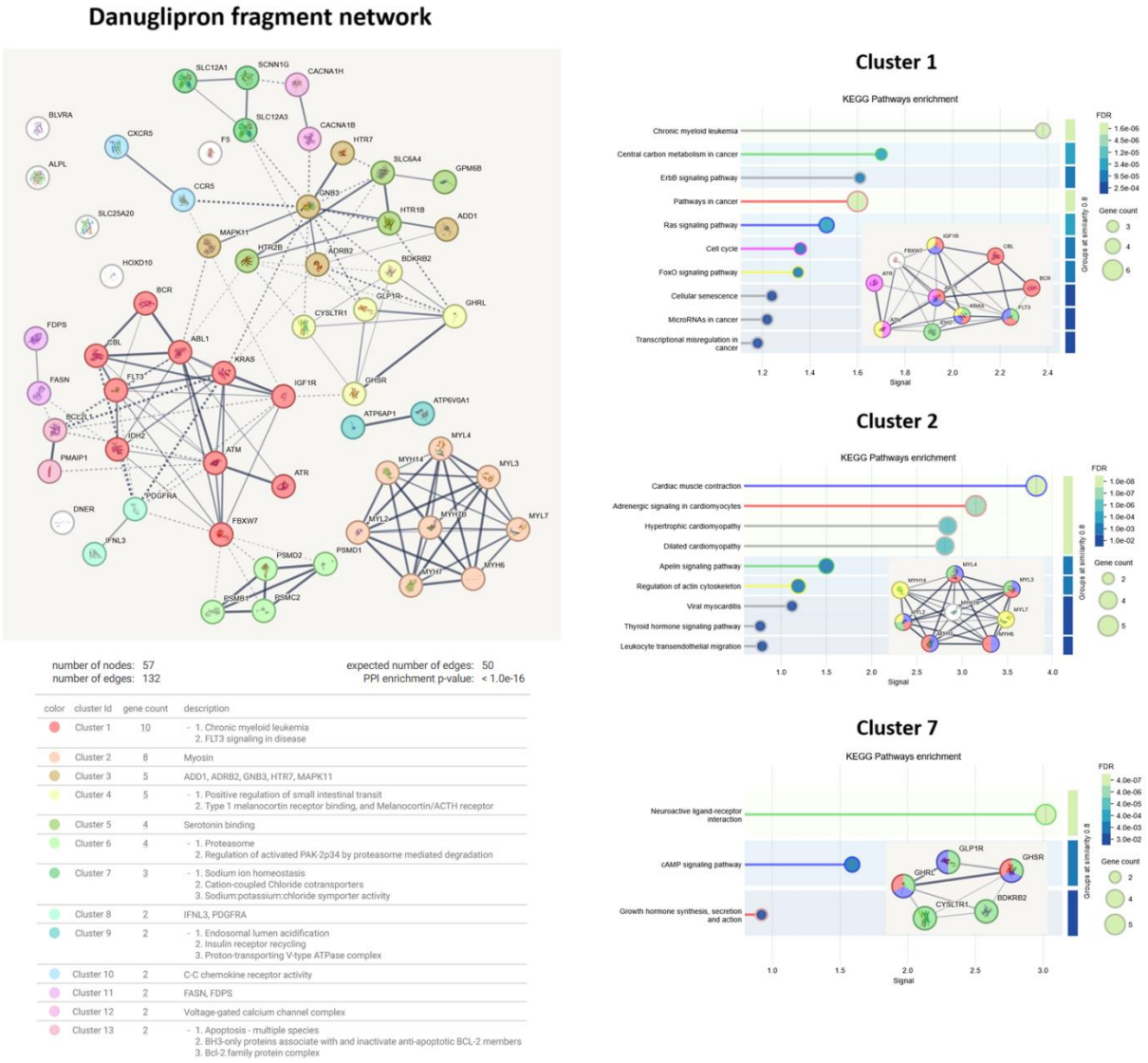
color	cluster Id	gene count	description
	Cluster 1	15	- 1. EGFR tyrosine kinase inhibitor resistance 2. Signaling by CSF3 (G-CSF) 3. Tyrosine kinase, catalytic domain
	Cluster 2	3	Prepulse inhibition
	Cluster 3	3	Aamy_C
	Cluster 4	3	Elongation factor
	Cluster 5	2	- 1. Positive regulation of small intestinal transit 2. Type 1 melanocortin receptor binding, and Melanocortin/ACTH receptor
	Cluster 6	2	MELK, TTK
	Cluster 7	2	- 1. Lysosphingolipid and LPA receptors 2. sphingosine-1-phosphate receptor activity
	Cluster 8	2	ACTA1, ATP2A1
	Cluster 9	2	G6PD, XDH

Figure 5. Ruxolitinib fragment neighborhood network and enrichment for selected biological processes and pathways. STRING database search: minimum required interaction score = 0.4, edges represent confidence. Ruxolitinib network of 45 nodes was clustered using the MCL algorithm with inflation parameter set to 3, where edges between clusters are shown with a dotted line. KEGG pathway enrichment is shown for individual clusters, along with the network fragment and highlighted pathways (enrichment: FDR<=0.05, signal>=0.5, strength>=0.5, min. network count n=2 background= whole genome; functional enrichment display: grouping by 0.8 similarity).

We also mapped interactions (minimum required interaction score=0.4) for the danuglipron neighborhood (n=62, DGIdb interaction threshold > 0.5; Suppl. Table 3), which resulted in a dense network of 57 nodes (PPI enrichment p-value: < 1.0e-16) (Fig. 6). GLP-1R cluster (Fig. 6, cluster 4) has several links with cluster 1 that maps to a number of cellular functions and cancer metabolism pathways. The GLP-1R network is situated closely to 5-hydroxytryptamine receptor 2B (HTR2B) which is a serotonin receptor regulating a number of physiological functions. Similarly, G protein subunit beta 3 (GNB3)

is another interactor of GLP-1R that has multiple metabolic functions. However, the largest cluster (n=10) of the studied protein set (cluster 1) is associated with cancer and cancer metabolism pathways. For instance, the activation of the feline McDonough sarcoma-like tyrosine kinase (FLT3) pathway is essential for cell survival, cell proliferation, and differentiation of hematopoietic progenitor cells (Grafone et al., 2012). The second-largest cluster (n=8) is part of the myosin network. This protein group is not only isolated from the rest of the clusters (no direct links at the set threshold), but it is also significantly enriched in cardiac health and homeostasis associated processes. Other smaller subgroups include chemokine receptors, solute carriers, ion channels and proteasomal as well as cellular transport regulators, illustrating the broad connectivity of the danuglipron fragment network.

By capturing the interactome of ruxolitinib and danuglipron, we can establish signposts for potential modulation of other targets which could be used for repurposing or adverse reaction prediction.



number of nodes: 57
number of edges: 132

expected number of edges: 50
PPI enrichment p-value: < 1.0e-16

color	cluster Id	gene count	description
	Cluster 1	10	1. Chronic myeloid leukemia 2. FLT3 signaling in disease
	Cluster 2	8	Myosin
	Cluster 3	5	ADD1, ADRB2, GNB3, HTR7, MAPK11
	Cluster 4	5	1. Positive regulation of small intestinal transit 2. Type 1 melanocortin receptor binding, and Melanocortin/ACTH receptor
	Cluster 5	4	Serotonin binding
	Cluster 6	4	1. Proteasome 2. Regulation of activated PAK-2p34 by proteasome mediated degradation
	Cluster 7	3	1. Sodium ion homeostasis 2. Cation-coupled Chloride cotransporters 3. Sodium:potassium:chloride symporter activity
	Cluster 8	2	IFNL3, PDGFRA
	Cluster 9	2	1. Endosomal lumen acidification 2. Insulin receptor recycling 3. Proton-transporting V-type ATPase complex
	Cluster 10	2	C-C chemokine receptor activity
	Cluster 11	2	FASN, FDPS
	Cluster 12	2	Voltage-gated calcium channel complex
	Cluster 13	2	1. Apoptosis - multiple species 2. BH3-only proteins associate with and inactivate anti-apoptotic BCL-2 members 3. Bcl-2 family protein complex

Figure 6. Danuglipron fragment neighborhood network and enrichment for selected biological processes and pathways. STRING database search: minimum required interaction score= 0.4, edges represent confidence. Danuglipron network of 57 nodes was clustered using the MCL algorithm with inflation parameter set to 3, where edges between clusters are shown with a dotted line. KEGG pathway enrichment is shown for individual clusters, along with the network fragment and highlighted pathways (enrichment: FDR<=0.05, signal>=0.5, strength>=0.5, min. network count n=2 background= whole genome; functional enrichment display: grouping by 0.8 similarity).

Expressome analysis confirms a complex regulatory network for the selected targets

To complete our investigation of ruxolitinib and danuglipron networks, we performed an analysis on available drug screening studies.

We selected a ruxolitinib study on Philadelphia chromosome-like acute lymphoblastic leukemia (Ph-like ALL) to assess the identified network node responses to the treatment and disease perturbations. This study evaluated gene expression profiles in Ph-like ALL patient xenograft models treated with control or ruxolitinib at different timepoints (GSE263709; Ding et al., 2025). We performed differentially expressed gene analysis using the R2 Platform between the control and the third as well as the seventh treatment days for a drug-sensitive model (Fig. 7a). We found nine genes to be significantly changed from our network (Fig. 5), which spanned different regulatory nodes. For example, membrane metallo-endopeptidase (MME), maternal embryonic leucine zipper kinase (MELK), TTK protein kinase (TTK), tubulin beta class I (TUBB), and fibroblast growth factor receptor 1 (FGFR1) were upregulated after three-day treatment. In addition, SH2B adaptor protein 3 (SH2B3), C-C motif chemokine receptor 1 (CCR1), ATP binding cassette subfamily A member 1 (ABCA1), and JAK2 were downregulated following the three-day treatment with ruxolitinib (Fig. 7a). When comparing these results with the differential expression in a seven-day model, the expressome profile changes: ABCA1 becomes upregulated, JAK2 returns to baseline, and a new gene, glucose-6-phosphate dehydrogenase (G6PD), becomes suppressed. Moreover, FGFR1, MME, SH2B3, CCR1 retain their differential expression. Such fluctuations likely result from the initial network perturbation and subsequent adjustment within the disease model.

In order to explore the danuglipron interactor network, we sought studies of similar profiles, as there were no publicly available datasets for this specific drug. We opted to investigate the gene expression perturbation in mice with induced metabolic syndrome that were treated with semaglutide - a GLP-1R agonist widely used in clinical practice (GSE266899; Cencioni et al., 2025). Despite the study's focus being pancreatic ductal adenocarcinoma, lesion sequencing provided additional insights into metabolic changes (Fig. 7b). Specifically, proteasome 26S subunit ubiquitin receptor, non-ATPase 2 (PSMD2), ATPase H+ transporting accessory protein 1 (ATP6AP1), and HTR2B were significantly downregulated when exposed to the GLP-1R agonist. In contrast, C-X-C motif chemokine receptor 5 (CXCR5) and calcium voltage-gated channel subunit alpha1 H (CACNA1H) showed a marked increase in expression. While the sequencing study focuses on oncological pathology, semaglutide induced alterations that can be mapped to our network underscore complex modulation across several regulatory hubs (Fig. 6 and 7b).

Even though these studies inherently had limitations and expression patterns are influenced by both pathology and tissue type, our findings reflect the need for more

research on how drug-induced expression changes can affect therapeutic outcomes and influence side effect emergence.

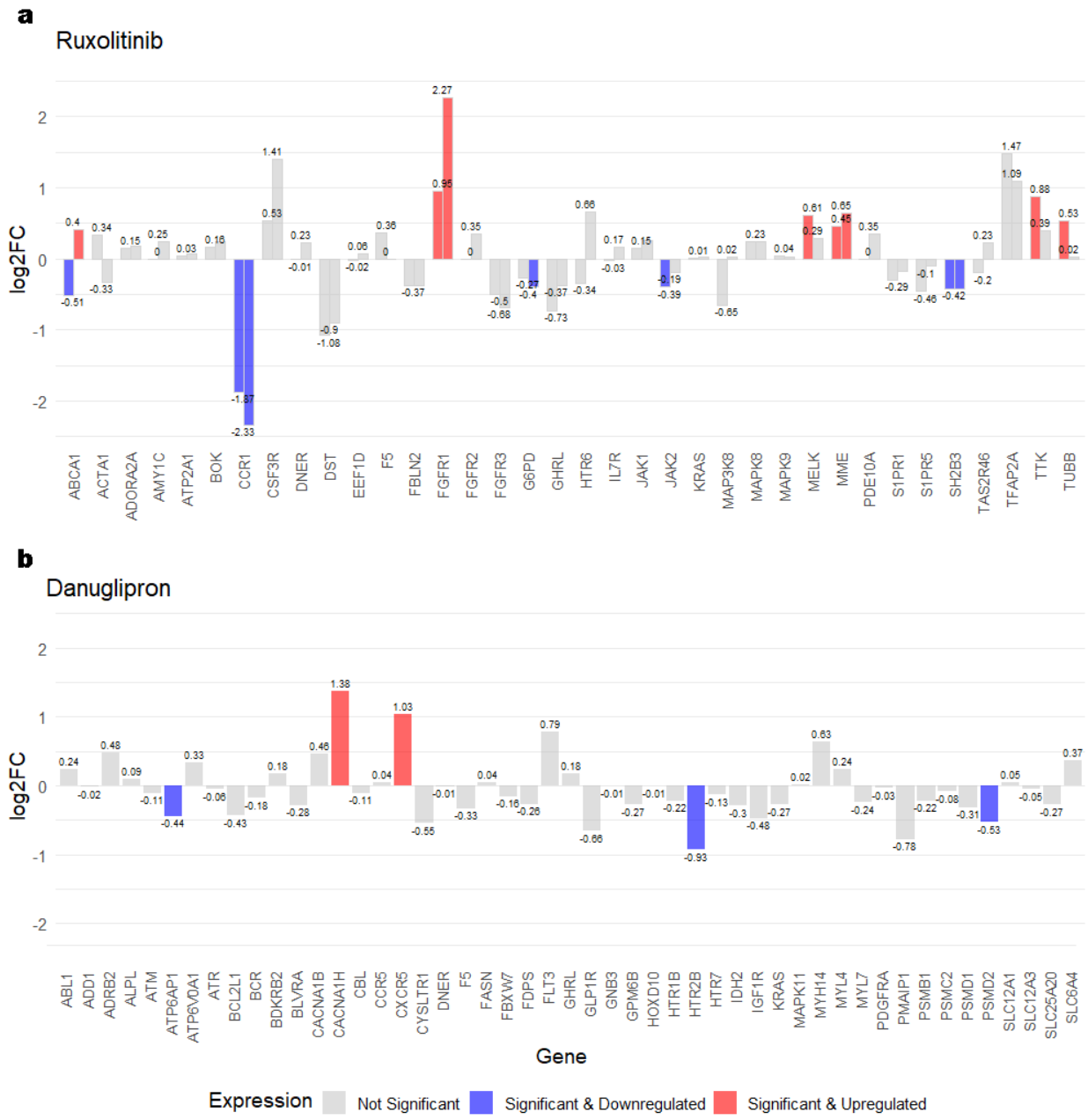


Figure 7. Differential expression (DE) profiles from external studies for assessing drug-induced perturbations within the fragment networks, reported as log2 fold change (log2FC). Only genes from networks with at least one non-zero log2FC are shown. (a) DE results comparing Day 3 (left bars) and Day 7 (right bars) after ruxolitinib treatment to control samples from Ph-like ALL PDX models. (b) DE results between semaglutide-treated

(a GLP-1R agonist similar to danuglipron) and control samples from a high-fat diet mouse model.

Discussion

In silico approaches in drug discovery have expanded the possibilities for new therapeutics identification. Nevertheless, there are a number of outstanding challenges for mapping the vast chemical space that by some estimates can even approach 10^{100} (Mullard, 2017; Zhang, 2025). In particular, not only the amount of data, but also the ability to uncover meaningful interactions are crucial for discovery success. Thus, while most current methods put emphasis on activity, ADME, *de novo* design, and synthesis planning, relatively little attention has been given to fragment profiling (Zhang, 2025). With the present work, we wanted to redefine perspectives on fragment-guided drug characterization and new molecule generation. As case studies, we selected ruxolitinib and danuglipron based on their therapeutic targets and network modulation potential, which enables broader clinical impact. Our developed model aligns potentially meaningful chemical groups without the restriction of these fragments being 'locked' in a compound, which could bias the search. Moreover, our proposed strategy significantly reduced the size of the search space to only 10,324 molecules. This contrasts sharply with commercial and other discovery campaigns that rely on compound libraries ranging from millions to billions (Lyu and Irwin, 2023). By taking advantage of fragment-informed chemical space mapping, we were able to generate new compounds that showed improved binding when compared to the parent molecule. In addition, fragment sampling for danuglipron yielded novel molecules that revealed a previously unrecognized residue important for binding and coordination. To complete the in silico pipeline, we incorporated biological space mapping to uncover new interactors that could be used for other modulator discovery or off-target effect prediction. Therefore, our analysis not only provided further insights into targeted therapy development but also outlined gene sets of particular importance for pathophysiology management.

To achieve our study goals, we introduced a new deep learning model in which the network's architecture is dynamically learned to identify the optimal latent space encoding through a Gaussian mixture (Fig. 1, Suppl. Fig. 2). Leveraging the expressivity of graph encoding (Kengkanna and Ohue, 2024), the first layer of molecular mapping was realized through Laplacian PE and GIN convolution layers (Suppl. Fig. 2). Specifically, the positional encoding and graph convolutions captured both chemical and topological interactions, allowing for richer fragment feature integration. Since our focus was on therapeutic entities, we demonstrated that this approach, even with a relatively small training set (~10,000 molecules), can lead to a discovery of completely new and structurally diverse

compounds with little similarity to any marketed drugs (Suppl. Fig. 6). This highlights the importance of developing new methods for drug design that do not solely rely on large chemical libraries (Gentile et al., 2022; Zhang, 2025).

We tested newly designed compounds by benchmarking them against parent molecules, namely ruxolitinib and danuglipron. Ligand docking is still challenging, owing to the fact that there are many considerations, ranging from flexibility to the scoring function's accuracy (Zhang et al., 2024). However, we wanted to explore how conformational 'snapshots' could accommodate molecules and whether this information could offer new insights. Thus, we employed two resolved structures for JAK2 and GLP-1R in which the parent molecule was bound, aiding with the docking site's identification. Our newly identified molecules for both target proteins demonstrated a number of new interactions that were not present in the original drug (Fig. 3). Consequently, these interactor groups could be used to establish modifications necessary for improved binding and new therapeutics design (Fig. 3, Table 1).

To further characterize novel compounds, we started our analysis with ruxolitinib (Suppl. Fig. 4, Suppl. document 2). Candidate 3 has several key anchoring points via dimethylamine. The same candidate also employed a number of polar and non-polar contacts via the three-ring heterocycle (Fig. 3). From an electrostatic and lipophilicity perspective (Fig. 4), we can see why these contacts are important in accommodating a molecule in the JAK2 binding pocket. Subsequently, these observations facilitated a pharmacophore search, which resulted in an even stronger binder (ZINC5203833) (Fig. 4). It is worth noting that new compound generation in combination with pharmacophore-based search can narrow down likely leads and guide targeted library search. In the case of danuglipron (Suppl. Fig. 5, Suppl. document 2), the binding pocket is much more constrained with limited degrees of freedom for molecule binding. Our analysis confirmed earlier findings that tryptophane (TRP33) is required for danuglipron signaling (Griffith et al., 2022) as binding is coordinated through this residue. However, we also identified phenylalanine (PHE381) which mediates anchoring for danuglipron in combination with TRP33, which was not highlighted previously (Griffith et al., 2022, Fig. 3). After comparing docking results of the candidates, we selected candidate 2 for pharmacophore search, as it exhibited not only a completely different docking pose, but also a more intricate quadruple coordination system involving TRP33, PHE381, lysine (LYS197), and its quinoline-derivative core (Fig. 3). This approach returned the ZINC32985857 molecule (Fig. 4, Table 1) with a markedly improved binding and the same coordination system of PHE381, TRP33, and compound's quinoline-like core. It is evident that both GLP-1R residues are necessary for optimal molecule accommodation within a binding pocket (Fig. 4). Thus, we not only identified new candidates with improved binding when benchmarked

against the parent molecule, but we also showed that an additional residue within the GLP-1R binding pocket is essential for binding. Importantly, such an exploratory refinement of candidate-to-pharmacophore demonstrates that we can efficiently identify novel therapeutic candidates and probe molecules to refine potential leads.

For the next step in our analysis, we investigated whether we could extend the chemical fragment space to the interactome to gain deeper insights into drug interactions, repurposing, or on/off-target effects. To achieve this, we analyzed the close neighborhood of our target fragments and sampled associated drugs. These drugs were utilized to capture additional targets for network construction. The mapped targets showed a significant enrichment for various interactome clusters (Fig. 5 and Fig. 6). This suggests that fragments carry their own 'functional meaning' that extends to biological processes. Additionally, functional similarities between fragments could be leveraged for therapeutic optimization, drug repurposing, or predicting potential side effects. Similarly, new functionalities or biomarkers could be explored by identifying novel interactors. Thus, ruxolitinib and danuglipron networks allowed us to better understand therapeutic profiles of these drugs and to theorize about newly designed modulator effects.

JAK2 inhibition involves a varied set of signaling events. Exploring beyond the nearest neighbors, such as KRAS, we discovered several key metabolic trajectories (Fig. 5). For example, ADORA2A stands out as a close neighbor to the JAK2 cluster, where it interacts with FGFR1 (Fig. 5, cluster 2 enrichment). Moreover, FGFR1 links ADORA2A and the adenosine signaling with KRAS and the JAK2 subclusters via the Rap1 pathway (KEGG database ID: hsa04015), which translates into cellular adhesion, migration, polarity, and survival. Targeting adenosine has been examined as a treatment strategy to block adenosine production and its receptor binding in oncological disorders (Leone et al., 2018; Deng and Flemming, 2022). However, recent clinical studies on JAK2 inhibition for anti-tumor effects have not established a link between JAK2/STAT3 and adenosine modulation effects, despite its potential effects in modifying tumor microenvironment and inducing immune reactions (Huang et al., 2022). Our expressome analysis revealed that FGFR1 was significantly upregulated when ALL xenografts were treated with ruxolitinib (Fig. 7). This could alter regulatory dynamics where, in oncological indications, FGFR1 might contribute to resistance together with sustained stimulation of ADORA2A. Nevertheless, in neurological disorders the FGFR1-ADORA2A axis could promote neuroplasticity and offer alternative treatment avenues (Latko et al., 2019; Fan et al., 2024). To our knowledge, there have been no reports in the clinical drug development setting identifying this link or exploring the potential of multi-modulation and different therapeutic outcomes, depending on the disease. Another member of JAK2 cluster, SH2B3, regulates IL7 receptor signaling (Cheng et al., 2016) and links inflammation with hypertension (Dale et al., 2016). Notably,

JAK2 has been implicated in salt-sensitive blood pressure and experimental studies established that JAK2 inhibition with ruxolitinib decreases pulmonary arterial pressure, restoring cardiac index (Saleem et al. 2024; Yerabolu et al., 2019). Additionally, ruxolitinib treatment in the ALL model downregulated both JAK2 and SH2B3 which points to another therapeutically relevant signaling branch (Fig. 7a). IL7R connects this interactor group with an adjacent small cluster consisting of sphingosine-1-phosphate receptors, that play a role in regulating inflammatory processes (Sun et al., 2024). Furthermore, experimental studies have indicated that S1PR1 activates the JAK2/STAT3 pathway via complex feedback loop (Lee et al., 2010). Thus, these observations underscore a broader impact of this network on several physiological functions and a complex interplay between different regulators, which could help establish new therapeutic strategies. Correspondingly, interactome analyses can flag potential side effects or parameters requiring clinical monitoring. For example, an initial drop in ABCA1 expression followed by upregulation (Fig. 7a) could have an effect on high density lipoprotein (HDL) levels. A study in mice demonstrated that ruxolitinib decreases cholesterol efflux, accelerating atherosclerosis (Tang et al., 2003; Dotan et al., 2022). Therefore, more research is needed to define clinically relevant alterations that could influence treatment protocols or pose a higher risk if other comorbidities are present. Taken together, a fragment-driven chemical and biological space mapping can help us better understand drug's role in both therapeutic and adverse outcomes.

Turning our attention to the GLP-1R network, we observe similarly enriched sub-clusters. The GLP-1R cluster (cluster 4) has several links with cluster 1 that maps to a number of cellular functions and cancer metabolism pathways. This is particularly noteworthy since effects of GLP-1R agonists on cancer biology have recently gained more attention (Ungvari et al., 2025). Our fragment-based network mapping highlights these links through several network trajectories, such as serotonergic, chemokine, and GPCR signaling (Fig. 6). For example, serotonin signaling has been implicated not only in immune cell regulation but also in oncological and metabolic disorders (Herr et al., 2017). Semaglutide-induced HTR2B downregulation in the pancreatic ductal adenocarcinoma model (Fig. 7b) (Cencioni et al., 2025) mirrors other studies demonstrating that HTR2B inhibition can suppress colorectal cancer tumor growth (Lee et al., 2024). The complex effects of serotonin are evident through other studies of HTR2B antagonism, which attenuated high-fat-diet-induced visceral adipose tissue inflammation, hepatic steatosis, and insulin resistance (Choi et al., 2021). Another therapeutically relevant group is formed by myosin (cluster 2), which includes the apelin signaling pathway (Fig. 6). Given that apelin is involved in the pathogenesis and protective effects in many diseases, such as coronary heart disease, heart failure, diabetes, and cancer, this network arrangement hints at modulation potential

of a number of parallel pathways (KEGG database ID: hsa04371) (Antushevich and Wójcik, 2018). Furthermore, GLP-1R has been documented for its cardioprotective effects; however, the exact mechanisms, for example in heart failure with preserved ejection fraction (HFpEF), remain to be elucidated (Hullon et al., 2025). A good example for such a therapeutic scope expansion is also the N-type calcium channel (CACNA1B) (cluster 12) which contributes to both cardio- and neuroprotective effects through the cervical ganglion (SCG) of the autonomic nervous system (Zhou et al., 2017). The role of SCG in cardiovascular diseases has been documented (Hernández-Ochoa et al, 2009; Cheng et al., 2018), and further studies are necessary to establish CACNA1B modulatory effects on cardiac tissue. One member of voltage-gated calcium channels, CACNA1H, was shown to be markedly upregulated in the pancreatic cancer study. This could provide further evidence for potential anti-cancer effects of GLP1-R agonists (Cencioni et al., 2025) (Fig. 7b) as CACNA1H inhibitors were found to reduce apoptosis induced by the ER stress and in oncological scenarios it could increase treatment sensitivity (Hu et al., 2019). Downregulation of PSMD2 and ATP6AP1 have a similar anti-proliferative effect (Fig. 7) (Matsuyama et al., 2011; Fei et al., 2024). As can be seen, by juxtaposing other medications, we can map a broader spectrum of beneficial functions across several therapeutic categories for GLP-1R. Currently, adverse effects, such as nausea, vomiting, or other gastrointestinal issues, associated with GLP-1R agonists have not been traced to specific pathway overstimulation, outside of the proposed GLP-1R+ neurons activation (Bettge et al., 2017; Drucker, 2025). In particular, none of the preclinical and clinical danuglipron studies have incorporated genetic profiling to address the complex signaling of GLP-1R (Saxena et al., 2021; Griffith et al., 2022; Saxena et al., 2023; Drucker, 2025). This lack of a more in-depth drug action characterization is evident across other GLP-1R clinical studies (Drucker, 2025). In our analysis, GLP-1R is in close proximity to amiloride-sensitive sodium channel subunit gamma (SCNN1G), which has been implicated in several disorders (Büsst et al., 2011). In particular, SCNN1G mediates luminal sodium and water diffusion through the apical membrane of epithelial cells, regulating electrolyte and blood pressure homeostasis (Büsst et al., 2011). Interestingly, it also has a role in taste perception (Dias et al., 2013). As a result, SCNN1G and other solute carriers in the network should be examined for adverse effects relating to hypertension, kidney tissue damage, gastrointestinal irritation, and electrolyte changes (Büsst et al., 2011). Moreover, our network analysis points to several pathways that can perturb liver metabolism leading to tissue injury (Fig. 6). For example, FoxO signaling pathway (KEGG database ID: hsa04068) is known to regulate liver stress response and hepatic homeostasis (Zhang et al., 2006). Several up- and down-stream members of this network, namely ATM, IGF1R, and KRAS, connect with the GLP-1R cluster (Fig. 6). Other examples include CCR5 which has been reported to play a role in chronic liver injury development (He et al., 2019). We also

reported an upregulation of CXCR5 which could signal B cell infiltration, and, in the liver, this can lead to a pro-inflammatory state (Fig. 7b) (Kim et al., 2025). These observations - in light of Pfizer's failed phase 3 clinical trial of danuglipron due to drug-induced liver injury - further emphasize missed opportunities in the drug's development (Pfizer, 2025). Therefore, integrating genetic parameters into clinical endpoint assessment could prevent clinical trial failure by tailoring studies for populations most likely to benefit from the treatment while also improving predictions for adverse effects (Drucker, 2025).

Both the JAK2 and GLP-1R networks share six common members (i.e., DNER, GHRL, GHSR, ALPL, F5, KRAS) accentuating their connection to pro-inflammatory states and metabolic regulation. Notably, the pharmacological intervention of the JAK-STAT pathway has emerged as a promising therapeutic approach for the treatment of metainflammation. This condition results in chronic and low-grade systemic inflammation and is characteristic of obesity, type 2 diabetes, and other disorders (Charles-Messance et al., 2020). Therefore, multi-modulatory approaches for complex diseases are increasingly being considered to achieve optimal treatment outcomes (Charles-Messance et al., 2020). For example, an experimental diabetes study in mice has demonstrated that liraglutide (GLP-1 analog) inhibited the JAK2/STAT3 pathway (Zitman-Gal et al., 2019). Other noteworthy regulatory networks in this context involve the neurological axis connecting both JAK2 and GLP-1R. Within the ruxolitinib network, SLC6A3, which is a dopamine transporter (Fig.5, cluster 2), links the JAK2 subnetwork with dopaminergic signaling. Dopamine is indispensable in regulating T cells, B cells, natural killer (NK) cells, macrophages, and dendritic cells (DCs) (Pia and Rodrigo, 2019; Matt and Gaskill, 2020). Consequently, dopamine agonists have been explored as a therapeutic strategy for rheumatoid arthritis (Lu et al., 2015). Similarly, the GLP-1R cluster extends into serotonergic signaling (Fig. 6, cluster 5), which also modulates many immune and metabolic responses (Martin et al., 2017; Chen et al., 2024). Given the discussed examples, it becomes evident that managing complex diseases could benefit from expanding the scope of therapeutic targets and their neighborhood mapping. Moreover, the growing number of multimorbidities will require improving personalized care to balance drug action (Skou et al., 2022). As a result, our introduced integrative approaches may offer better therapy matching and modelling of side effects.

Conclusion and future directions

The growing burden of complex diseases underscores the need for improved therapeutic solutions that employ diverse and targeted modulation strategies. While the advent of deep learning has significantly advanced drug discovery, current strategies are positioned to sift through large volumes of chemical data without optimizing alternative methods for

exploration. Consequently, enhancing drug discovery methods and characterization would significantly improve how we develop and use therapeutics in clinical settings. Our work not only offers a new perspective on targeted therapy development but also introduces an interactome mapping strategy, which takes advantage of molecular fragment profiling. We also developed a new self-adjusting neural network for this purpose that helped discover novel therapeutic candidates with improved binding. Furthermore, fragment-guided chemical space exploration enabled us to explore new metabolic trajectories. Our discussed clinical examples showcase how these insights could further guide drug repurposing efforts and advance the prediction of side effects. This is particularly important information to clinicians, who continuously refine treatment protocols for better outcomes and the minimization of adverse effects. Importantly, future research should expand on deep learning modelling methodologies as well as explore implications of broader network perturbations. Therefore, our work maps the first critical steps for this endeavor and highlights promising new avenues for drug discovery and clinical intervention development.

References

1. Drucker, D. J. (2018). Mechanisms of action and therapeutic application of glucagon-like peptide-1. *Cell metabolism*, *27*(4), 740-756.
2. Degen, J., Wegscheid-Gerlach, C., Zaliani, A., & Rarey, M. (2008). On the art of compiling and using'drug-like'chemical fragment spaces. *ChemMedChem*, *3*(10), 1503.
3. *Chugai Pharmaceutical Co. Ltd . Pyrazolopyridine derivative having GLP-1 receptor agonist effect. WO 2018056453, 2018.*
4. Dodington, D. W., Desai, H. R., & Woo, M. (2018). JAK/STAT–emerging players in metabolism. *Trends in Endocrinology & Metabolism*, *29*(1), 55-65.
5. Griffith, D. A., Edmonds, D. J., Fortin, J. P., Kalgutkar, A. S., Kuzmiski, J. B., Loria, P. M., ... & Tess, D. A. (2022). A small-molecule oral agonist of the human glucagon-like peptide-1 receptor. Journal of Medicinal Chemistry, 65(12), 8208-8226.
6. Sapre, M., Tremblay, D., Wilck, E., James, A., Leiter, A., Coltoff, A., ... & Gallagher, E. J. (2019). Metabolic effects of JAK1/2 inhibition in patients with myeloproliferative neoplasms. *Scientific Reports*, *9*(1), 16609.
7. Yin, Q., Wolkerstorfer, A., Lapid, O., Niessen, F. B., Van Zuijlen, P. P., & Gibbs, S. (2023). The JAK–STAT pathway in keloid pathogenesis: a systematic review with qualitative synthesis. *Experimental Dermatology*, *32*(5), 588-598.
8. Nauck, M. (2016). Incretin therapies: highlighting common features and differences in the modes of action of glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors. *Diabetes, Obesity and Metabolism*, *18*(3), 203-216.

9. Walton, K., Walker, K., Riddle, M., Koehn, B. H., Reff, J., Sagatys, E. M., ... & Betts, B. C. (2022). Dual JAK2/Aurora kinase A inhibition prevents human skin graft rejection by allo-inactivation and ILC2-mediated tissue repair. *American Journal of Transplantation*, *22*(3), 717-730.
10. Collotta, D., Franchina, M. P., Carlucci, V., & Collino, M. (2023). Recent advances in JAK inhibitors for the treatment of metabolic syndrome. *Frontiers in Pharmacology*, *14*, 1245535.
11. DGIdb 5.0: rebuilding the drug-gene interaction database for precision medicine and drug discovery platforms. Cannon M, Stevenson J, Stahl K, Basu R, Coffman A, Kiwala S, McMichael JF, Kuzma K, Morrisey D, Cotto KC, Mardis ER, Griffith OL, Griffith M, Wagner AH. Nucleic Acids Research. 2024 Jan 5; doi: https://doi.org/10.1093/nar/gkad1040. PMID: 37953380.
12. *Charles-Messance H., Mitchelson K. A. J., De Marco Castro E., Sheedy F. J., Roche H. M. (2020). Regulating metabolic inflammation by nutritional modulation. J. allergy Clin. Immunol. 146 (4), 706–720. 10.1016/j.jaci.2020.08.013*
13. Chen, M., Wei, Z., Huang, Z., Ding, B., & Li, Y. (2020, November). Simple and deep graph convolutional networks. In *International conference on machine learning* (pp. 1725-1735). PMLR.
14. Tonge PJ. Drug-Target Kinetics in Drug Discovery. ACS Chem Neurosci. 2018 Jan 17;9(1):29-39. doi: 10.1021/acschemneuro.7b00185. Epub 2017 Jul 14. PMID: 28640596; PMCID: PMC5767540.
15. Ryszkiewicz P, Malinowska B, Schlicker E. Polypharmacology: promises and new drugs in 2022. Pharmacol Rep. 2023 Aug;75(4):755-770. doi: 10.1007/s43440-023-00501-4. Epub 2023 Jun 6. PMID: 37278927; PMCID: PMC10243259.
16. Degen, J., Wegscheid-Gerlach, C., Zaliani, A., & Rarey, M. (2008). On the art of compiling and using'drug-like'chemical fragment spaces. ChemMedChem, 3 (10), 1503.
17. Schneider, G., & Clark, D. E. (2019). Automated de novo drug design: are we nearly there yet?. Angewandte Chemie International Edition, 58 (32), 10792-10803.
18. Lu, C., Liu, S., Shi, W., Yu, J., Zhou, Z., Zhang, X., ... & Wang, Y. (2022). Systemic evolutionary chemical space exploration for drug discovery. Journal of cheminformatics, 14 (1), 19. https://jcheminf.biomedcentral.com/articles/10.1186/s13321-022-00598-4
19. Leelananda, S. P., & Lindert, S. (2016). Computational methods in drug discovery. Beilstein journal of organic chemistry, 12 (1), 2694-2718.
20. Varolg¨une¸s, Y. B., Bereau, T., & Rudzinski, J. F. (2020). Interpretable embeddings from molecular simulations using Gaussian mixture variational autoencoders.

Machine Learning: Science and Technology, 1 (1), 015012. https://arxiv.org/pdf/1912.12175

21. Jang, E., Gu, S., & Poole, B. (2016). Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144. https://arxiv.org/pdf/1611.01144
22. Dwivedi, V. P., Joshi, C. K., Luu, A. T., Laurent, T., Bengio, Y., & Bresson, X. (2021). Benchmarking graph neural networks. Journal of Machine Learning Research, 24(43), 1-48.; https://arxiv.org/abs/2003.00982
23. Dwivedi, V. P., & Bresson, X. (2020). A generalization of transformer networks to graphs. arXiv preprint arXiv:2012.09699. https://arxiv.org/pdf/2012.09699
24. Dwivedi, V. P., Luu, A. T., Laurent, T., Bengio, Y., & Bresson, X. (2021). Graph neural networks with learnable structural and positional representations. arXiv preprint arXiv:2110.07875. https://arxiv.org/abs/2110.07875
25. Belkin, M.,& Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation, 15(6), 1373-1396.
26. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017, July). Neural message passing for quantum chemistry. In International conference on machine learning (pp. 1263-1272). PMLR. https://arxiv.org/pdf/1704.01212
27. Kensert, A., Desmet, G., & Cabooter, D. (2022). MolGraph: a Python package for the implementation of small molecular graphs and graph neural networks with TensorFlow and Keras. arXiv preprint arXiv:2208.09944. https://arxiv.org/abs/2208.09944
28. ChEMBL Database. Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, de Veij M, Ioannidis H, Lopez DM, Mosquera JF, Magarinos MP, Bosc N, Arcila R, Kizilören T, Gaulton A, Bento AP, Adasme MF, Monecke P, Landrum GA, Leach AR. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Res. 2024 Jan 5;52(D1):D1180-D1192. doi: 10.1093/nar/gkad1004.
29. Sakiyama, H., Fukuda, M., & Okuno, T. (2021). Prediction of blood-brain barrier penetration (bbbp) based on molecular descriptors of the free-form and in-blood-form datasets. *Molecules*, *26*(24), 7428.
30. ZINC15 Database. Sterling and Irwin, *J. Chem. Inf. Model, 2015 http://pubs.acs.org/doi/abs/10.1021/acs.jcim.5b00559*.
31. Knox C, Wilson M, Klinger CM, et al. *DrugBank 6.0: the DrugBank Knowledgebase for 2024.* Nucleic Acids Res. 2024 Jan 5;52(D1):D1265-D1275. doi: 10.1093/nar/gkad976.
32. PubMed: 37953279
33. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Annika GL, Fang T, Doncheva NT, Pyysalo S, Bork P‡, Jensen LJ‡, von Mering C‡. The STRING

database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023 Jan 6;51(D1):D638-646.

34. Bugnon M, Röhrig UF, Goullieux M, Perez MAS, Daina A, Michielin O, Zoete V. SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. Nucleic Acids Res. 2024, 52 (W1), W324-W332. DOI: 10.1093/nar/gkae300.
35. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J. Chem. Inf. Model., 2021, 61 (8), 3891–3898, DOI: 10.1021/acs.jcim.1c00203
36. PDB-101: Educational resources supporting molecular explorations through biology and medicine. Christine Zardecki, Shuchismita Dutta, David S. Goodsell, Robert Lowe, Maria Voigt, Stephen K. Burley. (2022) *Protein Science* **31**: 129-140 https://doi.org/10.1002/pro.4200
37. Bragina, ME., Daina, A., Perez, MAS., Michielin, O. & Zoete, V. SwissSimilarity 2021 Web Tool: Novel Chemical Libraries and Additional Methods for an Enhanced Ligand-Based Virtual Screening Experience., Int. J. Mol. Sci, 2022, 23(2), 811.
38. Daina, A., Michielin, O., & Zoete, V. J. D. O. S. (2017). *SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7: 42717*.
39. The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC.
40. UCSF ChimeraX: Tools for structure building and analysis. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. Protein Sci. 2023 Nov;32(11):e4792.
41. Bemis, G. W., & Murcko, M. A. (1996). The properties of known drugs. 1. Molecular frameworks. *Journal of medicinal chemistry*, *39*(15), 2887-2893.
42. Sunseri, J., & Koes, D. R. (2016). Pharmit: interactive exploration of chemical space. *Nucleic acids research*, *44*(W1), W442-W448.
43. Dwivedi, V. P., Joshi, C. K., Luu, A. T., Laurent, T., Bengio, Y., & Bresson, X. (2023). Benchmarking graph neural networks. *Journal of Machine Learning Research*, *24*(43), 1-48.
44. Kengkanna, A., & Ohue, M. (2024). Enhancing property and activity prediction and interpretation using multiple molecular graph representations with MMGX. *Communications Chemistry*, *7*(1), 74.
45. Kensert, A., Desmet, G., & Cabooter, D. (2025). MolGraph: a Python package for the implementation of molecular graphs and graph neural networks with TensorFlow and Keras. *Journal of Computer-Aided Molecular Design*, *39*(1), 1-11.

46. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K, Chun M, Li P, Gohara DW, Dolinsky T, Konecny R, Koes DR, Nielsen JE, Head-Gordon T, Geng W, Krasny R, Wei GW, Holst MJ, McCammon JA, Baker NA. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 2018 Jan;27(1):112-128. doi: 10.1002/pro.3280. Epub 2017 Oct 24. PMID: 28836357; PMCID: PMC5734301.
47. Broto, P., Moreau, G., & Vandycke, C. (1984). Molecular structures: perception, autocorrelation descriptor and sar studies: system of atomic contributions for the calculation of the n-octanol/water partition coefficients. *European journal of medicinal chemistry*, *19*(1), 71-78.
48. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. and Ishiguro-Watanabe, M.; KEGG: biological systems database as a model of the real world. Nucleic Acids Res. 53, D672-D677 (2025). [pubmed] [doi]
49. Cheng, Y., Chikwava, K., Wu, C., Zhang, H., Bhagat, A., Pei, D., ... & Tong, W. (2016). LNK/SH2B3 regulates IL-7 receptor signaling in normal and malignant B-progenitors. *The Journal of clinical investigation*, *126*(4), 1267-1281.
50. Ungvari, Z., Bartha, Á., Ungvari, A., Fekete, M., Bianchini, G., & Győrffy, B. (2025). Prognostic impact of glucagon-like peptide-1 receptor (GLP1R) expression on cancer survival and its implications for GLP-1R agonist therapy: an integrative analysis across multiple tumor types. *GeroScience*, 1-15.
51. Herr N, Bode C, Duerschmied D. The effects of serotonin in immune cells. *Front Cardiovasc Med*. 2017; 4: 48.
52. Zhang, K., Yang, X., Wang, Y., Yu, Y., Huang, N., Li, G., ... & Yang, S. (2025). Artificial intelligence in drug development. *Nature Medicine*, 1-15.
53. Mullard, A. The drug-maker's guide to the galaxy. *Nature* **549**, 445–447 (2017).
54. Liu, R., Wei, L. & Zhang, P. A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data. *Nat. Mach. Intell.* **3**, 68–75 (2021,a).
55. Liu, R. et al. Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature* **592**, 629–633 (2021,b).
56. Gentile, F. et al. Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking. *Nat. Protoc.* **17**, 672–697 (2022)
57. Zhang X, Shen C, Zhang H, Kang Y, Hsieh CY, Hou T. Advancing Ligand Docking through Deep Learning: Challenges and Prospects in Virtual Screening. Acc Chem Res. 2024 May 21;57(10):1500-1509. doi: 10.1021/acs.accounts.4c00093. Epub 2024 Apr 5. PMID: 38577892.
58. Hullon, D., Subeh, G. K., Volkova, Y., Janiec, K., Trach, A., & Mnevets, R. (2025). The role of glucagon-like peptide-1 receptor (GLP-1R) agonists in enhancing endothelial

function: a potential avenue for improving heart failure with preserved ejection fraction (HFpEF). *Cardiovascular Diabetology*, *24*, 70.

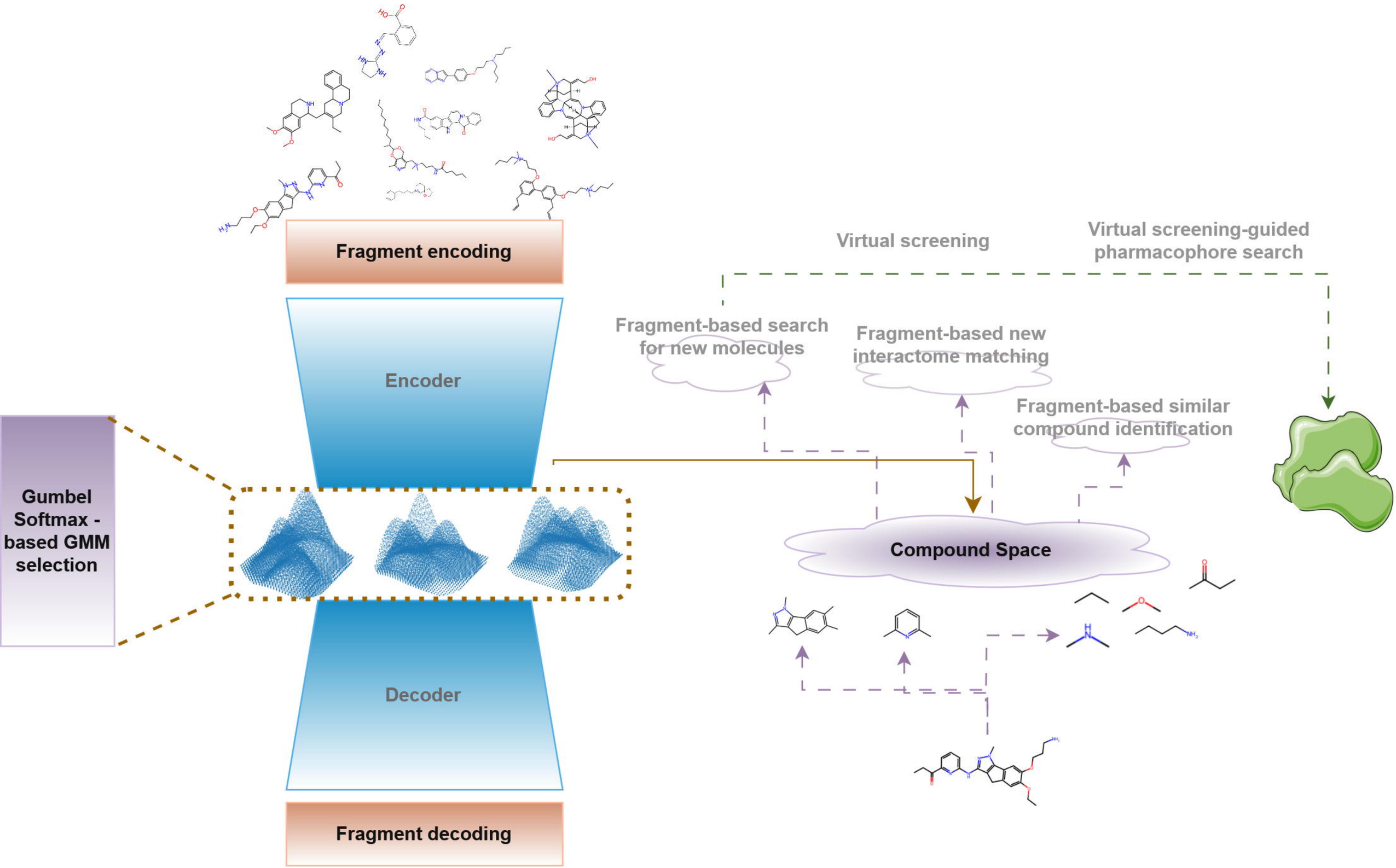
59. Bettge K, Kahle M, Abd El Aziz MS, Meier JJ, Nauck MA. Occurrence of nausea, vomiting and diarrhoea reported as adverse events in clinical trials studying glucagon-like peptide-1 receptor agonists: A systematic analysis of published clinical trials. Diabetes Obes Metab. 2017 Mar;19(3):336-347. doi: 10.1111/dom.12824. Epub 2016 Dec 19. PMID: 27860132.
60. Hernández-Ochoa E. O., Prosser B. L., Wright N. T., Contreras M., Weber D. J., Schneider M. F. (2009). Augmentation of Cav1 channel current action potential duration after uptake of S100A1 in sympathetic ganglion neurons. Am. J. Physiol. Cell Physiol. 297 C955–C970. 10.1152/ajpcell.00140.2009
61. Cheng L, Wang X, Liu T, Tse G, Fu H, Li G. Modulation of Ion Channels in the Superior Cervical Ganglion Neurons by Myocardial Ischemia and Fluvastatin Treatment. Front Physiol. 2018 Sep 10;9:1157. doi: 10.3389/fphys.2018.01157. PMID: 30246810; PMCID: PMC6139347.
62. Zhou X., Wang W., Zhang S., Wang X., Tang Z., Gu J., et al. (2017). CACNA1B (Cav2.2) overexpression and its association with clinicopathologic characteristics and unfavorable prognosis in non-small cell lung cancer. Dis. Markers 2017:6136401. 10.1155/2017/6136401
63. Grafone, T., Palmisano, M., Nicci, C., & Storti, S. (2012). An overview on the role of FLT3-tyrosine kinase receptor in acute myeloid leukemia: biology and treatment. *Oncology reviews*, *6*(1), e8.
64. Dale BL, Madhur MS. Linking inflammation and hypertension via LNK/SH2B3. Curr Opin Nephrol Hypertens. 2016 Mar;25(2):87-93. doi: 10.1097/MNH.0000000000000196. PMID: 26717315; PMCID: PMC4763980.
65. Saleem, M., Aden, L. A., Mutchler, A. L., Basu, C., Ertuglu, L. A., Sheng, Q., ... & Kirabo, A. (2024). Myeloid-specific JAK2 contributes to inflammation and salt sensitivity of blood pressure. *Circulation Research*, *135*(9), 890-909.
66. Leone R. D., Emens L. A. (2018). Targeting Adenosine for Cancer Immunotherapy. J. Immunotherapy Cancer 6 (1), 57. 10.1186/s40425-018-0360-8
67. Deng J, Fleming JB. Inflammation and Myeloid Cells in Cancer Progression and Metastasis. Front Cell Dev Biol. 2022 Jan 21;9:759691. doi: 10.3389/fcell.2021.759691. PMID: 35127700; PMCID: PMC8814460.
68. Huang, B., Lang, X., & Li, X. (2022). The role of IL-6/JAK2/STAT3 signaling pathway in cancers. *Frontiers in oncology*, *12*, 1023177.
69. Sun, G., Wang, B., Wu, X., Cheng, J., Ye, J., Wang, C., ... & Liu, X. (2024). How do sphingosine-1-phosphate affect immune cells to resolve inflammation?. *Frontiers in Immunology*, *15*, 1362459.

70. Lee, H., Deng, J., Kujawski, M., Yang, C., Liu, Y., Herrmann, A., ... & Yu, H. (2010). STAT3-induced S1PR1 expression is crucial for persistent STAT3 activation in tumors. *Nature medicine*, *16*(12), 1421-1428.
71. Zitman-Gal, T., Einbinder, Y., Ohana, M., Katzav, A., Kartawy, A., & Benchetrit, S. (2019). Effect of liraglutide on the Janus kinase/signal transducer and transcription activator (JAK/STAT) pathway in diabetic kidney disease in db/db mice and in cultured endothelial cells. *Journal of Diabetes*, *11*(8), 656-664.
72. M.V. Pia, P. Rodrigo. Targeting the dopaminergic system in autoimmunity. J. Neuroimmune Pharmacol. (2019), 10.1007/s11481-019-09834-5
73. Matt, S. M., & Gaskill, P. J. (2020). Where is dopamine and how do immune cells see it?: dopamine-mediated immune cell function in health and disease. Journal of Neuroimmune Pharmacology, 15, 114-164.
74. Lu JH, Liu YQ, Deng QW, Peng YP, Qiu YH. Dopamine D2 Receptor Is Involved in Alleviation of Type II Collagen-Induced Arthritis in Mice. *BioMed Res Int* (2015) 2015:496759. doi: 10.1155/2015/496759
75. Chen, L., Huang, S., Wu, X., He, W., & Song, M. (2024). Serotonin signalling in cancer: Emerging mechanisms and therapeutic opportunities. *Clinical and Translational Medicine*, *14*(7), e1750.
76. Martin, A. M., Young, R. L., Leong, L., Rogers, G. B., Spencer, N. J., Jessup, C. F., & Keating, D. J. (2017). The diverse metabolic roles of peripheral serotonin. *Endocrinology*, *158*(5), 1049-1063.
77. Drucker, D. J. (2025). GLP-1-based therapies for diabetes, obesity and beyond. *Nature Reviews Drug Discovery*, 1-20.
78. Latko M, Czyrek A, Porębska N, Kucińska M, Otlewski J, Zakrzewska M, Opaliński Ł. Cross-Talk between Fibroblast Growth Factor Receptors and Other Cell Surface Proteins. Cells. 2019 May 14;8(5):455. doi: 10.3390/cells8050455. PMID: 31091809; PMCID: PMC6562592.
79. Fan S, Chen Y, Wang W, Xu W, Tian M, Liu Y, Zhou Y, Liu D, Xia Q, Dong L. Pharmacological and Biological Targeting of FGFR1 in Cancer. Curr Issues Mol Biol. 2024 Nov 18;46(11):13131-13150. doi: 10.3390/cimb46110783. PMID: 39590377; PMCID: PMC11593329.
80. Tang C, Vaughan AM, Oram JF. Janus kinase 2 modulates the apolipoprotein interactions with ABCA1 required for removing cellular cholesterol. J Biol Chem. 2004 Feb 27;279(9):7622-8. doi: 10.1074/jbc.M312571200. Epub 2003 Dec 9. PMID: 14668333.
81. Choi WG, Choi W, Oh TJ, Cha HN, Hwang I, Lee YK, Lee SY, Shin H, Lim A, Ryu D, Suh JM, Park SY, Choi SH, Kim H. Inhibiting serotonin signaling through HTR2B in visceral adipose tissue improves obesity-related insulin resistance. J Clin Invest.

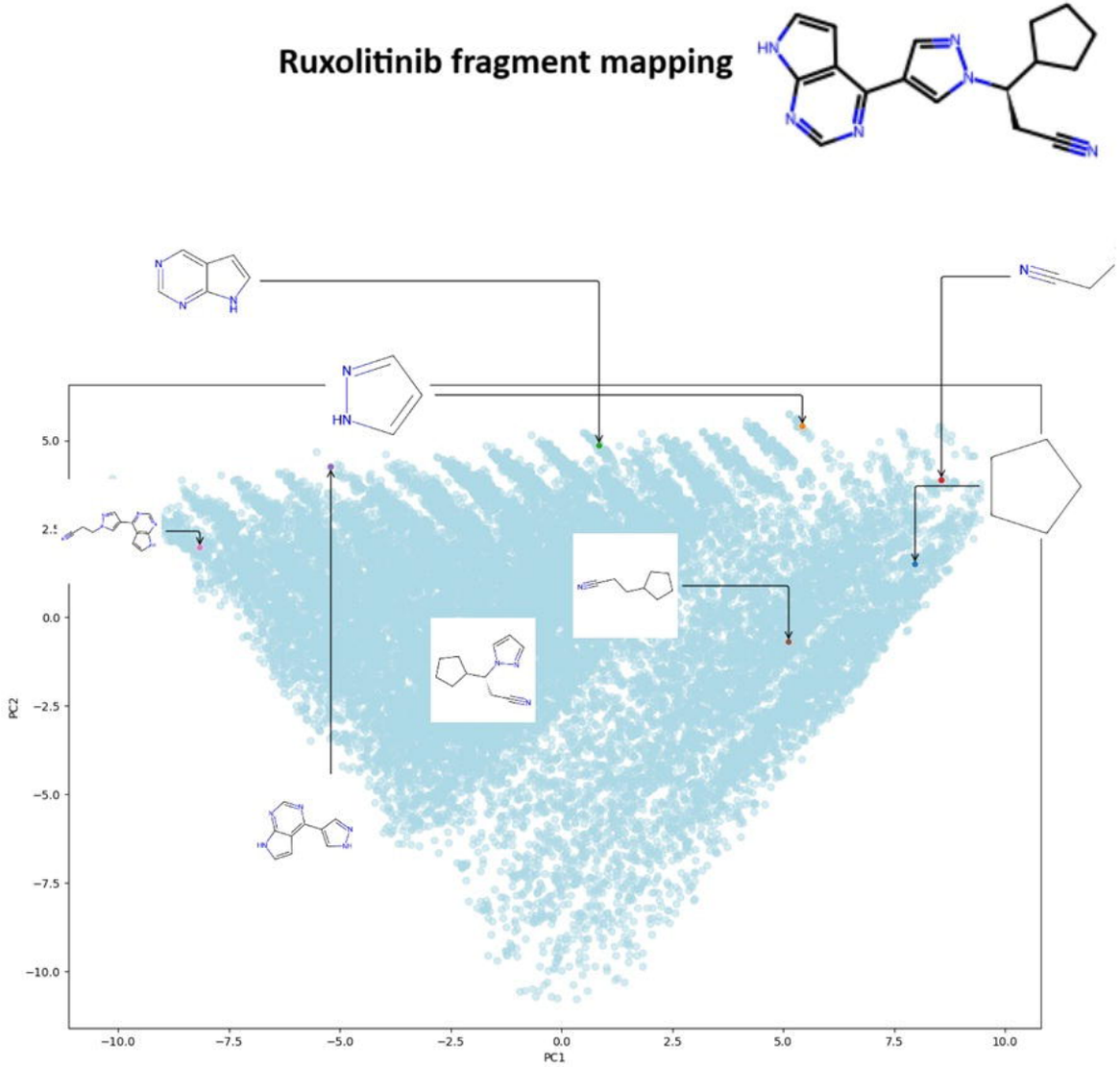
2021 Dec 1;131(23):e145331. doi: 10.1172/JCI145331. PMID: 34618686; PMCID: PMC8631597.
82. Lee JY, Park S, Park EJ, Pagire HS, Pagire SH, Choi BW, Park M, Fang S, Ahn JH, Oh CM. Inhibition of HTR2B-mediated serotonin signaling in colorectal cancer suppresses tumor growth through ERK signaling. Biomed Pharmacother. 2024 Oct;179:117428. doi: 10.1016/j.biopha.2024.117428. Epub 2024 Sep 9. PMID: 39255737.
83. Hu J, Wu Q, Wang Z, Hong J, Chen R, Li B, Hu Z, Hu X, Zhang M. Inhibition of CACNA1H attenuates doxorubicin-induced acute cardiotoxicity by affecting endoplasmic reticulum stress. Biomed Pharmacother. 2019 Dec;120:109475. doi: 10.1016/j.biopha.2019.109475. Epub 2019 Sep 30. PMID: 31580970.
84. Fei Y, Yan X, Liang M, Zhou S, Xu D, Li L, Xu W, Song Y, Zhu Z, Zhang J. Lysosomal gene ATP6AP1 promotes doxorubicin resistance via up-regulating autophagic flux in breast cancer. Cancer Cell Int. 2024 Dec 3;24(1):394. doi: 10.1186/s12935-024-03579-9. PMID: 39627767; PMCID: PMC11616228.
85. Matsuyama Y, Suzuki M, Arima C, Huang QM, Tomida S, Takeuchi T, Sugiyama R, Itoh Y, Yatabe Y, Goto H, Takahashi T. Proteasomal non-catalytic subunit PSMD2 as a potential therapeutic target in association with various clinicopathologic features in lung adenocarcinomas. Mol Carcinog. 2011 Apr;50(4):301-9. doi: 10.1002/mc.20632. PMID: 21465578.
86. Kim NH, Kim MY, Yang YM, Jeong WI, Lee HW, Kim W, Kang SG, Han YH. Bacterial components-driven intrahepatic CXCR5^{hi} B cells are important population for MASH progression through inducing inflammation. FASEB J. 2025 Jan 31;39(2):e70322. doi: 10.1096/fj.202401256R. PMID: 39812617.
87. Cencioni C, Malatesta S, Vigiano Benedetti V, Licursi V et al. The GLP-1R agonist semaglutide reshapes pancreatic cancer associated fibroblasts reducing collagen proline hydroxylation and favoring T lymphocyte infiltration. J Exp Clin Cancer Res 2025 Jan 20;44(1):18. PMID: 39828692
88. Dotan I, Yang J, Ikeda J, Roth Z, Pollock-Tahiri E, Desai H, Sivasubramaniyam T, Rehal S, Rapps J, Li YZ, Le H, Farber G, Alchami E, Xiao C, Karim S, Gronda M, Saikali MF, Tirosh A, Wagner KU, Genest J, Schimmer AD, Gupta V, Minden MD, Cummins CL, Lewis GF, Robbins C, Jongstra-Bilen J, Cybulsky M, Woo M. Macrophage Jak2 deficiency accelerates atherosclerosis through defects in cholesterol efflux. Commun Biol. 2022 Feb 15;5(1):132. doi: 10.1038/s42003-022-03078-5. PMID: 35169231; PMCID: PMC8847578.
89. Yerabolu D, Weiss A, Kojonazarov B, Boehm M, Schlueter BC, Ruppert C, Günther A, Jonigk D, Grimminger F, Ghofrani HA, Seeger W, Weissmann N, Schermuly RT. Targeting Jak-Stat Signaling in Experimental Pulmonary Hypertension. Am J Respir Cell Mol Biol. 2021 Jan;64(1):100-114. doi: 10.1165/rcmb.2019-0431OC. PMID: 33052714.

90. He, Z., Wei, G., Li, N., Niu, M., Gong, S., Wu, G., ... & Chen, P. (2019). CCR2 and CCR5 promote diclofenac-induced hepatotoxicity in mice. *Naunyn-Schmiedeberg's Archives of Pharmacology*, *392*, 287-297.
91. Zhang W, Patil S, Chauhan B, Guo S, Powell DR, Le J, Klotsas A, Matika R, Xiao X, Franks R, Heidenreich KA, Sajan MP, Farese RV, Stolz DB, Tso P, Koo SH, Montminy M, Unterman TG. FoxO1 regulates multiple metabolic pathways in the liver: effects on gluconeogenic, glycolytic, and lipogenic gene expression. J Biol Chem. 2006 Apr 14;281(15):10105-17. doi: 10.1074/jbc.M600272200. Epub 2006 Feb 21. PMID: 16492665.
92. Antushevich H, Wójcik M. Review: Apelin in disease. Clin Chim Acta. 2018 Aug;483:241-248. doi: 10.1016/j.cca.2018.05.012. Epub 2018 May 8. PMID: 29750964.
93. Büsst CJ, Bloomer LD, Scurrah KJ, Ellis JA, Barnes TA, Charchar FJ, Braund P, Hopkins PN, Samani NJ, Hunt SC, Tomaszewski M, Harrap SB. The epithelial sodium channel γ-subunit gene and blood pressure: family based association, renal gene expression, and physiological analyses. Hypertension. 2011 Dec;58(6):1073-8. doi: 10.1161/HYPERTENSIONAHA.111.176370. Epub 2011 Oct 17. PMID: 22006290; PMCID: PMC3220739.
94. Dias AG, Rousseau D, Duizer L, Cockburn M, Chiu W, Nielsen D, El-Sohemy A. Genetic variation in putative salt taste receptors and salt taste perception in humans. Chem Senses. 2013 Feb;38(2):137-45. doi: 10.1093/chemse/bjs090. Epub 2012 Nov 1. PMID: 23118204.
95. Saxena, A. R., Gorman, D. N., Esquejo, R. M., Bergman, A., Chidsey, K., Buckeridge, C., ... & Kim, A. M. (2021). Danuglipron (PF-06882961) in type 2 diabetes: a randomized, placebo-controlled, multiple ascending-dose phase 1 trial. *Nature medicine*, *27*(6), 1079-1087.
96. Saxena, A. R., Frias, J. P., Brown, L. S., Gorman, D. N., Vasas, S., Tsamandouras, N., & Birnbaum, M. J. (2023). Efficacy and safety of oral small molecule glucagon-like peptide 1 receptor agonist danuglipron for glycemic control among patients with type 2 diabetes: a randomized clinical trial. *JAMA network open*, *6*(5), e2314493-e2314493.
97. Mohammed S, Bindu A, Viswanathan A, Harikumar KB. Sphingosine 1-phosphate signaling during infection and immunity. Prog Lipid Res. 2023 Nov;92:101251. doi: 10.1016/j.plipres.2023.101251. Epub 2023 Aug 24. PMID: 37633365.
98. Lyu J, Irwin JJ, Shoichet BK. Modeling the expansion of virtual screening libraries. Nat Chem Biol. 2023 Jun;19(6):712-718. doi: 10.1038/s41589-022-01234-w. Epub 2023 Jan 16. PMID: 36646956; PMCID: PMC10243288.
99. Pfizer Provides Update on Oral GLP-1 Receptor Agonist Danuglipron. Pfizer Inc., 2025, April 14; https://www.pfizer.com/news/press-release/press-release-detail/pfizer-provides-update-oral-glp-1-receptor-agonist
100. Ochiai, T., Inukai, T., Akiyama, M. et al. Variational autoencoder-based chemical latent space for large molecular structures with 3D complexity. Commun Chem 6, 249 (2023). https://doi.org/10.1038/s42004-023-01054-6
101. Ding YY, Sussman JH, Madden K, Loftus JP et al. Targeting senescent stemlike subpopulations in Philadelphia chromosome-like acute lymphoblastic leukemia. Blood 2025 Mar 13;145(11):1195-1210. PMID: 39774844

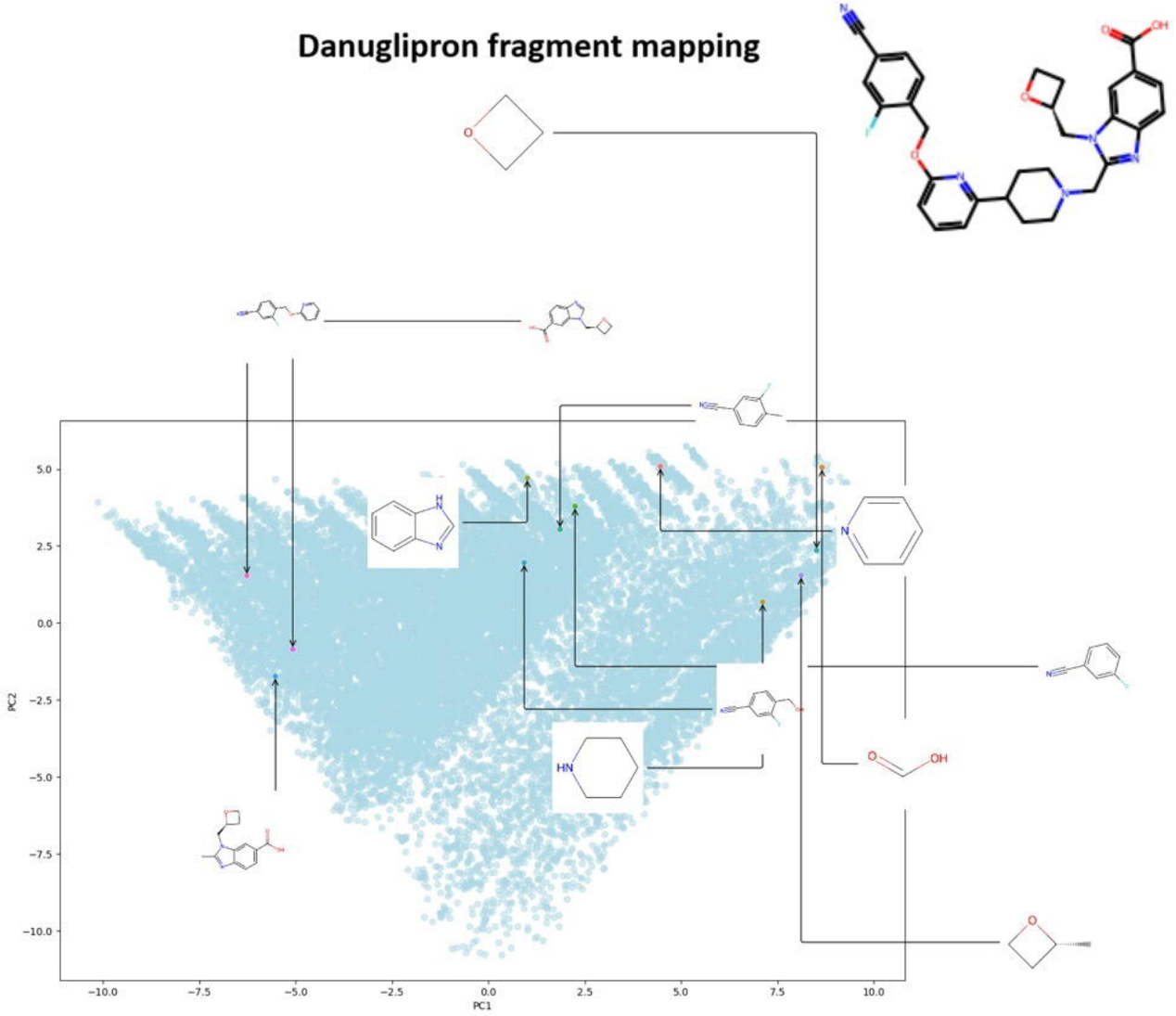
102. Skou ST, Mair FS, Fortin M, Guthrie B, Nunes BP, Miranda JJ, Boyd CM, Pati S, Mtenga S, Smith SM. Multimorbidity. Nat Rev Dis Primers. 2022 Jul 14;8(1):48. doi: 10.1038/s41572-022-00376-4. PMID: 35835758; PMCID: PMC7613517.
103. Anderson B, Rosston P, Ong HW, Hossain MA, Davis-Gilbert ZW, Drewry DH. How many kinases are druggable? A review of our current understanding. Biochem J. 2023 Aug 30;480(16):1331-1363. doi: 10.1042/BCJ20220217. PMID: 37642371; PMCID: PMC10586788.
104. Sriram K, Insel PA. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? Mol Pharmacol. 2018 Apr;93(4):251-258. doi: 10.1124/mol.117.111062. Epub 2018 Jan 3. PMID: 29298813; PMCID: PMC5820538.

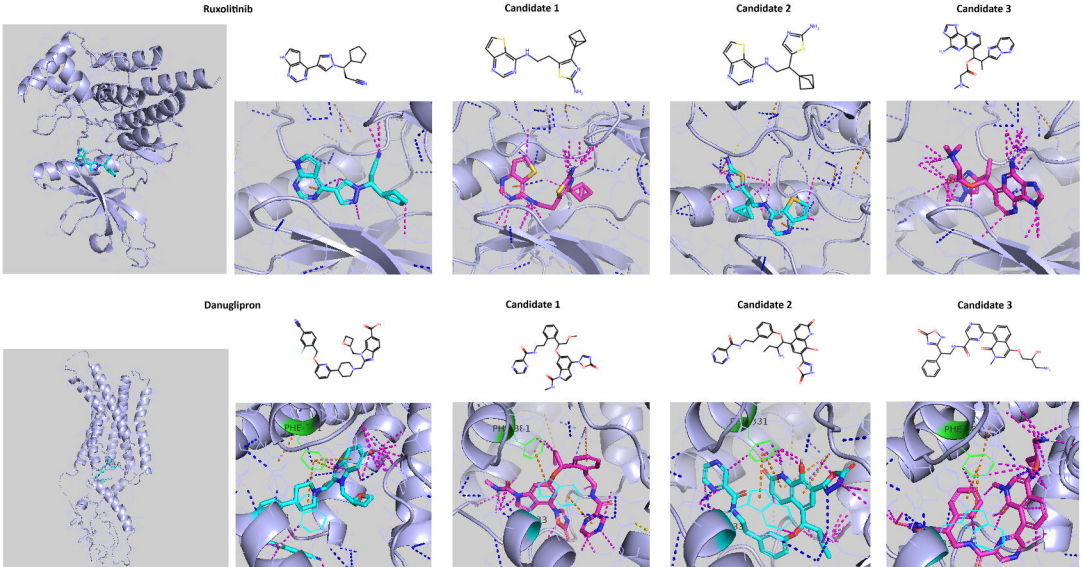

Fragment encoding
Encoder
Gumbel Softmax - based GMM selection
Decoder
Fragment decoding
Virtual screening
Virtual screening-guided pharmacophore search
Fragment-based search for new molecules
Fragment-based new interactome matching
Fragment-based similar compound identification
Compound Space

Ruxolitinib fragment mapping

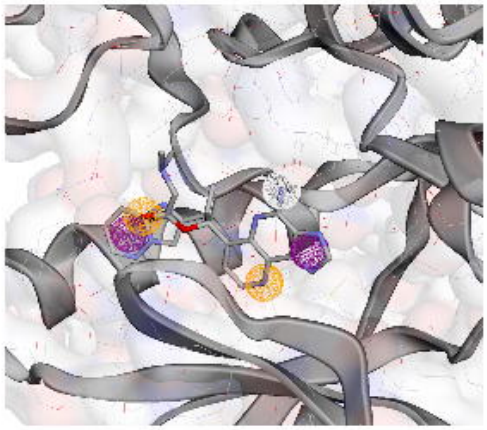


Danuglipron fragment mapping

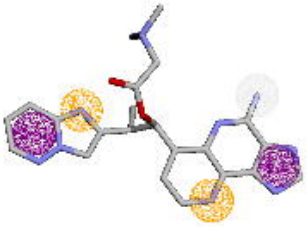



Ruxolitinib
Candidate 1
Candidate 2
Candidate 3
Danuglipron
Candidate 1
Candidate 2
Candidate 3

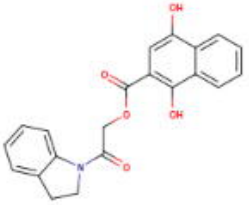
JAK2 inhibitor search

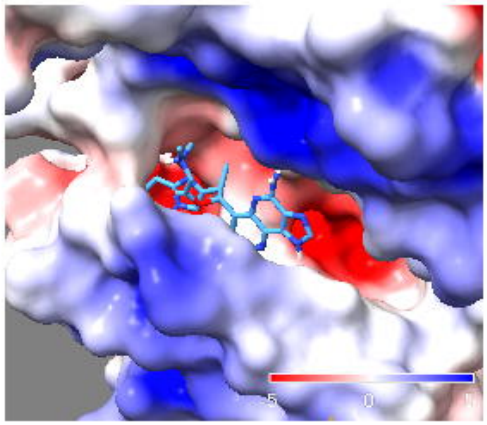


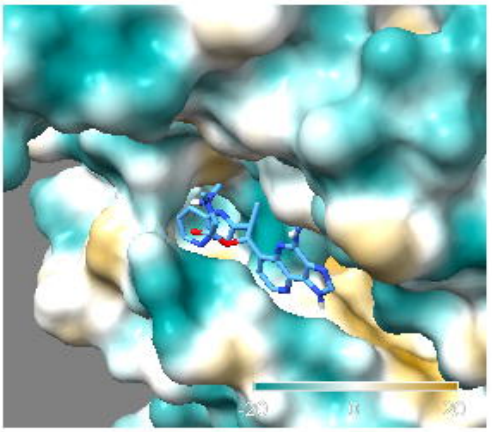
Candidate 3-based pharmacophore

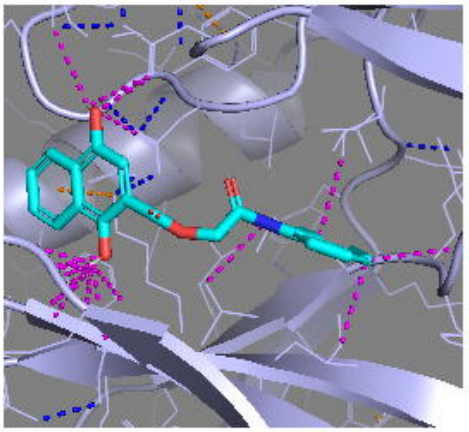


ZINC5203833

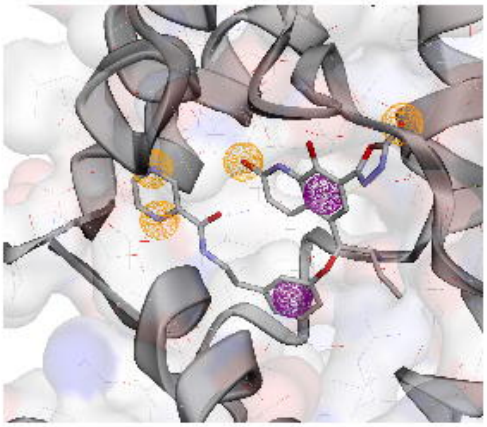




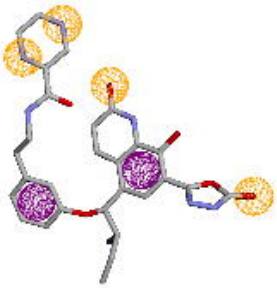




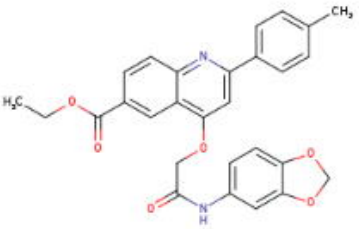
GLP1-R inhibitor search

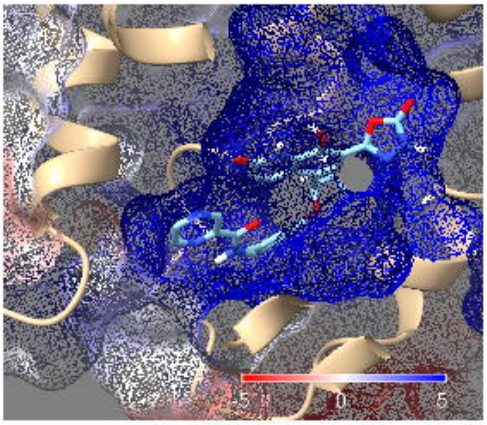


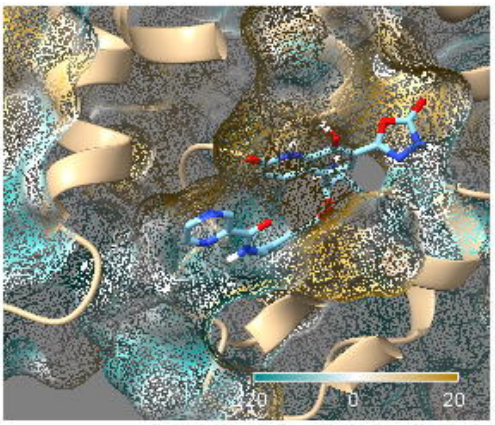
Candidate 2-based pharmacophore

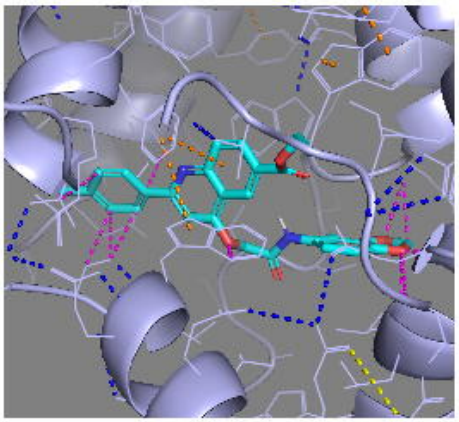


ZINC32985857

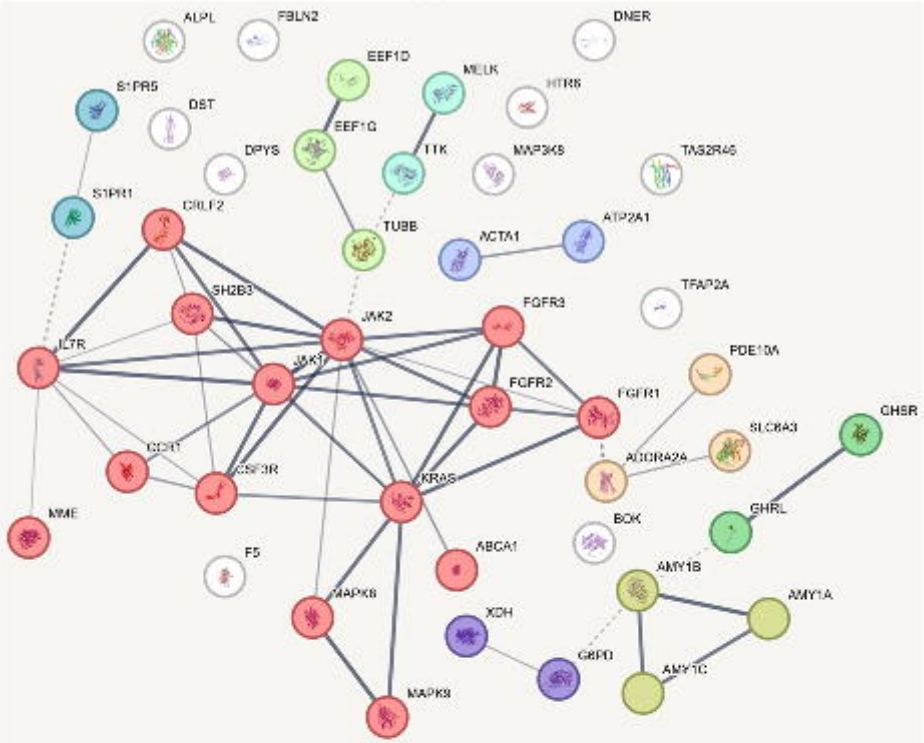








Ruxolitinib fragment network

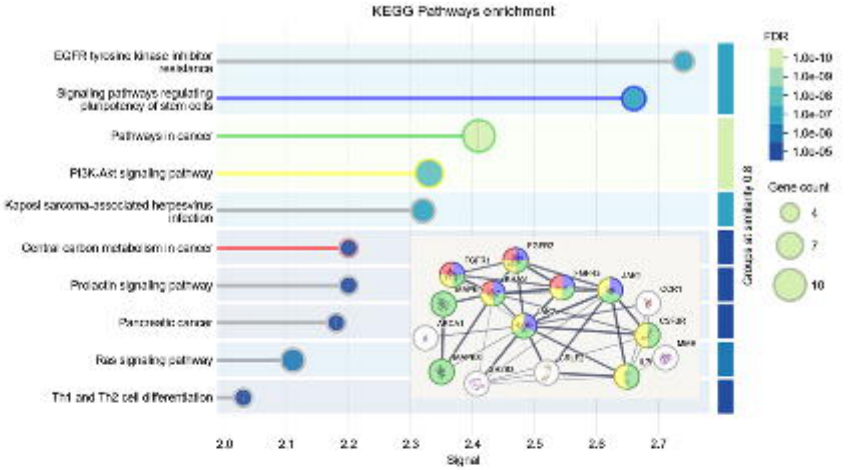


number of nodes: 45
number of edges: 56

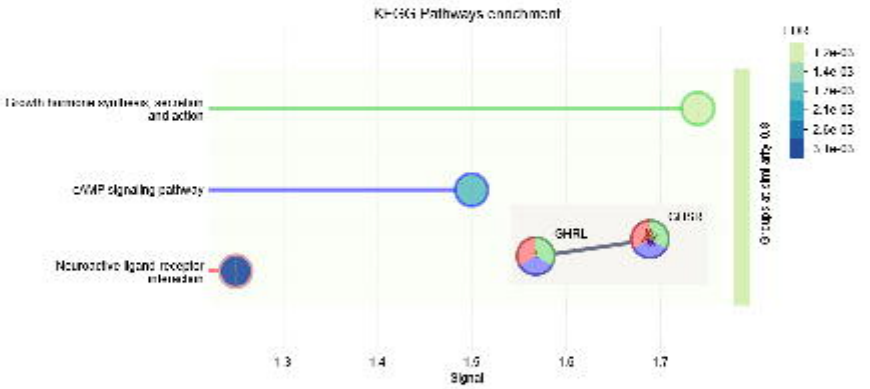
expected number of edges: 27
PPI enrichment p-value: 5.77e-07

color	cluster Id	gene count	description
	Cluster 1	15	1. EGFR tyrosine kinase inhibitor resistance 2. Signaling by CSF3 (G-CSF) 3. Tyrosine kinase, catalytic domain
	Cluster 2	3	Prepulse inhibition
	Cluster 3	3	Aamy_C
	Cluster 4	3	Elongation factor
	Cluster 5	2	1. Positive regulation of small intestinal transit 2. Type 1 melanocortin receptor binding, and Melanocortin/ACTH receptor
	Cluster 6	2	MELK, TTK
	Cluster 7	2	1. Lysosphingolipid and LPA receptors 2. sphingosine-1-phosphate receptor activity
	Cluster 8	2	ACTA1, ATP2A1
	Cluster 9	2	G6PD, XDH

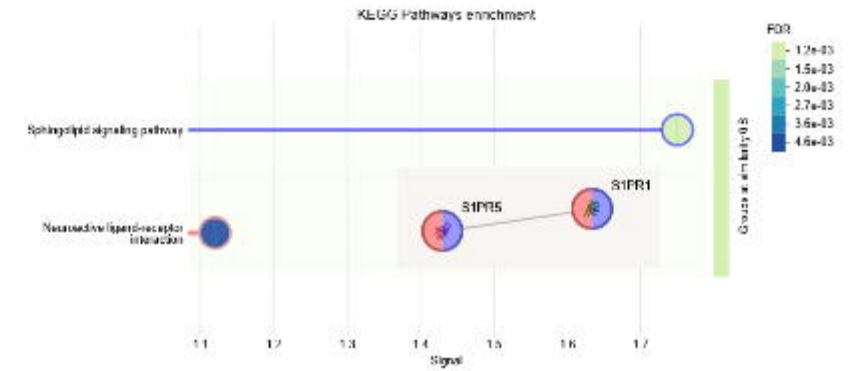
Cluster 1



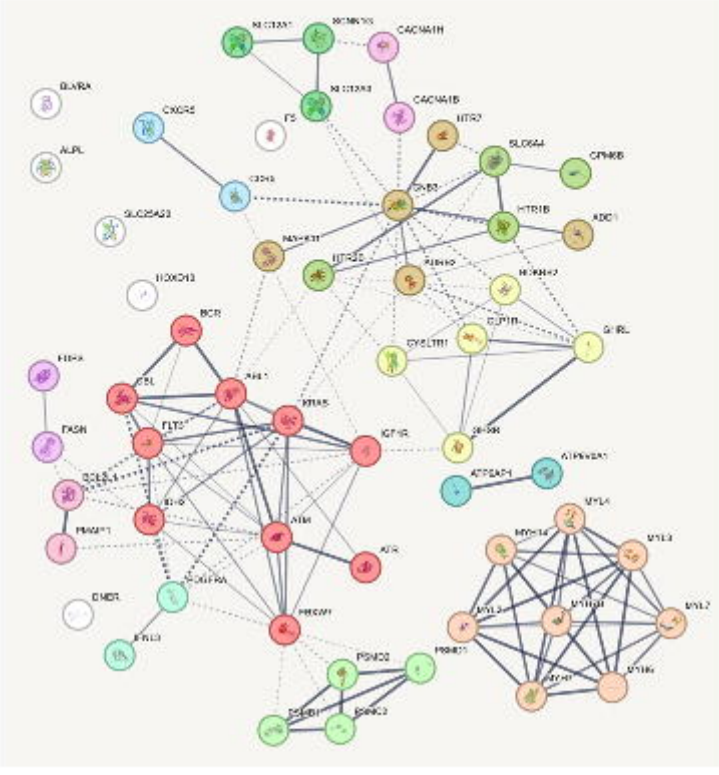
Cluster 5



Cluster 7



Danuglipron fragment network

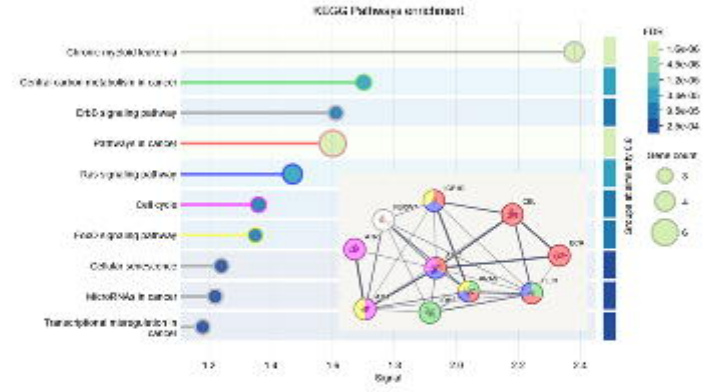


number of nodes: 57
number of edges: 139

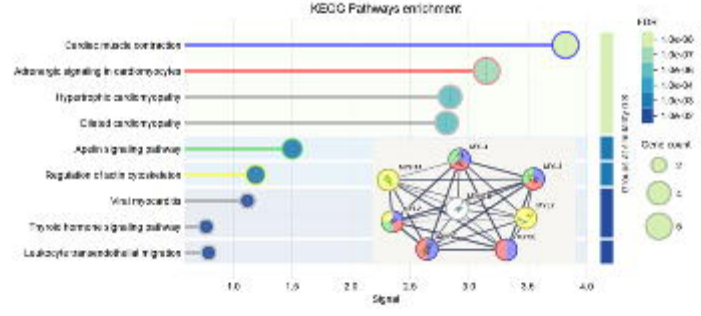
expected number of edges: 50
PPI enrichment p-value: < 1.0e-16

color	clusterId	gene count	description
	Cluster 1	10	1. Chronic myeloid leukemia 2. PI3 signaling in disease
	Cluster 2	8	Myosin
	Cluster 3	5	ADD1, ADRB2, GNB3, HTR7, MAPK11
	Cluster 4	5	1. Positive regulation of small intestinal transit 2. Type 1 melanocortin receptor binding, and Melanocortin/ACTH receptor
	Cluster 5	4	Serotonin binding
	Cluster 6	4	1. Proteasome 2. Regulation of activated PAK-2p34 by proteasome mediated degradation
	Cluster 7	3	1. Sodium ion homeostasis 2. Cation-coupled Chloride cotransporters 3. Sodium potassium chloride symporter activity
	Cluster 8	2	IFNL3, PDGFRA
	Cluster 9	2	1. Endosomal lumen acidification 2. Insulin receptor recycling 3. Proton-transporting V-type ATPase complex
	Cluster 10	2	C-C chemokine receptor activity
	Cluster 11	2	FASN, FDPS
	Cluster 12	2	Voltage-gated calcium channel complex
	Cluster 13	2	1. Apoptosis - multiple species 2. BH3-only proteins associate with and inactivate anti-apoptotic BCL-2 members 3. Bcl-2 family protein complex

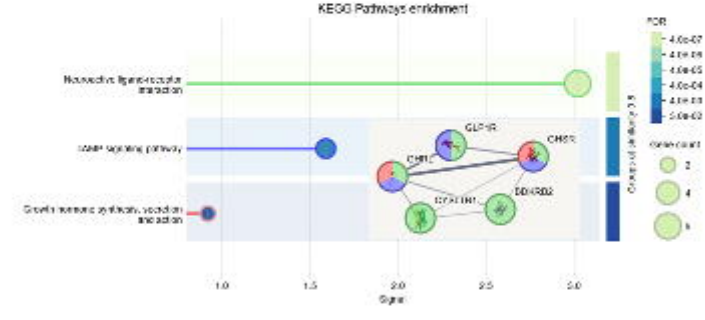
Cluster 1

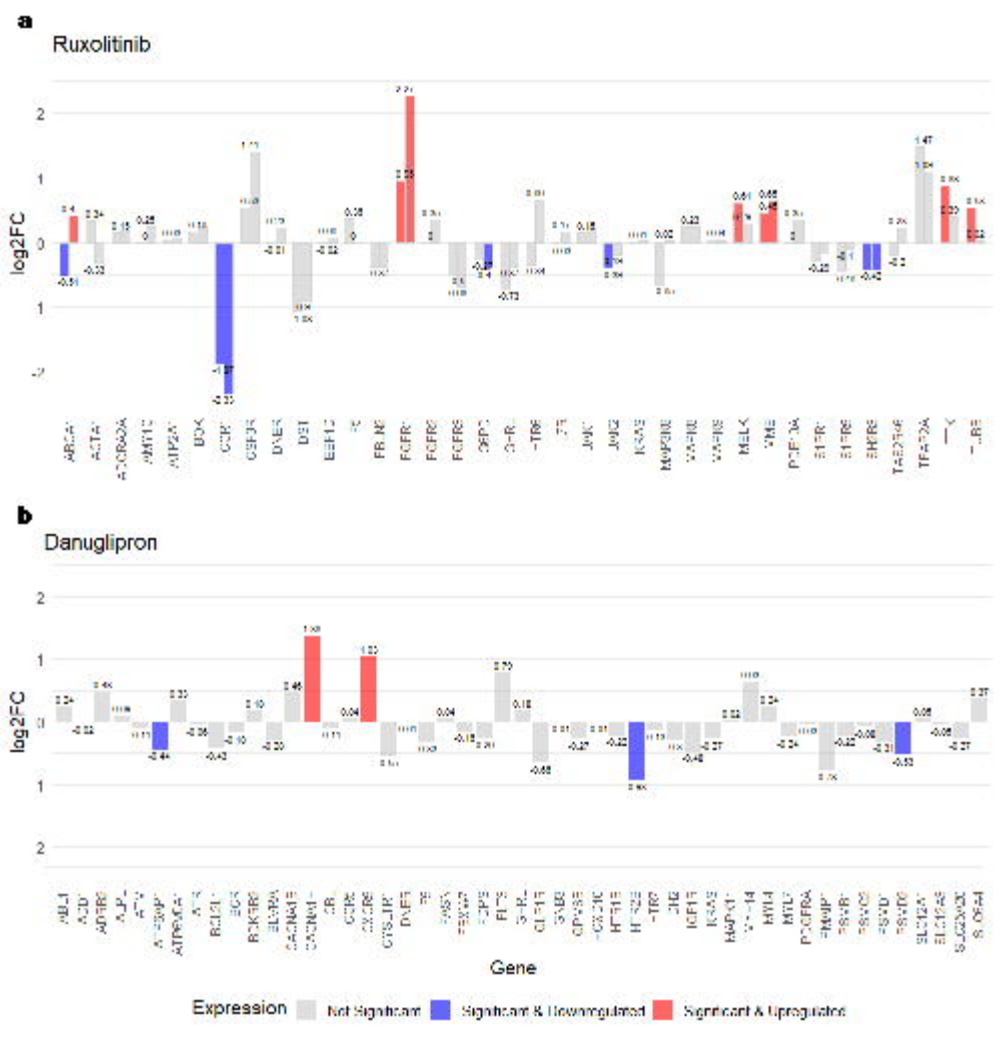


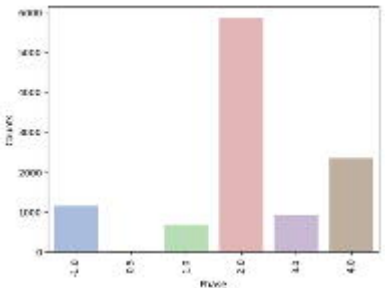
Cluster 2

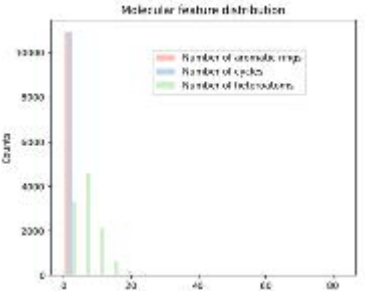


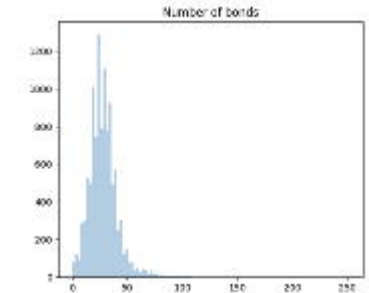
Cluster 7

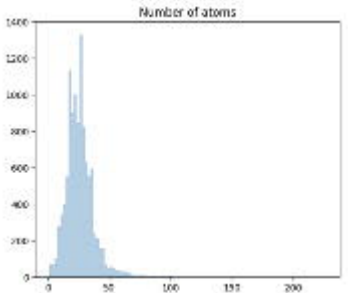


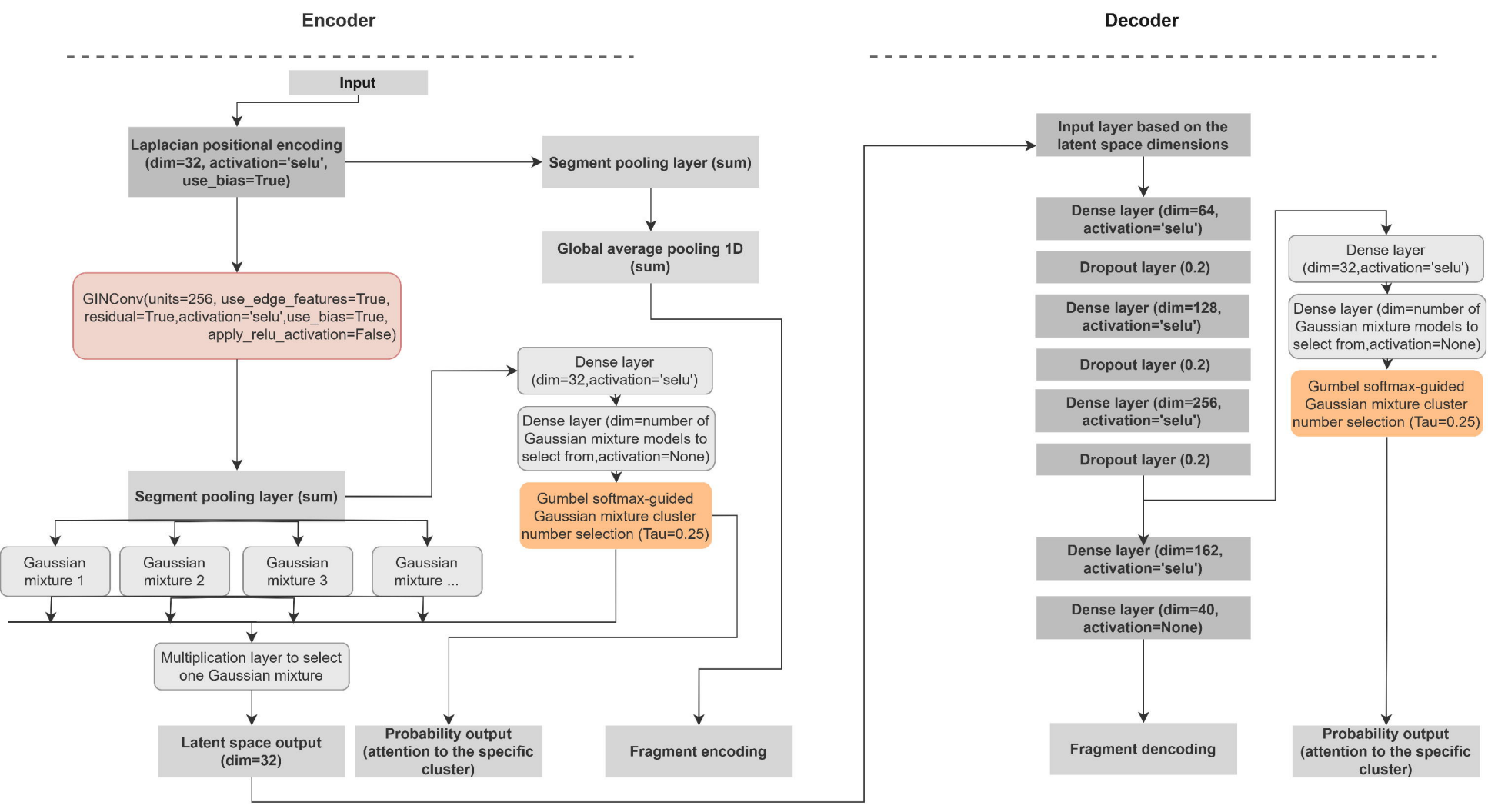

a
Ruxolitinib
log2FC
b
Danuglipron
log2FC
Gene
Expression
Not Significant
Significant & Downregulated
Significant & Upregulated

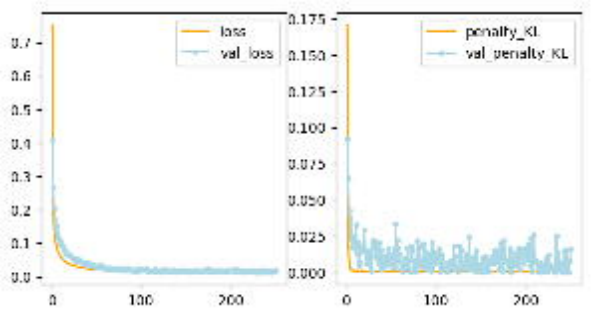

Counts
Phase

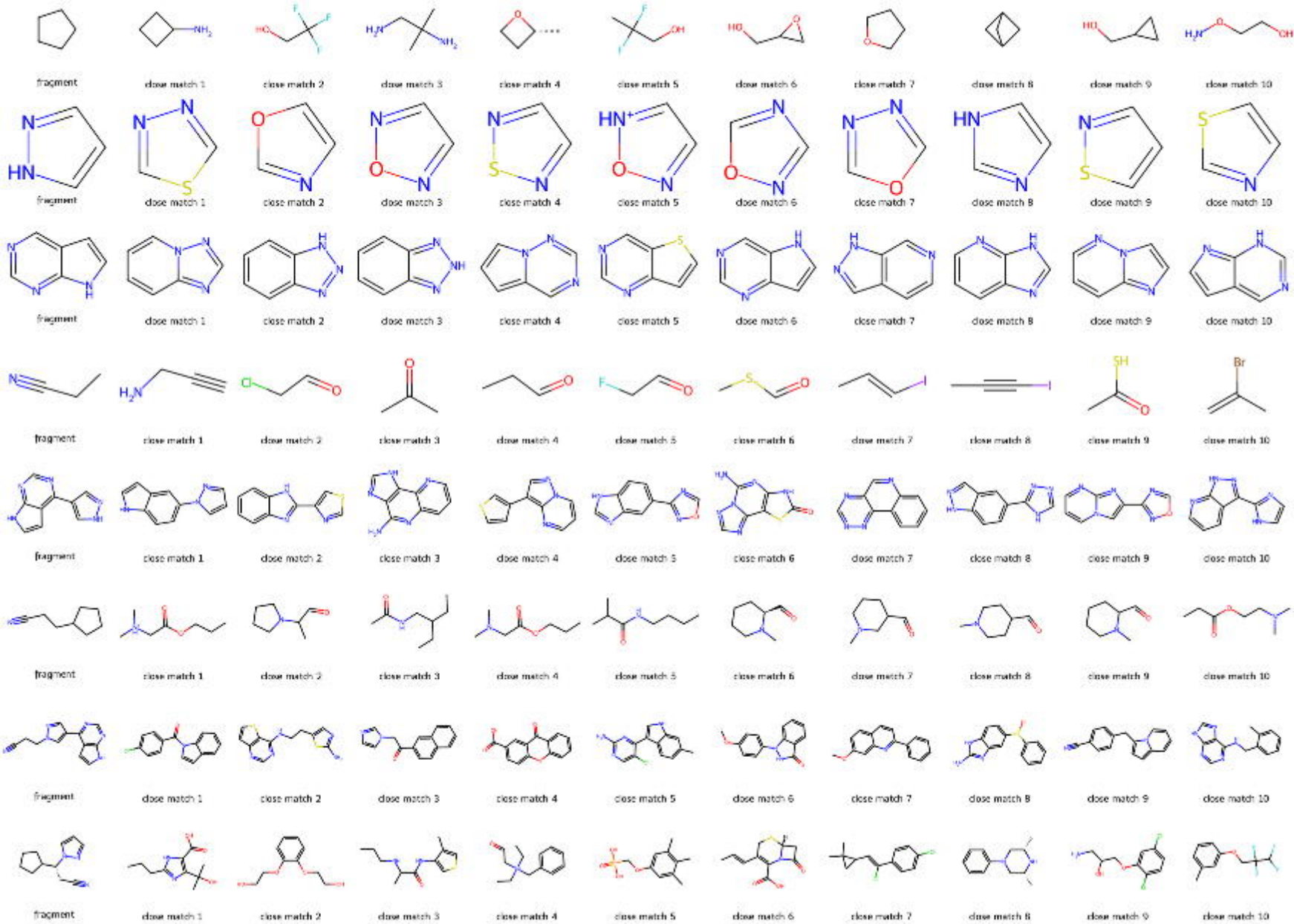

Molecular feature distribution
Number of aromatic rings
Number of cycles
Number of heteroatoms
Counts

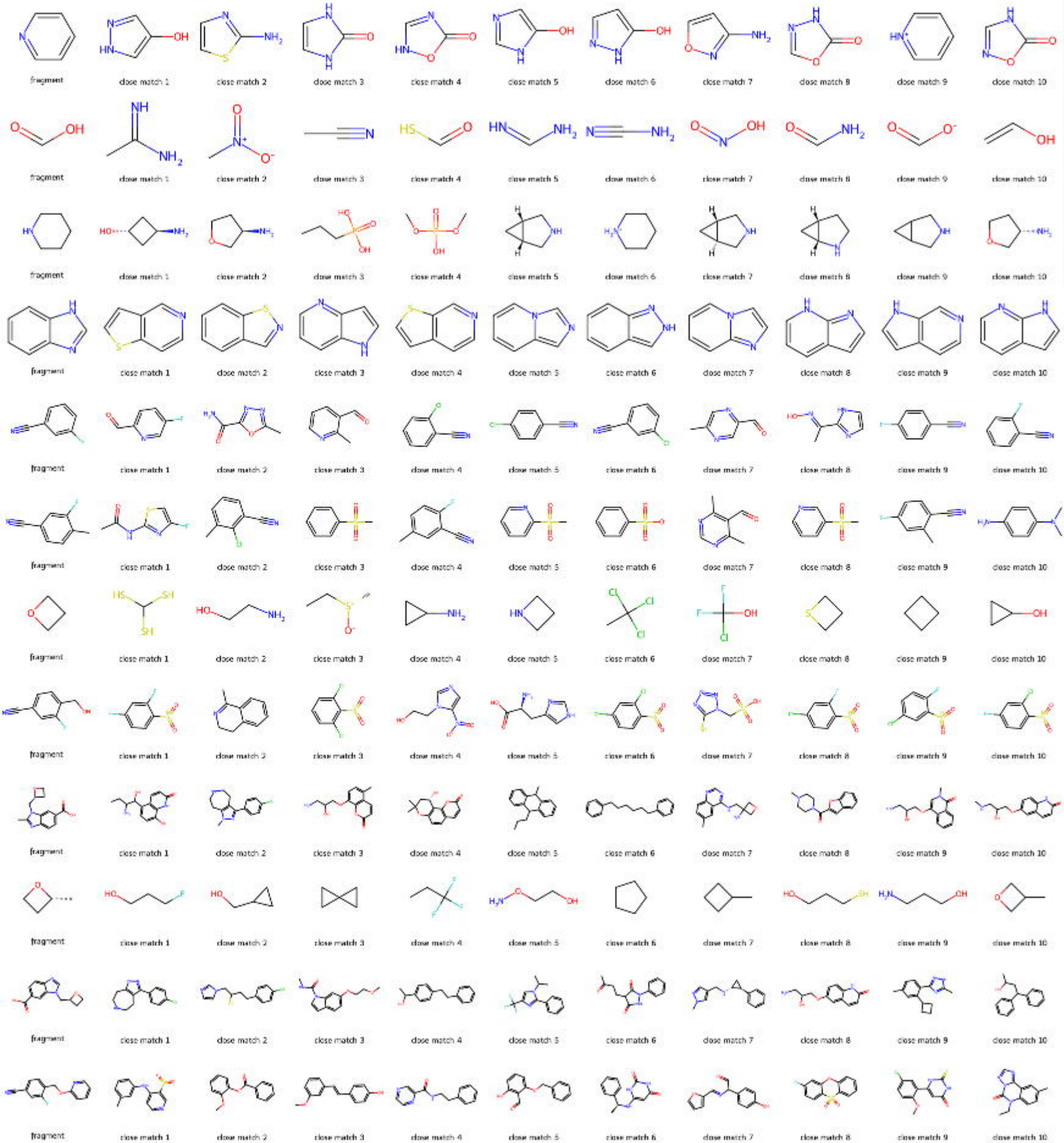

Number of bonds


Number of atoms


Encoder
Decoder
Input
Laplacian positional encoding (dim=32, activation='selu', use_bias=True)
Segment pooling layer (sum)
Global average pooling 1D (sum)
GINConv(units=256, use_edge_features=True, residual=True,activation='selu',use_bias=True, apply_relu_activation=False)
Dense layer (dim=32,activation='selu')
Dense layer (dim=number of Gaussian mixture models to select from,activation=None)
Gumbel softmax-guided Gaussian mixture cluster number selection (Tau=0.25)
Segment pooling layer (sum)
Gaussian mixture 1
Gaussian mixture 2
Gaussian mixture 3
Gaussian mixture ...
Multiplication layer to select one Gaussian mixture
Latent space output (dim=32)
Probability output (attention to the specific cluster)
Fragment encoding
Input layer based on the latent space dimensions
Dense layer (dim=64, activation='selu')
Dropout layer (0.2)
Dense layer (dim=128, activation='selu')
Dropout layer (0.2)
Dense layer (dim=256, activation='selu')
Dropout layer (0.2)
Dense layer (dim=162, activation='selu')
Dense layer (dim=40, activation=None)
Fragment dencoding
Dense layer (dim=32,activation='selu')
Dense layer (dim=number of Gaussian mixture models to select from,activation=None)
Gumbel softmax-guided Gaussian mixture cluster number selection (Tau=0.25)
Probability output (attention to the specific cluster)

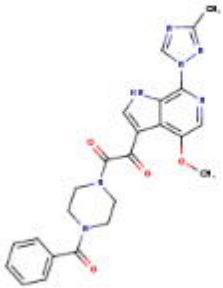

loss
val_loss
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
100
200
penalty_KL
val_penalty_KL
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
100
200


fragment
close match 1
close match 2
close match 3
close match 4
close match 5
close match 6
close match 7
close match 8
close match 9
close match 10

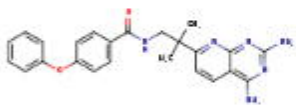

fragment
close match 1
close match 2
close match 3
close match 4
close match 5
close match 6
close match 7
close match 8
close match 9
close match 10

Similar compounds to the ruxolitinib candidate (3)

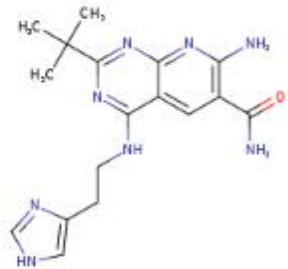
DB14675, Temsavir
Score : 0.383



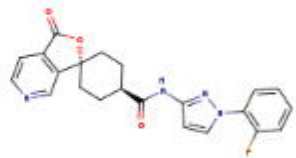
DB07041
Score : 0.316



DB07042
Score : 0.290

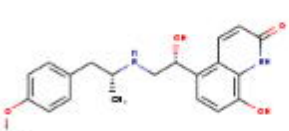


DB12168, MK-0557

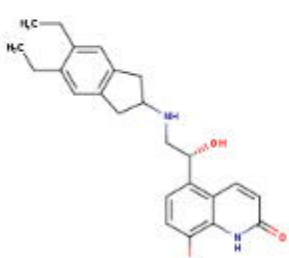


Similar compounds to the danuglipron candidate (2)

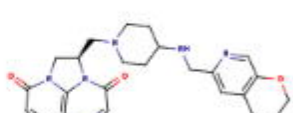
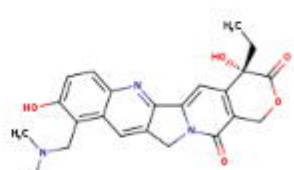
DB15784, Carmoterol
Score : 0.246



DB05039, Indacaterol
Score : 0.242



DB12134, Gepotidacin
Score : 0.200



DB01030, Topotecan
Score : 0.173