

Università degli Studi del Sannio

Dipartimento di Scienze e Tecnologie
Corso di Laurea Magistrale in Biologia

Curriculum Biosanitario

Tesi di laurea in Bioinformatica

Analisi bioinformatica delle strutture comuni in piccole molecole associate al carcinoma ovarico

Relatore
Ch.mo Prof.
Francesco Napolitano

Candidata
Erica Cogliano
Mat 506001932

Carcinoma ovarico - Incidenza

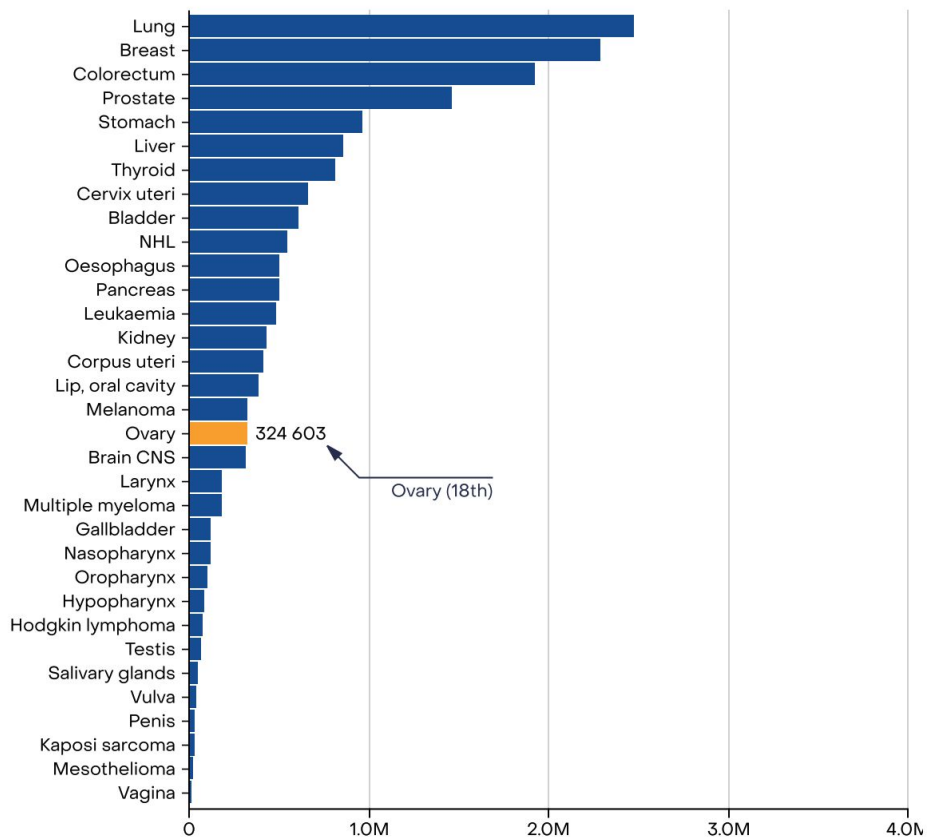
Rank 18

Casi 324 603

ASR (World) 6.7

Numero di nuovi casi, entrambi i sessi, tutte le età

International Agency for Research on Cancer - WHO (2022)



Carcinoma ovarico - Mortalità

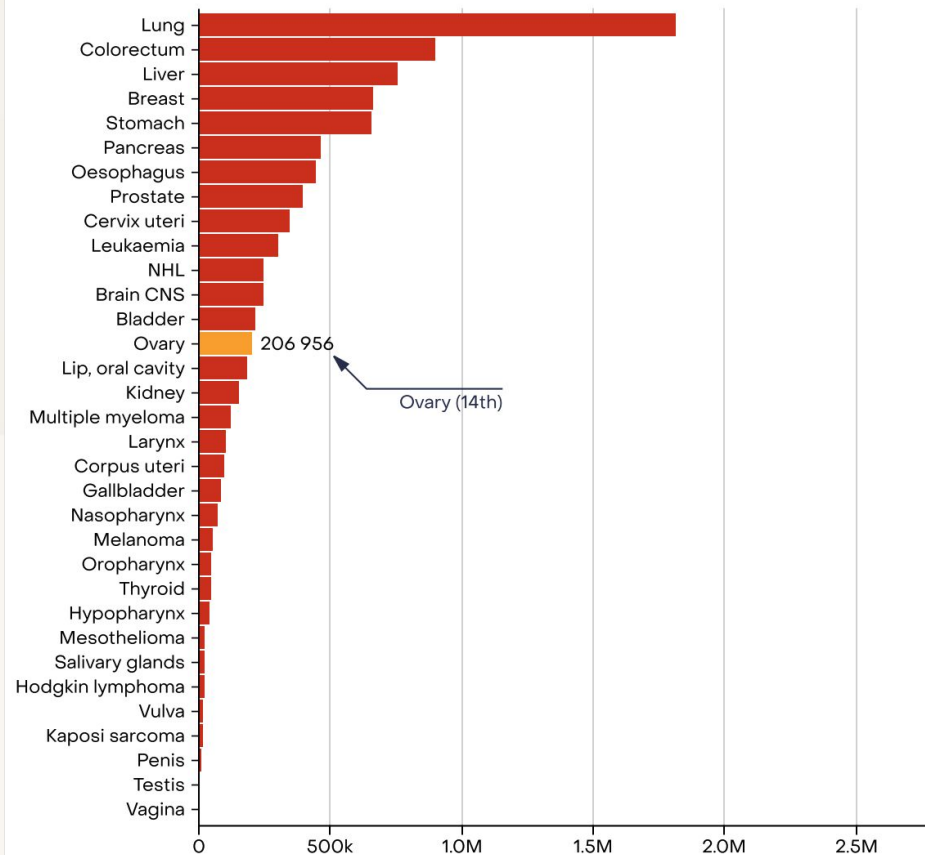
Rank 14

Casi 206 956

ASR (World) 4.0

Numero di morti, entrambi i sessi, tutte le età

International Agency for Research on Cancer - WHO (2022)



Origine ed eterogeneità della patologia

Classificazione dei carcinomi ovarici epiteliali

- Sieroso di alto grado
- Sieroso di basso grado
- Mucinoso
- Endometrioide
- Cellule chiare

TIPO I

- Lenti e meno aggressivi
- Aspetto microscopico ordinato
- Lesione precursore benigna

TIPO II

- Rapidi e molto aggressivi
- Crescita rapida e tendenza metastatica
- Origine nella fimbria tubarica

HGSC

Sottotipo di maggiore interesse biologico,
rappresenta il 95% dei tumori epiteliali

TP53:

forte associazione con alterazioni di TP53, gene oncosoppressore coinvolto nel controllo del ciclo cellulare

BRCA1/BRCA2:

coinvolgimento di geni della ricombinazione omologa rilevante nei casi familiari o geneticamente predisposti

Valore clinico:

profilo molecolare utile per orientare la scelta di trattamenti mirati

Difficoltà cliniche principali

In questo contesto e' utile affiancare alla lettura clinica una lettura strutturale dei composti associati alla malattia.

Diagnosi tardiva

I sintomi iniziali sono spesso vaghi e la malattia viene riconosciuta frequentemente in fase avanzata

Recidiva frequente

anche dopo una buona risposta iniziale, molte pazienti vanno incontro a recidiva

Resistenza al platino

recidiva entro 6 mesi o progressione durante la terapia rendono la gestione clinica più complessa

Obiettivo della tesi

≤ 100 atomi, focalizzazione sulle piccole molecole per lavorare su strutture più omogenee e confrontabili.

1. **Selezione** dei composti associati al cancro ovarico
- ↓
2. **Recuperare** le strutture molecolari in formati utilizzabili per le analisi chemioinformatiche
- ↓
3. **Confronto** tramite fingerprint e similarità strutturale
- ↓
4. **Raggruppamento** dei composti in cluster affini
- ↓
5. **Interpretazione** dei cluster tramite MCS e confronto con molecole esterne

Therapeutic Target Database (TTD)

Banca dati pubblica che raccoglie informazioni sui target terapeutici e sui composti associati alle diverse patologie.

1. input: P1-05-Disease.txt
2. parsing dei dati in formato leggibile
3. filtro dei record per indicazione ed ICD11
4. output: dataset ordinato e utilizzabile nelle fasi successive

TTD - Therapeutic Targets Database Full Data Download File
Title - Drug to disease mapping with ICD identifiers
Version 10.1.01 (2024.03.30)
Provided by IDRB Lab of Innovative Drug Research and Bioinformatics
College of Pharmaceutical Sciences
Zhejiang University
<https://idrblab.org/>
Provided by BIDD Bioinformatic and Drug Design Group
Department of Pharmacy
National University of Singapore
<https://bidd.group/>
Any question about data provided here, please contact with:
Dr. Zhou (zhou_ying@zju.edu.cn) and Dr. Zhang (zhangyintao@zju.edu.cn)

Abbreviations:

TTDDRUID TTD Drug ID

DRUGNAME Drug Name

INDICATI Indication Disease entry ICD-11 Clinical status

TTDDRUID DZB84T

DRUGNAME Maralixibat

INDICATI Pruritus ICD-11: EC90 Approved

INDICATI Progressive familial intrahepatic cholestasis ICD-11: 5C58.03

Phase 3

INDICATI Alagille syndrome ICD-11: LB20.0Y Phase 2

TTDDRUID DZA90G

DRUGNAME BNT162b2

INDICATI Coronavirus Disease 2019 (COVID-19) ICD-11: 1D6Y Approved

TTDDRUID DZ8DF0

DRUGNAME Nedosiran

INDICATI Primary hyperoxaluria type 1 ICD-11: 5C51.20 Approved

INDICATI Hyperoxaluria ICD-11: 5C51.2 Phase 2

TTDDRUID DY49OT

DRUGNAMEandexxa

INDICATI Reversal of anticoagulation ICD-11: N.A. Approved

TTDDRUID DY38JR

DRUGNAME Astodimer

INDICATI Bacterial vaginosis ICD-11: MF3A Approved

[...]

Recupero delle strutture: SDF, Pubchem

SDF:

file che contiene una o più strutture molecolari in formato MOLfile

Pubchem:

banca dati pubblica utilizzata per ottenere e verificare strutture, identificativi e sinonimi.

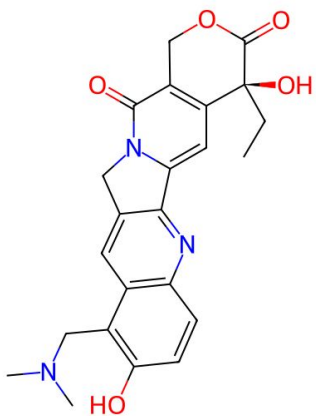
MOLfile:

oggetto manipolabile dal software rendendo possibile il trattamento computazionale per atomi, legami e anelli.

```
Carboplatin
RDKit 2D
13 10 0 0 0 0 0 0 0 9999 V2000
4.8059 3.8526 0.0000 Pt 0 0 0 0 0 15 0 0 0 0 0
5.6720 3.3526 0.0000 O 0 0 0 0 0 0 0 0 0 0 0
3.9399 3.3526 0.0000 O 0 0 0 0 0 0 0 0 0 0 0
6.5380 1.8526 0.0000 O 0 0 0 0 0 0 0 0 0 0 0
3.0739 1.8526 0.0000 O 0 0 0 0 0 0 0 0 0 0 0
0.5369 3.5463 0.0000 N 0 0 0 0 0 0 0 0 0 0 0
4.8059 6.4726 0.0000 N 0 0 0 0 0 0 0 0 0 0 0
4.8059 1.8526 0.0000 C 0 0 0 0 0 0 0 0 0 0 0
5.5130 1.1455 0.0000 C 0 0 0 0 0 0 0 0 0 0 0
4.0988 1.1455 0.0000 C 0 0 0 0 0 0 0 0 0 0 0
4.8059 0.4384 0.0000 C 0 0 0 0 0 0 0 0 0 0 0
5.6720 2.3526 0.0000 C 0 0 0 0 0 0 0 0 0 0 0
3.9399 2.3526 0.0000 C 0 0 0 0 0 0 0 0 0 0 0
2 12 1 0
3 13 1 0
4 12 2 0
5 13 2 0
8 9 1 0
8 10 1 0
8 12 1 0
8 13 1 0
9 11 1 0
10 11 1 0
M CHG 3 1 2 2 -1 3 -1
M END
```

Fingerprint molecolari

Rappresentazione numerica della struttura molecolare, usata per confrontare rapidamente molecole diverse.



Morgan Fingerprint



(0,0,1,0,1,0,0,...,0,1,0,0)

Similarità di Tanimoto

Il coefficiente di Tanimoto misura quanto due molecole condividono caratteristiche strutturali comuni a partire dai fingerprint.

$$S_{AB} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

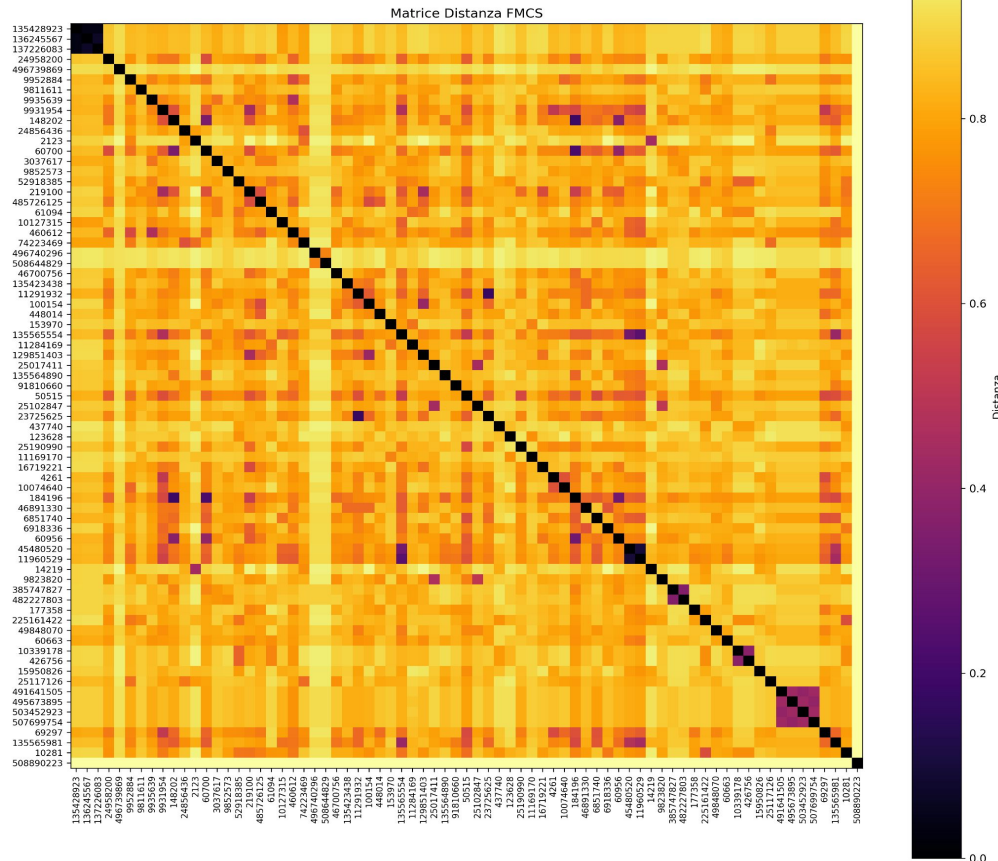
Matrice delle distanze

Esprime quanto sono dissimili le molecole.

$$D = 1 - \text{SIMILARITÀ}$$

Valori bassi, strutture affini.

Valori alti, strutture diverse.

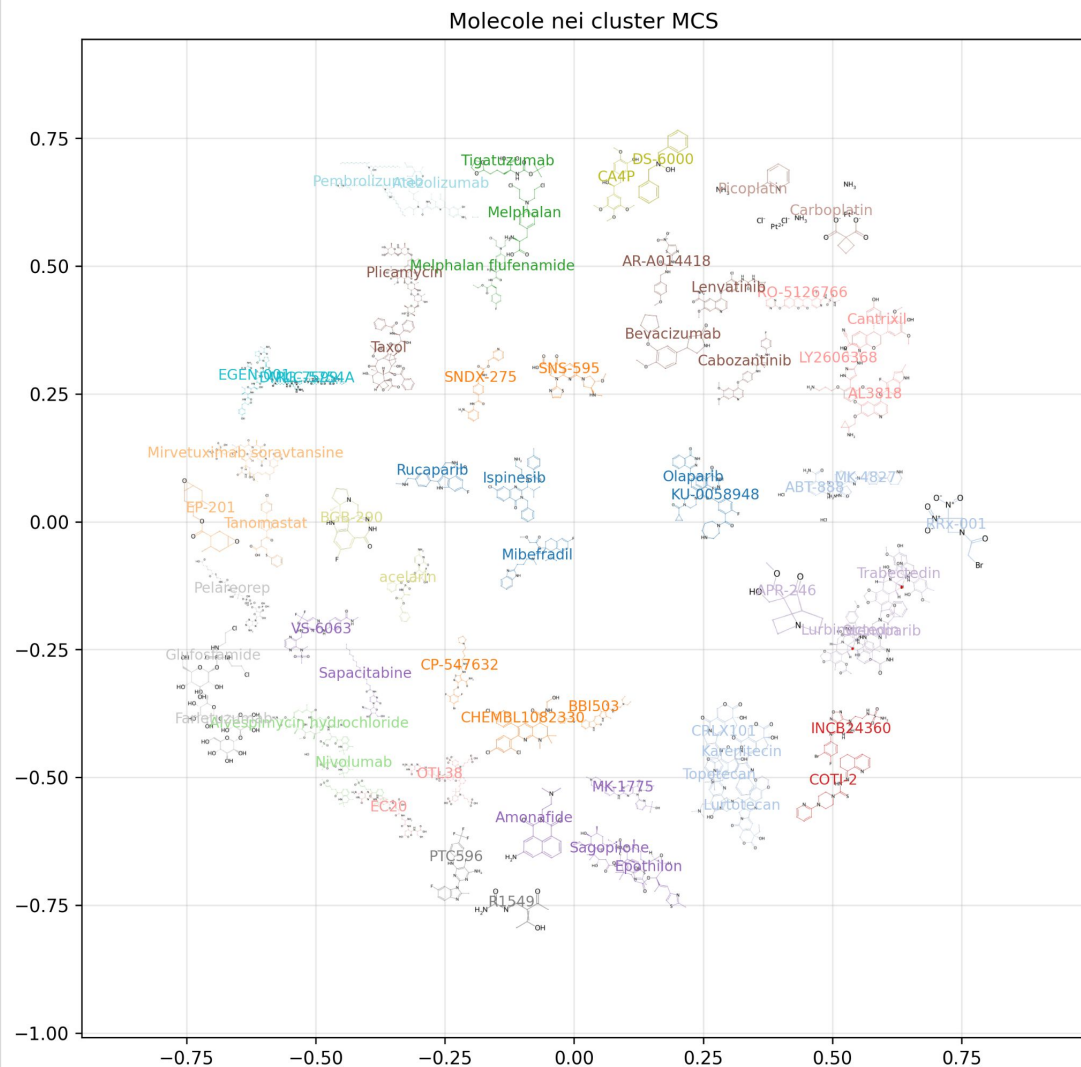


Visualizzazione e clustering

MDS: proietta le molecole in uno spazio a bassa dimensionalità e ne preserva le distanze

k-means: permette di raggruppare le molecole in sottogruppi a partire dalle proiezioni delle molecole nello spazio n-dimensionale.

Interpretazione: la visualizzazione e il clustering rendono più leggibile un insieme inizialmente disperso e aiutano ad identificare sottogruppi strutturali.

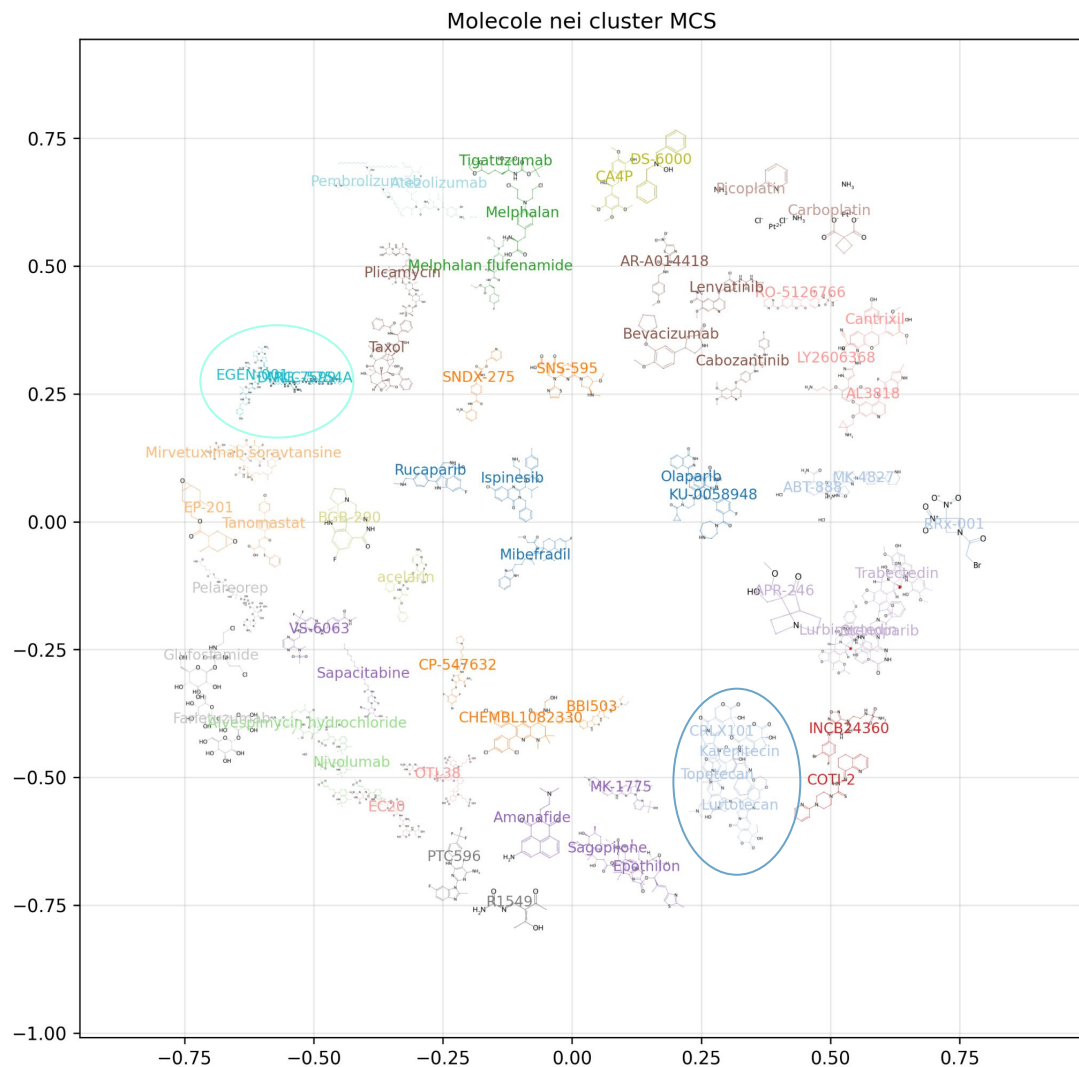


Visualizzazione e clustering

MDS: proietta le molecole in uno spazio a bassa dimensionalità e ne preserva le distanze

k-means: permette di raggruppare le molecole in sottogruppi a partire dalle proiezioni delle molecole nello spazio n-dimensionale.

Interpretazione: la visualizzazione e il clustering rendono più leggibile un insieme inizialmente disperso e aiutano ad identificare sottogruppi strutturali.

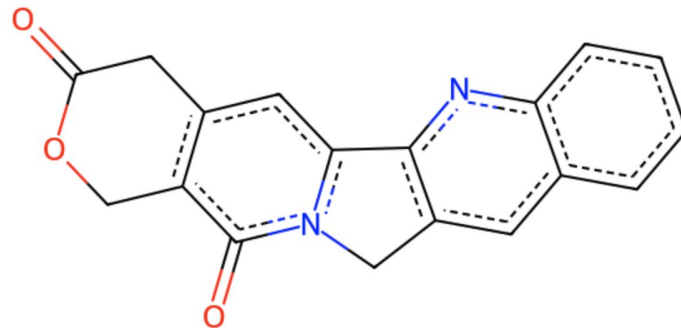
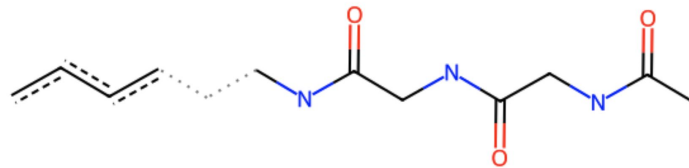


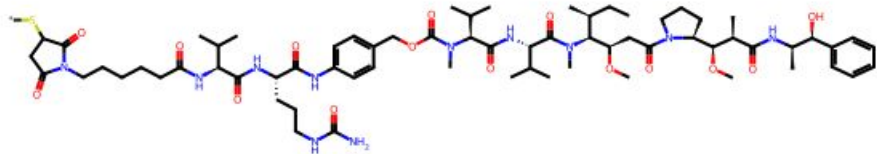
MCS e Scaffold Bemis Murko

MCS: identificano la massima sottostruttura comune per ogni clusters

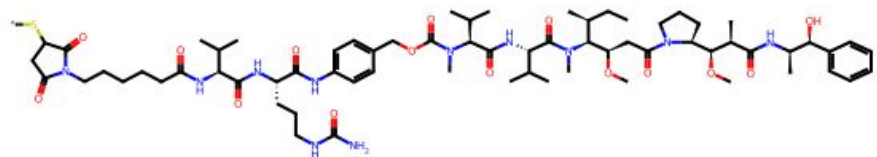
Bemis-Murko: scaffold della MCS.
Mantiene il nucleo centrale della molecola eliminando i sostituenti periferici.

Confronto: MCS vs P1-05-Drug_disease.txt residuo

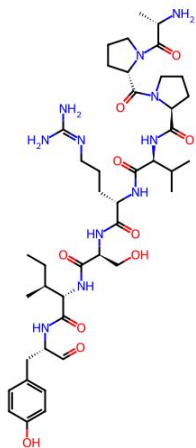




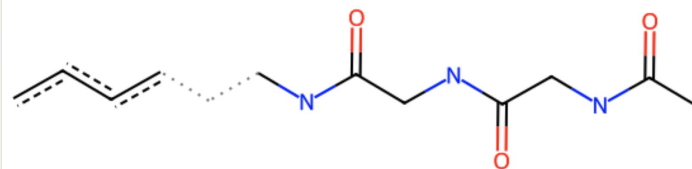
DMUC-5754A



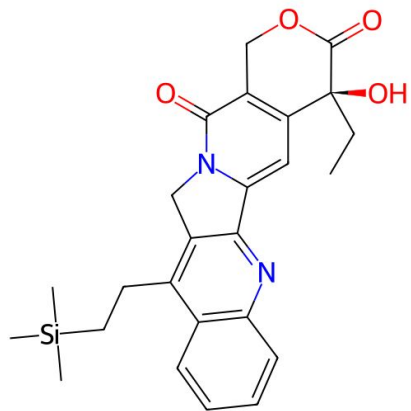
RG-7599



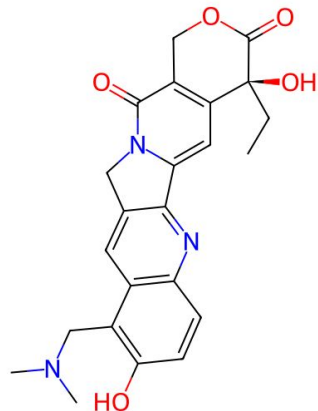
EGEN-001



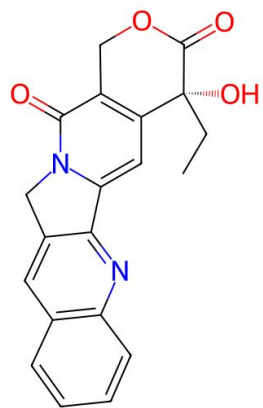
MCS Cluster 18 (size 3)
 - DMUC-5754A
 - RG-7599
 - EGEN-001



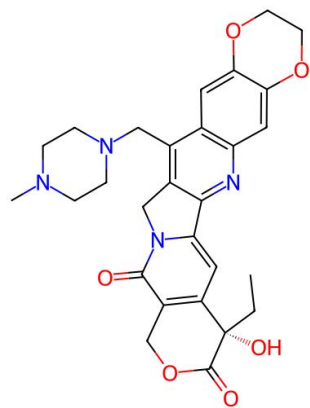
Karenitecin



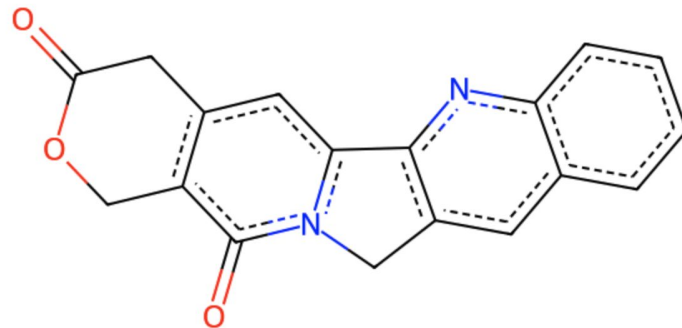
Topotecan



CRLX101



Lurtotecan



MCS Cluster 21 (size 4)
 - Karenitecin
 - Topotecan
 - CRLX101
 - Lurtotecan

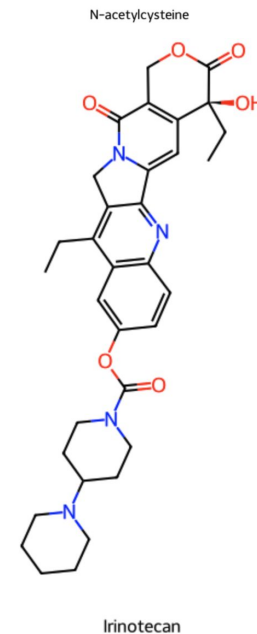
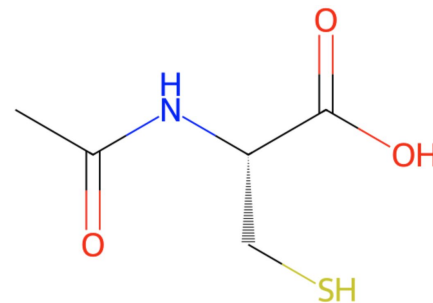
Risultati principali e confronto in letteratura

Primo output: un dataset strutturale ottenuto da PubChem, in formato SDF, derivato dal filtro farmacologico iniziale.

Secondo output: Scaffold/MCS dei cluster un set di core chimici interpretabili.

Fase successiva di confronto con le molecole residue ha permesso di evidenziare hit coerenti con evidenze già presenti in letteratura.

Altri riscontri: treosulfan, vorinostat, disulfiram, thalidomide.



Significato del lavoro e conclusioni

Metodo: descrive i composti associati alla patologia e li confronta. In questo modo, un insieme eterogeneo di molecole diventa uno spazio chimico più leggibile, nel quale si possono riconoscere gruppi affini e motivi comuni.

Valore: approccio può mettere in evidenza anche molecole esterne al dataset iniziale, ma strutturalmente affini ai composti associati alla patologia

Approccio: genera ipotesi esplorative per selezionare composti utili da approfondire, non prove definitive di meccanismo o efficacia clinica.

The background is a light beige color. It features several large, overlapping white circles. On the right side, there are three clusters of hexagonal patterns. Each cluster consists of a central hexagon with lines extending from its vertices to form a larger, more complex geometric structure. The text "Grazie per l'attenzione!" is centered in the upper half of the image.

Grazie per l'attenzione!

bioinformatics Lab @unisannio