

Applicazione del TSR per il drug repurposing

Giovanni Firrincieli

Dipartimento di Matematica e Informatica
Corso di Laurea Magistrale in Informatica [LM-18]

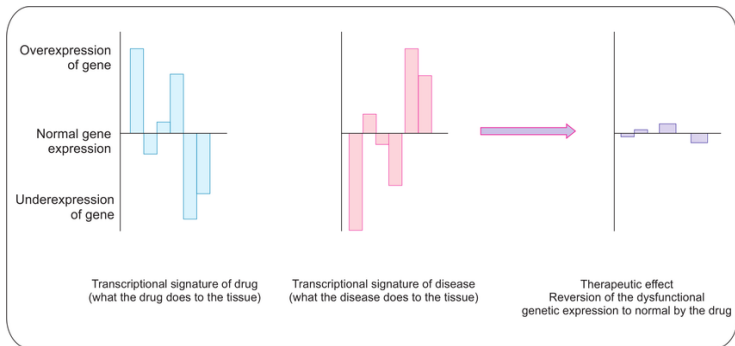
Relatore:
Prof. Alfredo Pulvirenti
Correlatori:
Prof. Giovanni Micale
Prof. Salvatore Alaimo

- Concetto di trascrittoma, firma trascrittomica e firma trascrittomica inversa.
- Workflow, dataset utilizzati, materiali e metodo.
- Pipeline R.
- Interfaccia grafica.
- Risultati.
- Conclusioni.
- Sviluppi futuri.

- Il Trascrittoma è l'insieme completo delle trascrizioni genetiche presenti in una cellula in un dato momento.
- Firma trascrittomica: profilo dell'espressione genica in un contesto specifico.
- TSR: Firma trascrittomica inversa (Transcription Signature Reversion), pattern anticomplementari per regolarizzare l'espressione genica.

- Il Trascrittoma è l'insieme completo delle trascrizioni genetiche presenti in una cellula in un dato momento.
- Firma trascrittomica: profilo dell'espressione genica in un contesto specifico.
- TSR: Firma trascrittomica inversa (Transcription Signature Reversion), pattern anticomplementari per regolarizzare l'espressione genica.

- Il Trascrittoma è l'insieme completo delle trascrizioni genetiche presenti in una cellula in un dato momento.
- Firma trascrittomica: profilo dell'espressione genica in un contesto specifico.
- TSR: Firma trascrittomica inversa (Transcription Signature Reversion), pattern anticomplementari per regolarizzare l'espressione genica.



Fonte: Scientific Figure on ResearchGate

- Una pipeline R per una lista di **farmaci riposizionabili** su linee cellulari.
- Un interfaccia per la consultazione dei risultati e l'avvio di nuove analisi.

Risultati sperimentali

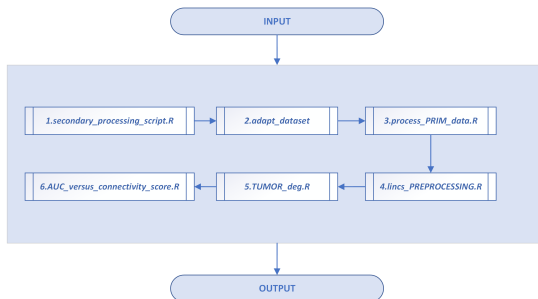
Analisi condotte sulle seguenti linee tumorali:

- A549_LUNG
- HCT116_LARGE_INTESTINE

A549 è una linea cellulare di adenocarcinoma polmonare umano.

HCT116 è una linea cellulare di carcinoma coloretale umano

- Diversi script in R per calcolo del punteggio.
- Ogni farmaco testato avrà un punteggio che misura il grado di inibizione verso una linea cellulare.



- **The PRISM repurposing dataset:** dataset farmaco-genomico
 - 578 linee tumorali testate con 4518 farmaci ad una singola concentrazione di $2,5 \mu M$
 - 1448 dei farmaci più attivi, nuovamente testati ad 8 concentrazioni che vanno da 610 pM a $10 \mu M$.
 - Utilizzo del parametro AUC (Area Under the Curve) per il progetto.
 - Dimensione dataset: **252MB**

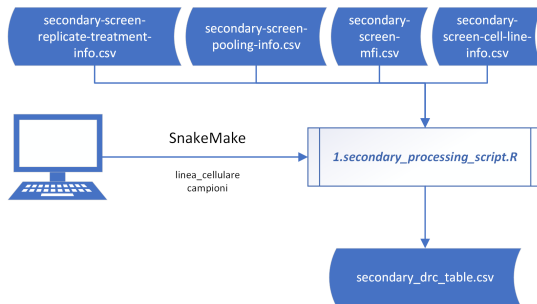
- La **Library of Integrated Network-Based Cellular Signatures (LINCS)** è un progetto di ricerca nel campo della biologia dei sistemi e della genomica funzionale.
- Contiene firme cellulari che includono dati molecolari e genetici le quali rappresentano le risposte delle cellule a stimoli variabili, come trattamenti farmacologici o modifiche ambientali.
- Dimensione dataset: **60GB**

- **The Cancer Genome Atlas (TCGA)** è il progetto di ricerca per la mappatura e analisi delle alterazioni genomiche associate a diversi tipi di cancro.
- Contiene analisi di dati genomici, epigenomici, proteomici e clinici da campioni di tessuto tumorale provenienti da pazienti affetti da vari tipi di cancro.
- Dimensione dataset: **8GB**

- MacBook Pro M1 con 16GB di Ram
- R version 4.3.1
- Python 3.1.2 e PyQt5
- **Snakemake**: uno strumento per creare analisi di dati riproducibili e scalabili.
 - Flussi di lavoro descritti in linguaggio Python-based, contenenti la descrizione del software necessario.
 - Flussi facilmente adattabili ad ambienti server, cluster e cloud senza la necessità di modificare la definizione del flusso di lavoro.

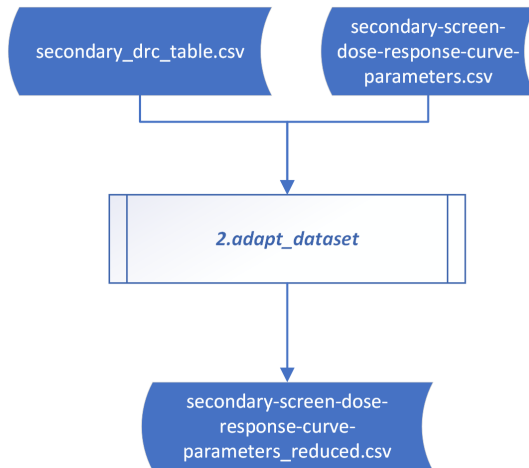
1. secondary_processing_script.R

- Linea cellulare e numero di campioni in input.
- Creazione dati sulla risposta di linee cellulari a trattamenti farmacologici sperimentali.
- Dati salvati su un file csv.
- Tempo di esecuzione: **1h/10000 campioni**



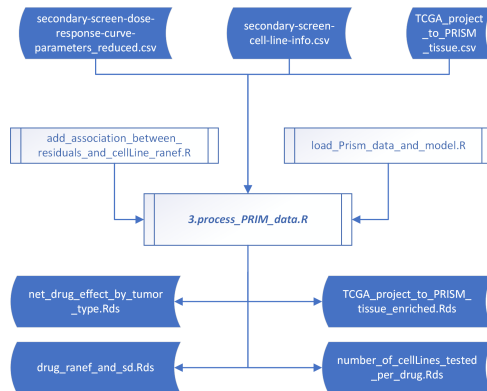
2. adapt_dataset.R

- Aggiunta di informazioni supplementari al dataset generato in precedenza.
 - Nome linea cellulare
 - Nome farmaco
 - MOA
 - Target e area interessata
- Dati salvati su un file csv.
- Tempo di esecuzione: **2min**

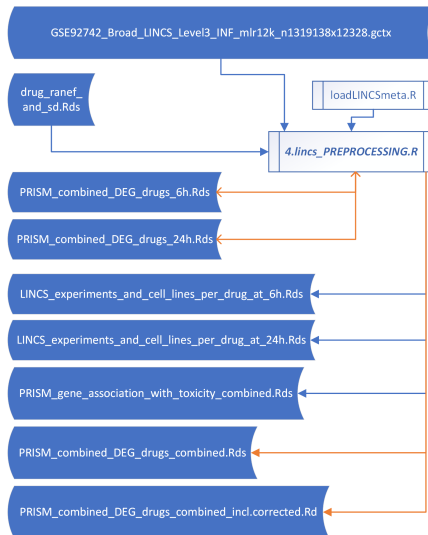


MOA = Mechanism of Action

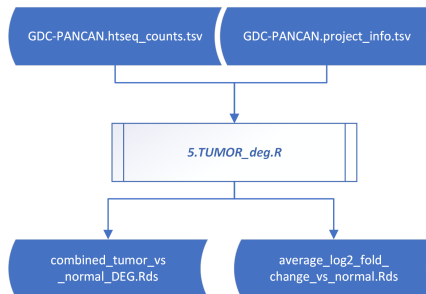
- Applicazione del modello lineare a effetti misti (**lmer**) per valutare l'effetto del farmaco sulla linea cellulare.
- Associazione linea cellulare a tipologia di cancro.
- Dati salvati su file RDS contenente dati sperimentali sugli effetti dei farmaci per la linea cellulare.
- Tempo di esecuzione: **10min**



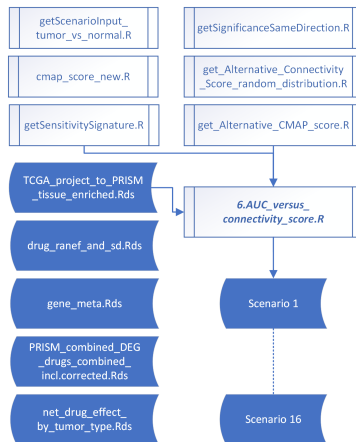
- Generazione dataset correlati con la funzione **loadLINCSmeta.R**
- Utilizza geni essenziali ("**landmark genes**") per la ricostruzione della rete perturbata.
- Testa ogni farmaco a $10\mu M$ per 6 e 24 ore.
- Modellazione lineare a effetti misti per ogni gene, farmaco e linea cellulare.
- Dati salvati su file RDS.
- Tempo di esecuzione: **48h**



- Analisi Differenziale tra Tessuto Tumorale e Normale.
- Utilizza dati dal progetto TCGA per confronto di espressione genica.
- Dati salvati su file RDS.
- Tempo di esecuzione:
30min



- Calcolo degli score di connettività tra le firme molecolari dei farmaci e la linea cellulare.
- Creazione e salvataggio di 16 scenari diversi (in formato RDS)
- Metodi di selezione
 - 50 MSSG
 - 100 MSSG
 - 150 MSSG
 - Bin Chen method
- Experiment Duration
 - 6h/24h
 - 6h_24h
 - 6h_24h_corrected
- Tempo di esecuzione: **20min**




MSSG = Most Significant Statistically Genes

- Sviluppata in Python e PyQt5
- rpy2 per il porting degli script R in Python
- csv per il caricamento e la gestione del file contenente le informazioni su linea cellulare e tessuto associato.
- Due modalità operative:
 - Repurposing
 - Preprocessing

TSR for drug repurposing

Drugs Repurposed Tool New Line Preprocessing Tool

 **Drug Repurposing from existing Experiment**

Choose a Tissue:

LUNG

Choose a Cellular Line...

A549

Choose a method

☐ 50 MSSG ☐ 100 MSSG ☒ 150 MSSG ☐ Bin Chen Method

Experiment duration:

☐ 6h ☐ 24h ☐ 6h_24h ☒ 6h_24_corrected

Extract data

Selected Tissue: LUNG
Selected CCLE: A549
Selected Method: 150 MSSG
Experiment duration: 6h_24h_c

- Punteggio negativo = Buona capacità inibitoria del farmaco sulla linea cellulare.
- P-value inferiore a 0,05 conferisce significatività al punteggio.


Tissue: LUNG || CCL6:A549 || Method: 150 MSSG || Timelength: 34h_c

drug_name	connectivity_score_with_LINCS_profile	connectivity_score_with_LINCS_profile.pvalue	alt_connectivity_score_with_LINCS_profile	alt_connectivity_score_with_LINCS_profile.pvalue
panobinostat	-0.44954500000000003	0.0010774808611196124	-0.2568722783260253	0.01883
MK-2206	-0.43916	3.360092197652134e-05	-0.43449459484914754	1e-05
cediranib	-0.4262	0.00015612204643292003	-0.382311108134411923	6e-05
BIBU-1361	-0.42547	7.402222530795056e-05	-0.2704563768022108	0.01145
tacedinaline	-0.41444	0.0024590599138579808	-0.3248126772502276	0.00106
bosutinib	-0.41215	0.000722263546437805	-0.25384325585776004	0.02095
doxorubicin	-0.39473	0.0086652980744662	-0.2395912910618793	0.0334
AZ-628	-0.39180499999999996	0.0014333744933439302	-0.33044020946196503	0.00086
saracatinib	-0.385385	0.0018700145482183537	-0.23918577393597532	0.03385
NSC-3852	-0.37919	0.003933007221782407	-0.27993301261027454	0.00783
belinostat	-0.37443499999999996	0.005138618975802369	-0.298250649009635083	0.00377
amsonafide	-0.36994499999999997	0.0015831635155580638	-0.3306377345458732	0.00083
votinostat	-0.36778	0.03245023296882102	-0.27420675617894146	0.00985
aminopurvalanol-a	-0.36717999999999995	0.0012764812642488763	-0.07232126200372546	0.27961
homohamiltonine	-0.36619	0.021850507798323063	-0.24673754931611406	0.02044
NSC-632839	-0.363555	0.017950613648722367	-0.2760286670039871	0.00921
sunitinib	-0.36250000000000004	0.0031786494911611215	-0.25841382630262566	0.01775
everolimus	-0.36236999999999997	0.01281171809939499	-0.3837967216321149	6e-05
BI-2536	-0.361805	0.041052753106628835	-0.3547973984114194	0.00022
gefiratinib	-0.358955	0.00662591528779542	-0.3608817548688218	0.00019
trichostatin-a	-0.35616	0.2432059940907918	-0.1871670635208305	0.12842
PCI-24781	-0.355605	0.004861240977950540	-0.26097843291438616	0.01625
PHA-865752	-0.35347	0.0031107011054757193	-0.284866710863679	0.0064
WZ-3146	-0.35047	0.018106155661753727	-0.2685071632638112	0.01319

Save as CSV

TSR for drug repurposing

Drugs Repurposed Tool New Line Preprocessing Tool

 **Preprocessing Tissue with specific CCLE**

Choose a Tissue:

LARGE_INTESTINE

Choose a Cellular Line:

LOVO

Insert samples (preprocessing):

10000

Preprocess

Selected Tissue: LARGE_INTESTINE
Selected CCLE: LOVO
Samples: 10000

A549 è una linea cellulare di adenocarcinoma polmonare umano

Risultati per un estrazione con 150 MSSG e 6h_24h_corrected:

drug_name	c_score_with_L_profile	c_s_w_L_profile.pvalue
panobinostat	-0.44954500000000003	0.0010767692798337577
MK-2206	-0.43916	5.510647464460808e-05
cediranib	-0.4262	0.00027153033996577723
BIBU-1361	-0.42547	7.428268915082463e-05
tacedinaline	-0.41444	0.0020856028175623596
belinostat	-0.37443499999999996	0.005322774676560825
vorinostat	-0.36778	0.023057130891617744
sunitinib	-0.36250000000000004	0.002819439904686258
everolimus	-0.36236999999999997	0.013643552691102458

Tabella: Estratto risultati A549_LUNG

Panobinostat: Applying HDACis to increase SSTR2 expression and radiolabeled DOTA-TATE uptake: from cells to mice
(2023 Dec 1:334:122173. doi: 10.1016/j.lfs.2023.122173. Epub 2023 Oct 29)

Cedinarib: Cediranib Induces Apoptosis, G1 Phase Cell Cycle Arrest, and Autophagy in Non-Small-Cell Lung Cancer Cell A549 In Vitro
(Volume 2021 | Article ID 5582648 | <https://doi.org/10.1155/2021/5582648>)

Everolimus: Everolimus inhibits the proliferation and migration of epidermal growth factor receptor-resistant lung cancer cells A549 via regulating the microRNA-4328/phosphatase and tensin homolog signaling pathway
(2019 Nov;18(5):5269-5276. doi: 10.3892/ol.2019.10887. Epub 2019 Sep 19)

HCT116 è una linea cellulare di carcinoma coloretale umano

Risultati per un estrazione con 150 MSSG e 6h_24h_corrected:

drug_name	c_score_with_L_profile	c_s_w_L_profile.pvalue
trichostatin-a	-0.557495	0.0004108450029859808
panobinostat	-0.549925	1.5824620614642984e-06
vorinostat	-0.541775	8.124240498608127e-05
PD-184352	-0.5108550000000001	5.931058434928109e-05
PCI-24781	-0.48372499999999996	2.005829037379543e-05
thiostrepton	-0.44981	3.884672166074261e-05
dasatinib	-0.44594500000000004	6.625666218610047e-05
tacedinaline	-0.444555	0.0001929959417132669
sunitinib	-0.438925	0.0007851167331943106

Tabella: Estratto risultati HCT116_LARGE_INTESTINE

Trichostatin-a: Trichostatin A induces p53-dependent endoplasmic reticulum stress in human colon cancer cells

(2019 Jan;17(1):660-667. doi: 10.3892/ol.2018.9641. Epub 2018 Oct 30)

Panobinostat Zap-Pano: a Photocaged Prodrug of the KDAC Inhibitor Panobinostat

(14 July 2021 <https://doi.org/10.1002/cmdc.202100403>)

Tacedinaline: Tacedinaline (CI-994) has undergone Phase II clinical trials for multiple myeloma, among other cancers.

(Medicinal Chemistry of Anticancer Drugs (Third Edition), 2023)

- Esiste una correlazione tra i risultati ottenuti dalla pipeline e gli articoli presenti in letteratura.
- Aumentando il numero di campioni nel primo step della pipeline, i risultati migliorano (con tempi di esecuzione maggiore).
- L'introduzione di Snakemake **automatizza e velocizza** l'esecuzione della Pipeline R nell'analisi di nuove linee cellulari.
- Il codice prodotto è disponibile su GitHub.

- Rilascio app tramite container Docker.
- Ottimizzazione del codice.
- Sviluppare Pipeline su *"una linea cellulare custom"*.

Grazie!