1. **DrugBank**: [https://go.drugbank.com/releases/latest](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgo.drugbank.com%2Freleases%2Flatest&data=04%7C01%7CTerry.Lin%40sas.com%7C3b9c4ec5b5e24bad3e4008d8a5c6e60b%7Cb1c14d5c362545b3a4309552373a0c2f%7C0%7C0%7C637441621381789722%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=mHtmlN8X%2BKBVZALmVWeSeJGuNt8Zw90Y9eFHf7lLI0s%3D&reserved=0)

**Utility**: the **drug and target relationship** could be curated from this database. Meanwhile, some general drug information such as therapeutic categories, drug indication, the chemical structure was provided.

**Database structure**: the download file was an XML file.

Therefore, a parser code (Jupyter file: drug\_target\_relationship\_DrugBank.ipynb) was needed to extract the information we needed. I attached our python script for this purpose for your information.

1. **COVID-19 related pathogen and host protein information**

Data resources: this data was widely distributed in the literature and publicly available resources. I listed a few examples:

1. Gordon, D.E. et al. (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583(7816):459-468.

26 SARS-CoV-2 virus proteins and 332 host proteins relationship were provided. I attached the relationship for your information (HCoV-associated host proteins.csv)

1. [https://www.ncbi.nlm.nih.gov/sars-cov-2/](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.ncbi.nlm.nih.gov%2Fsars-cov-2%2F&data=04%7C01%7CTerry.Lin%40sas.com%7C3b9c4ec5b5e24bad3e4008d8a5c6e60b%7Cb1c14d5c362545b3a4309552373a0c2f%7C0%7C0%7C637441621381789722%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=HcEMGp1F%2BUe7NGOzZIrVbKsGhVsfp5Kj2%2Blz%2F%2BOPp8w%3D&reserved=0)

This is COVID-19 data atlas where covers different biological profiles related to COVID19

1. **Protein-protein interaction data – STRING database**

[https://string-db.org/cgi/download?sessionId=blXUTlQ4pHMK](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fstring-db.org%2Fcgi%2Fdownload%3FsessionId%3DblXUTlQ4pHMK&data=04%7C01%7CTerry.Lin%40sas.com%7C3b9c4ec5b5e24bad3e4008d8a5c6e60b%7Cb1c14d5c362545b3a4309552373a0c2f%7C0%7C0%7C637441621381799719%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=3pjixgG1A0p1AnFlAKVZqIZKuqcCQnYp6a4MCv0IfXY%3D&reserved=0)

**I will show a demo how to handle this dataset.**

1. **Ground truths: COVID19 study in clinicaltrial.gov**

**Utility: as label to train deep learning model**

[https://clinicaltrials.gov/ct2/results?cond=COVID-19](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fclinicaltrials.gov%2Fct2%2Fresults%3Fcond%3DCOVID-19&data=04%7C01%7CTerry.Lin%40sas.com%7C3b9c4ec5b5e24bad3e4008d8a5c6e60b%7Cb1c14d5c362545b3a4309552373a0c2f%7C0%7C0%7C637441621381799719%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=rtWdD7cSHItel62U7fdo56gR8coDca9dGVGdRgI4gNs%3D&reserved=0)

I attached the data table (clinicaltrial\_covid.csv) and will explain in the meeting

Talk to you more at the meeting.

Best,

Zhichao

You now have the:

Drug - Target relationship [Drug Bank – all drugs XML download]

Host – Virus relationship [Gordon, Nature publication]

* Provided HCoV-associated host proteins.csv, but not sure if this is the right file.

Human – Human PPI relationship [Use STRING Homosapiens]

* Score = 0 to 1000, the higher the score the better - used as the weight.

Ground Truth that have the potential to treat Covid-19 [ClinicalTrials.gov]

* Use ground Truth to train the data – 1 drug formally approved so far with the rest in CT. All good potential candidates need to calculate enrichment rate. Intervention column is key – what kind of drug is in the trial – very important to scrub. Ground truth to challenge our model.

Goal: Develop an AI model to enrich the repositioning candidates

Ground Truth to train the data – 1 drug formally approved – rest in CT. All good candidates, all good potential – to calculate enrichment rate. Intervention column – what kind of drug is in the trial – very important. Ground truth to challenge our model.

Drug Bank download page XML – free account

* Chemical structure SDF format – need it for molecular docking – chemical structure side, do we want to step in to it? Focus on a drug at a time. Abacavir? Example.
* Target vs Enzyemes? Drug can perturb that therapeutic targets
* Try to find target effect

Key low hanging fruit Questions we want to address?

String – use Home sapiens – you can download PPI data – the first one is good enough. 71.2 MB.

Scored links – use score as weights. Human proteins only.

Ground Truth – to train algorithm / good or bad / 1 FDA approved so far – bunch in clinical trial with good potential – also use this to calculate enrichment rate? Interventions is very important. To challenge your model. Ignore the Ranking.

Drug – Target relationship

Host – Virus relationship

PP

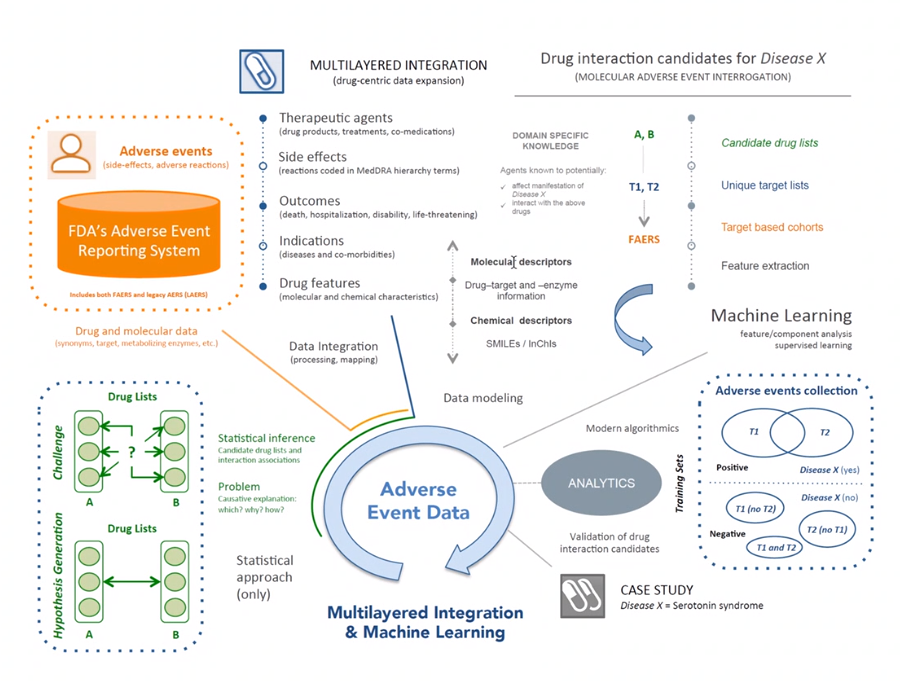
GT of the drug w/ their potential to treat

Complex drug-target-host-pathogen

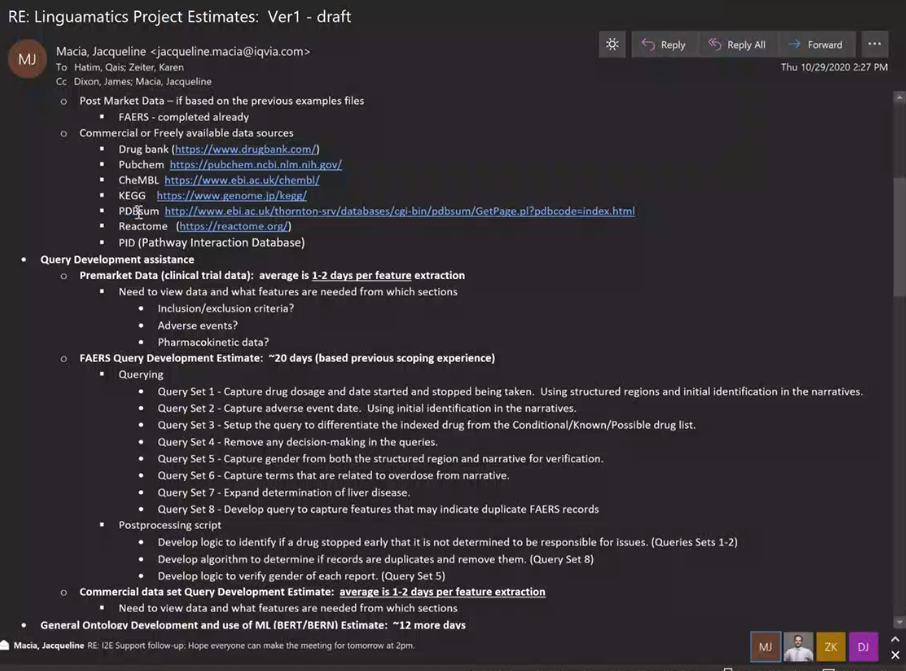
Interpret together – to enrich the repurposing candidates via AI algorithms

AI Powered Network Pharmacology to ID novel drug treatments for Covid-19?

Qais – data analysis – earlier this year w/ Germany (Molecular effect) company – group of scientists.



Platform – utilizing ML on predicting DDI – already have whitepapers – really can share with SAS – involve them to utilize



Qais took SS DDI – how we can get interactions, outcomes of the drug. Diff info about the drug.

* Will share w/ Dr Liu.
* Theo PHD to bio molecular – pay them out of this one – more accessibility
* Dr Liu – great info but now we need to kickoff this project w/ the simple data to gradually to expand the model – b/c so many databases and resources – keep it in the parking lot.
* Crawl Walk Run approach

Neal – DrugBank – target info for these drugs – experiment verified info

Planning ahead – putting together scenarios – Qais working with Janis Deloitte team looking at clinical trial to bring that in. Those have people have schedule priorities. JANIS team – CT in the EDR most of the data is not standardize – drug / NDA – retrieve for you in JMP Clinical format – analyze it – standardize from the CTs – domain based – every CT has several study – filtered them by certain criteria in a way we have – Harvard D2D interactions 25 – 1900 studies filtering on demo..etc. –submitted JMP format – using it.

Ryan – analyzing CT – what public data we can utilize? On his to-do list. Completed vs screening processes. 1 good way to categorize them first – how many CT to treat covid-19, what phase..etc. ClinicalTrials.gov is the site.

Jorge – chemical compound similarity – many ways to measure it. Similar chem structure – biological profile or Transcriptomic pairwise similarity. Exploratory techniques based on the pairwise distances.

FDA formula – similarity profile. @Qais to send to Jorge.

Share with us? Think about something else. Robert/Hamza – something we need to read

JMP – Russ – Carlos.

Key questions to address:

* Lots of info to digest. Still Forming – Storming phase
* Norming and storming
* Norming and performing

Drug – Target relationship

Host – Virus relationship

PP

GT of the drug w/ their potential to treat

Complex drug-target-host-pathogen

Interpret together – to enrich the repurposing candidates via AI algorithms

How we can leverage AI Powered Network Pharmacology to ID novel drug treatments for Covid-19?

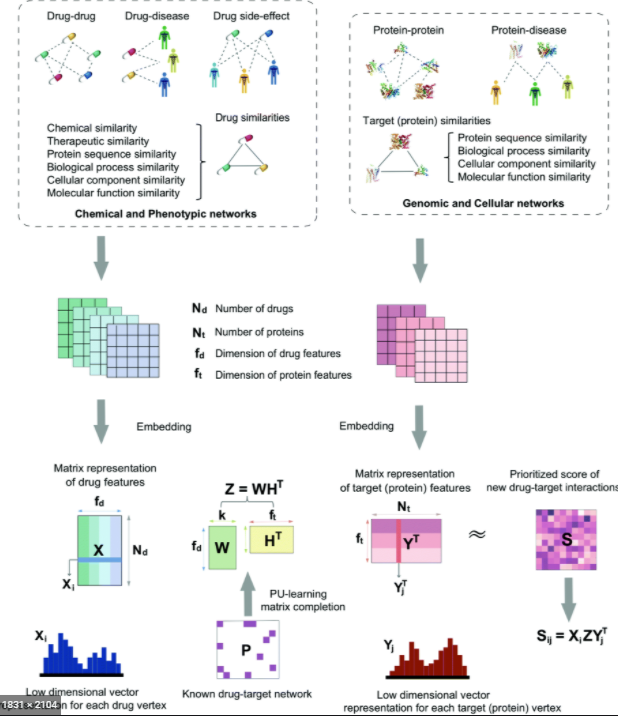
Next session Jan 8th

@Dr Liu to provide a shared login account for Drug Bank access. Back up plan is to use Jorge’s EDU account. <https://go.drugbank.com/covid-19#drugs>

@Qais to provide the standard FDA formula for chemical compound similarity measurements to Jorge & team to assess exploratory techniques based on pairwise distances..etc.

@Qais will identify and work with the backup FDA COR on the security clearances.

@Ryan will assess the publicly accessible clinical trial data (ClinicalTrials.gov) we can leverage as input to our holistic evaluation. How do we better categorize them first, how many clinical trials are needed, what phases are they in to treat Covid-19?

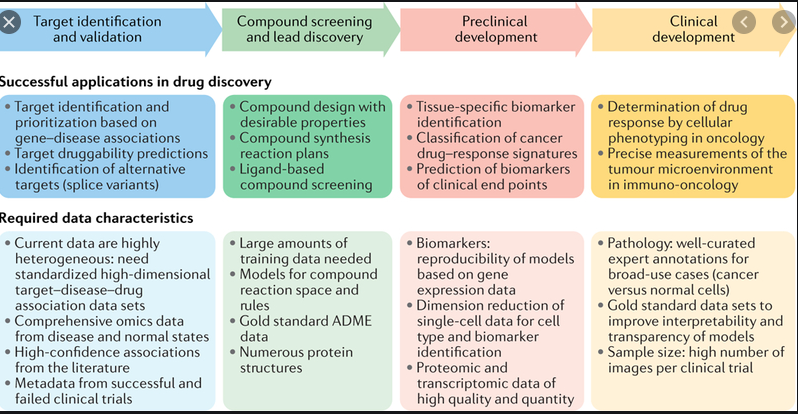


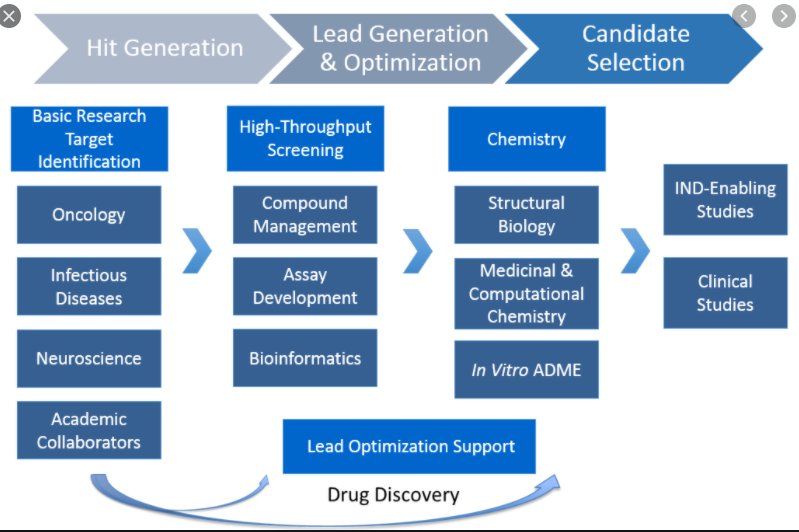
<https://pubs.rsc.org/en/content/articlehtml/2020/sc/c9sc04336e>

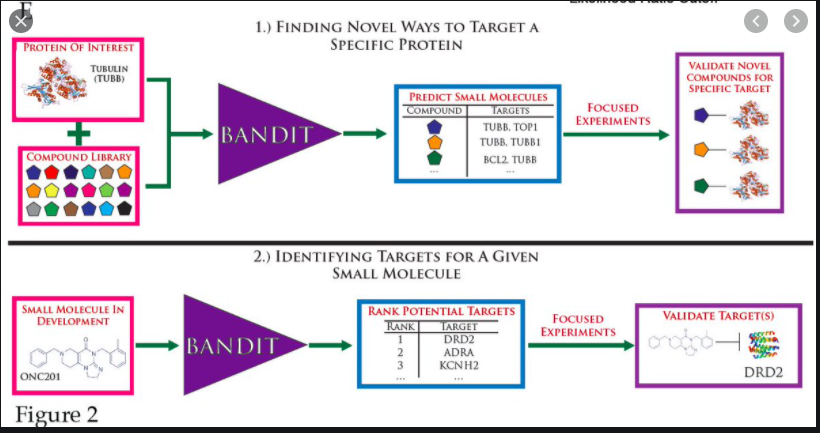
We present deepDTnet, a novel, network-based deep learning methodology for target identification and drug repurposing, which systematically embeds 15 types of chemical, genomic, phenotypic, and cellular networks, and predicts new molecular targets among known drugs under a PU-learning framework. Most importantly, we experimentally validated that topotecan predicted by deepDTnet has a high inhibitory activity against human ROR-γt. We subsequently showed that topotecan has potential therapeutic effects in a mouse model of multiple sclerosis.

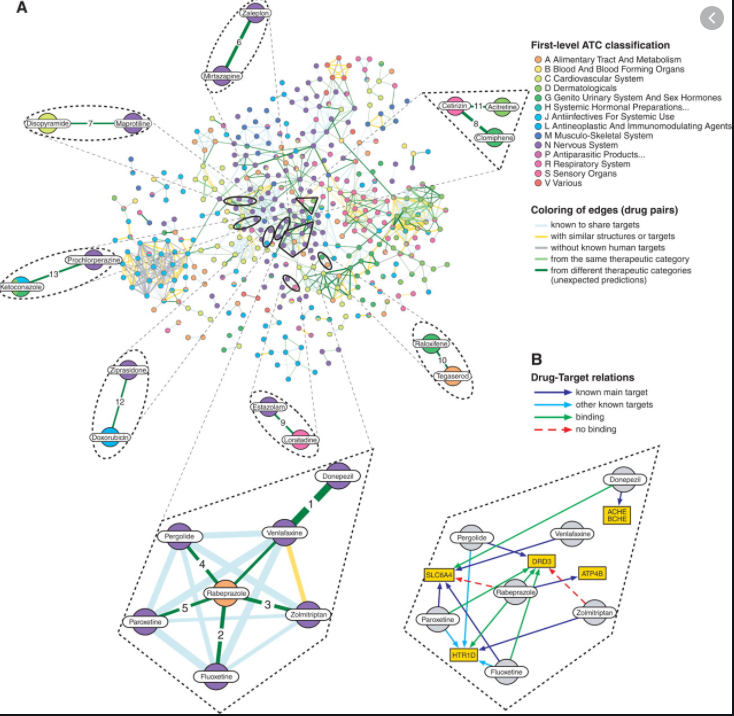
To the best knowledge of the authors, this is a systematic deep learning study that integrates the largest biomedical network datasets for target identification, drug repurposing, and testing of findings experimentally. In this way, we can minimize the translational gap between pre-clinical testing results in animal models and clinical outcomes in humans, which is a significant problem in current drug development.

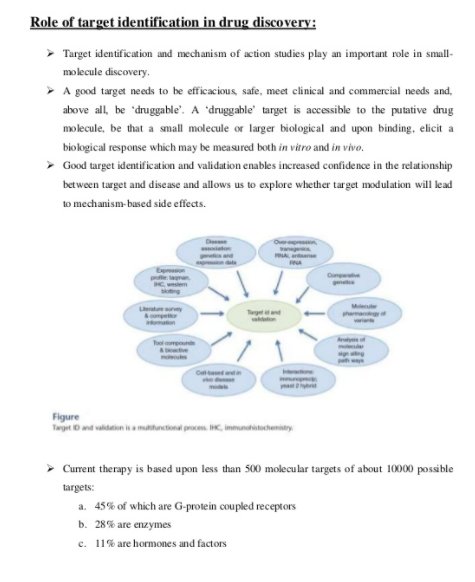
In summary, our findings suggest that target identification and drug repurposing can benefit from network-based, rational deep learning prediction in order to explore the relationship between drugs and targets in a heterogeneous drug–gene–disease network. From a translational perspective, if broadly applied, the network-based deep learning tools presented here could help develop novel, efficacious treatment strategies for multiple complex diseases.

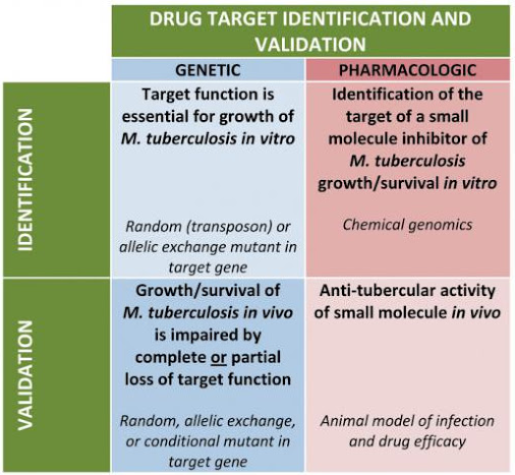




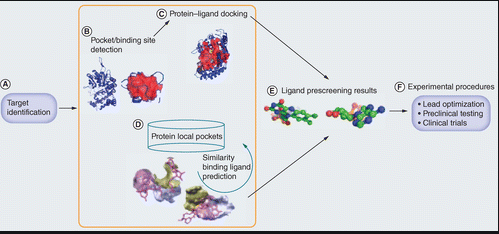


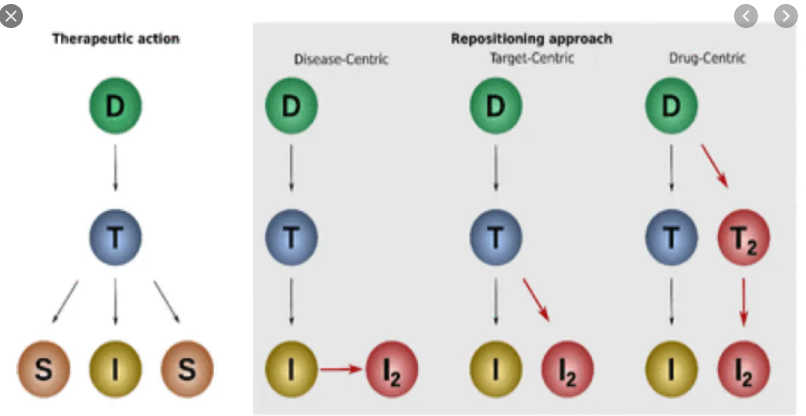




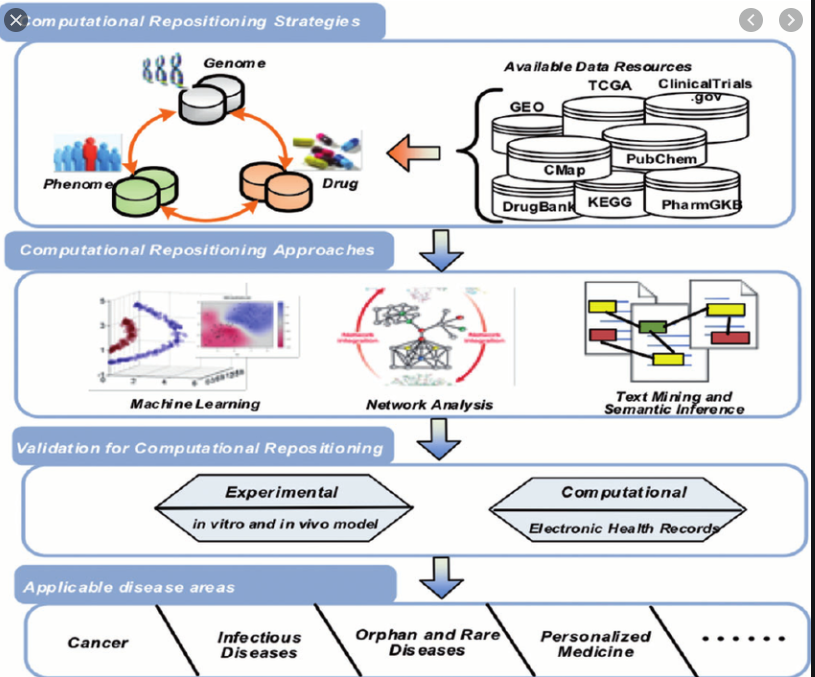


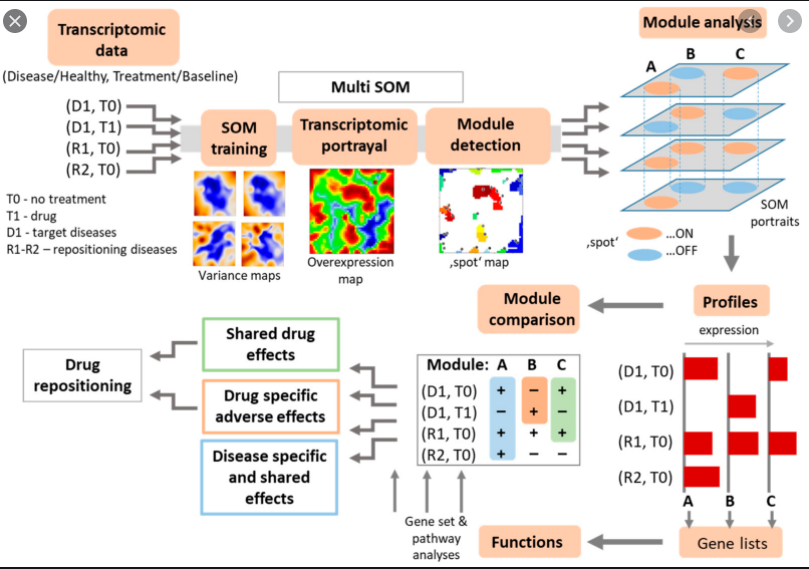






<https://engineering.osu.edu/news/2020/07/drug-repurposing-knowledge-graph-could-help-find-covid-19-treatments>





**Transcriptomics** is the study of the transcriptome—the complete set of RNA transcripts that are produced by the genome, under specific circumstances or in a specific cell—using high-throughput methods, such as microarray analysis.

@Dr Liu will provide a shared account to access Drug Bank

<https://go.drugbank.com/covid-19#drugs>