Graph in Drug Discovery

Advanced Training
2020. 11. 26

Erkhembayar J. Ph.D

Korea AI Center for Drug Discovery and Development

https://www.nature.com/articles/nbt0717-604

AGENDA

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

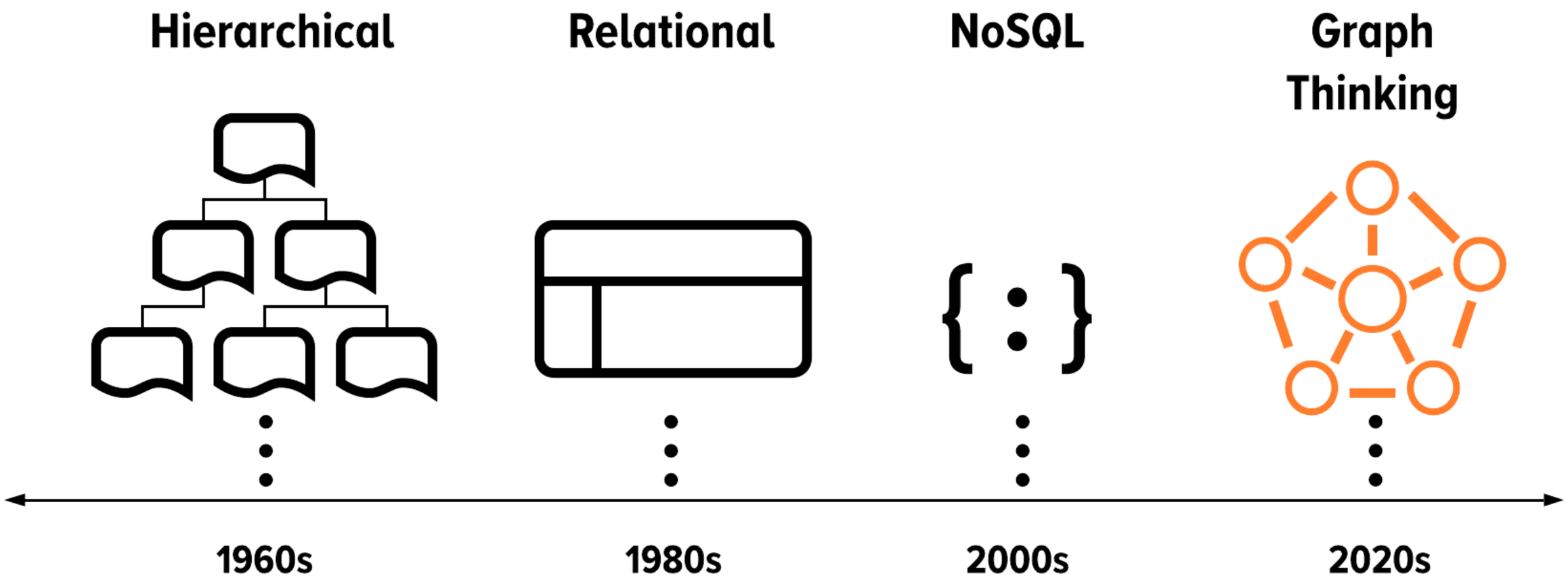
	HOURS	CONTENT
GRAPH IN DRUG DISCOVERY	**GENERAL OVERVIEW** 09.00 – 9.40	Why Graph is important ? How the graphs are used in Drug Discovery? COVID19: Drug Repositioning Knowledge Graph
CODING SESSSION (colab)	**CODING SESSION** 10:00 - 11.40	How to Construct your own Graph? Building Compound to Graph from Scratch
CODING SESSION (colab)	**CODING SESSION** 11:40 - 12.00	Knowledge –based Drug Repositioning Source Code Understanding
	LUNCH	

GRAPH IN DRUG DISCOVERY

General Overview of Graph Application in the Field

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Era of Data Representation



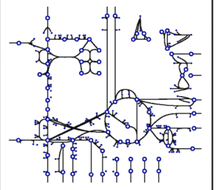
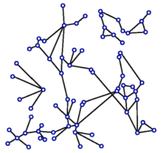
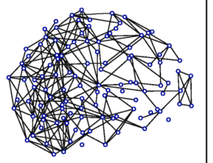
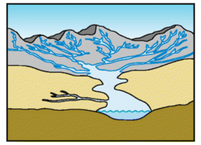
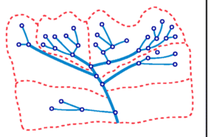
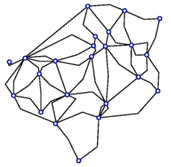
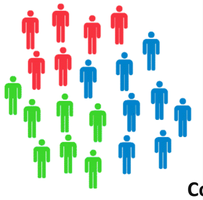
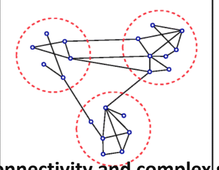
Denise et. al. 2020

Graphs

- Complex problems
  - are the individual problems that are observable and measurable within **complex systems**
- Complex systems
  - System that composed of many **individual components** that interact each other

GRAPHs are good at solving complex problem in complex systems

- Each component represented as graph
- Interaction between component could defined as edges
- ML/AI applied to individual components/or whole system to answer different questions

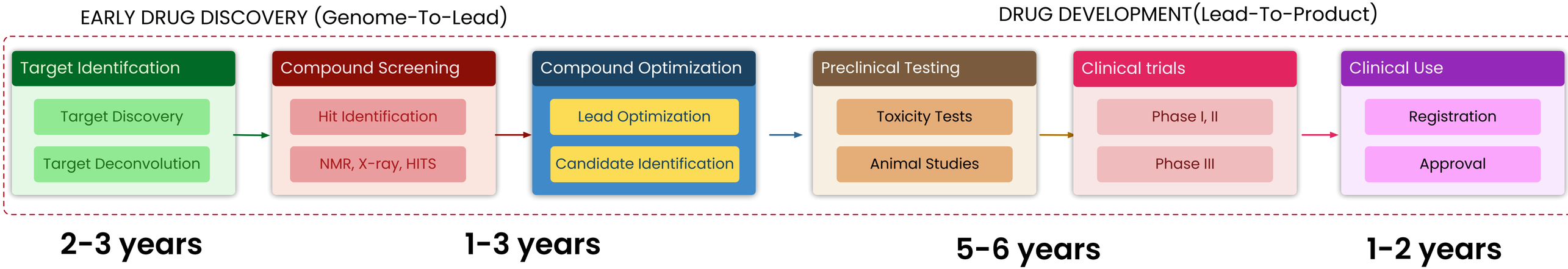
Discipline	System	Network Representation	Network Specifics
Systems Biology			Metabolic pathway map drawn from biochemical knowledge and biological intuition vs. metabolite-centric network representation of this pathway map. ○ Node: gene or regularory enzyme — Links represent regulation, gene enzyme association or metabolic reaction.
Neuroscience & Computational Neuroscience			○ Node: individual neuron or cortical area — Links between nodes. In spatially embedded networks distance between nodes is important.
Geomorphology			○ Node: spatial location within geomorphc network — Link: unidirectional weighted links --- Sub-regions of the network represent spatially discrete sub-catchments
Ecology			○ Node: representing a patch, organisms or population — Link: bi-directional links representing connected pathways between patches
Social Network Science			○ Node: representing a person (or interaction) — Links: non-directional or bi-directional links --- Groups of nodes forming a community

Connectivity and complex systems: learning from a multi-disciplinary perspective 2018.

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Drug Discovery is Complex System

- Drug Discovery is complex system with complex problem



2-3 years **1-3 years** **5-6 years** **1-2 years**

Expensive

~$2.8B

DiMasi et al.: Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. J. Health Econ. 2016, 47, 20–33

Time Consuming

~11-16 years

Matthews, et al.:"Omics"-Informed Drug and Biomarker Discovery: Opportunities, Challenges and Future Perspectives. Proteomes 2016, 4, 28.

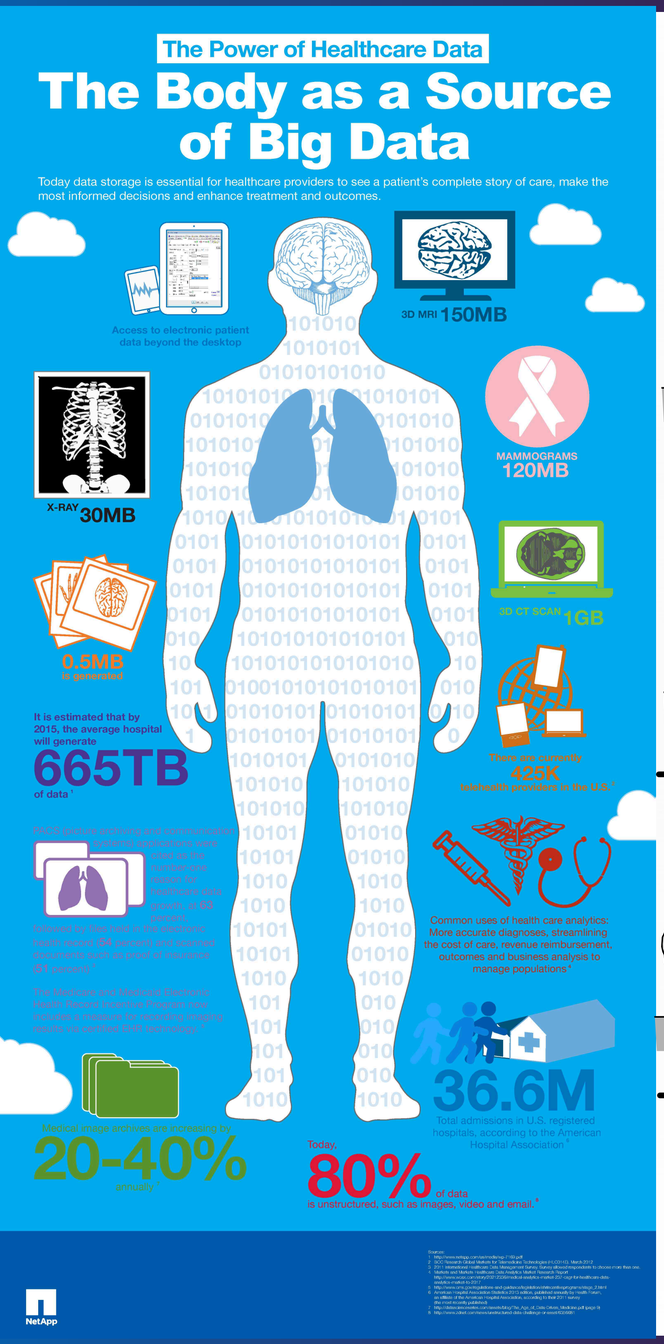
High Failure Rate

90%

Dowden, et al.: Trends in Clinical Success Rates and Therapeutic Focus. Nat. Rev. Drug Discovery 2019, 18, 495–496.

GRAPHs are good at solving complex problem in complex systems

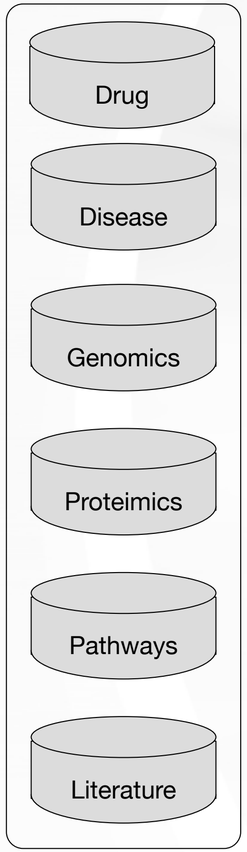
RECENTLY GRAPH + DL HAVE PROMISING RESULT in **EACH DRUG DISCOVERY COMPONENTS**

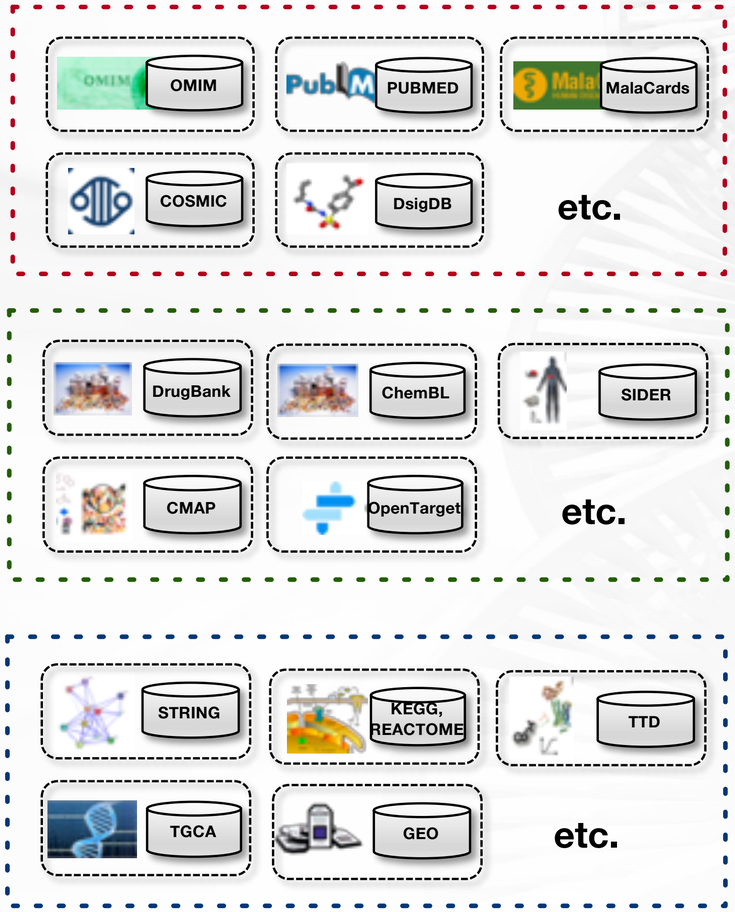


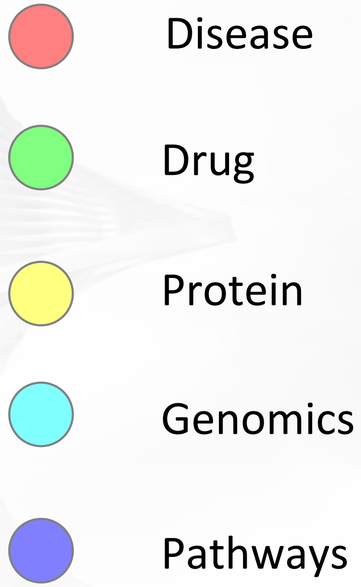
Drug Discovery Data Space

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

high-content, high-dimensional biological data





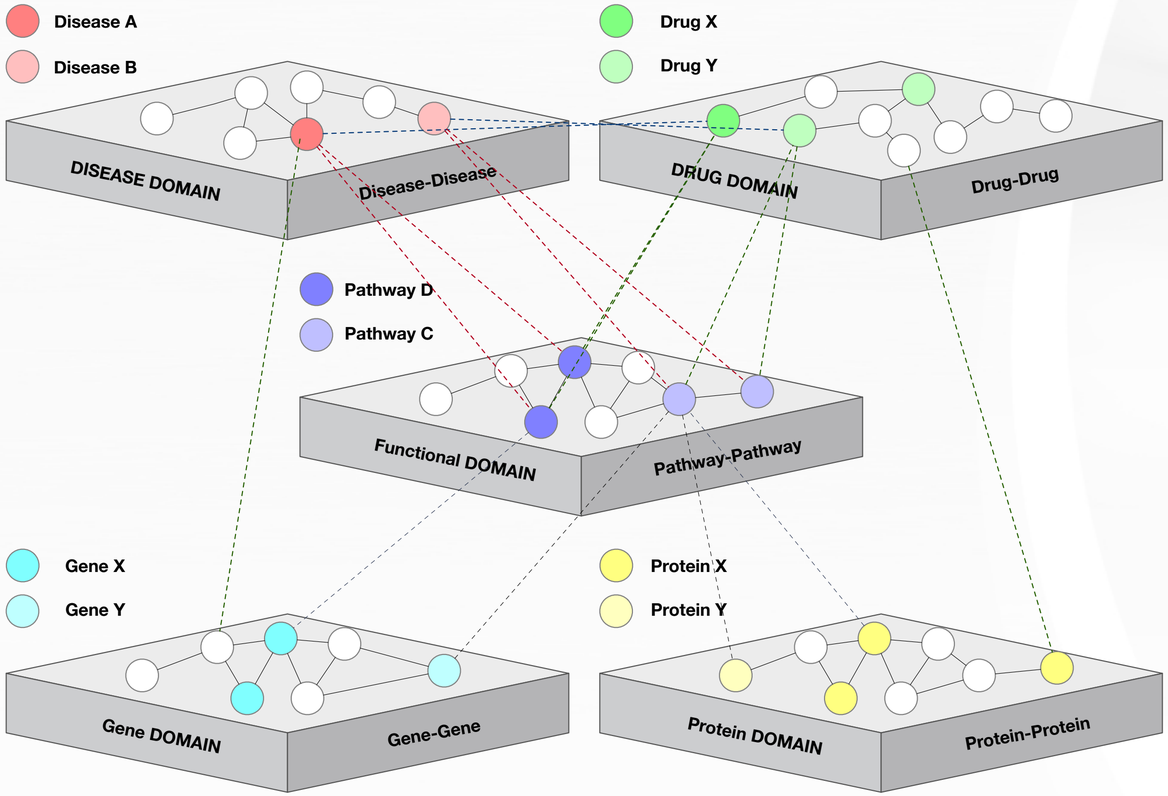


GRAPHs are representing complex data domain

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Graph Representation

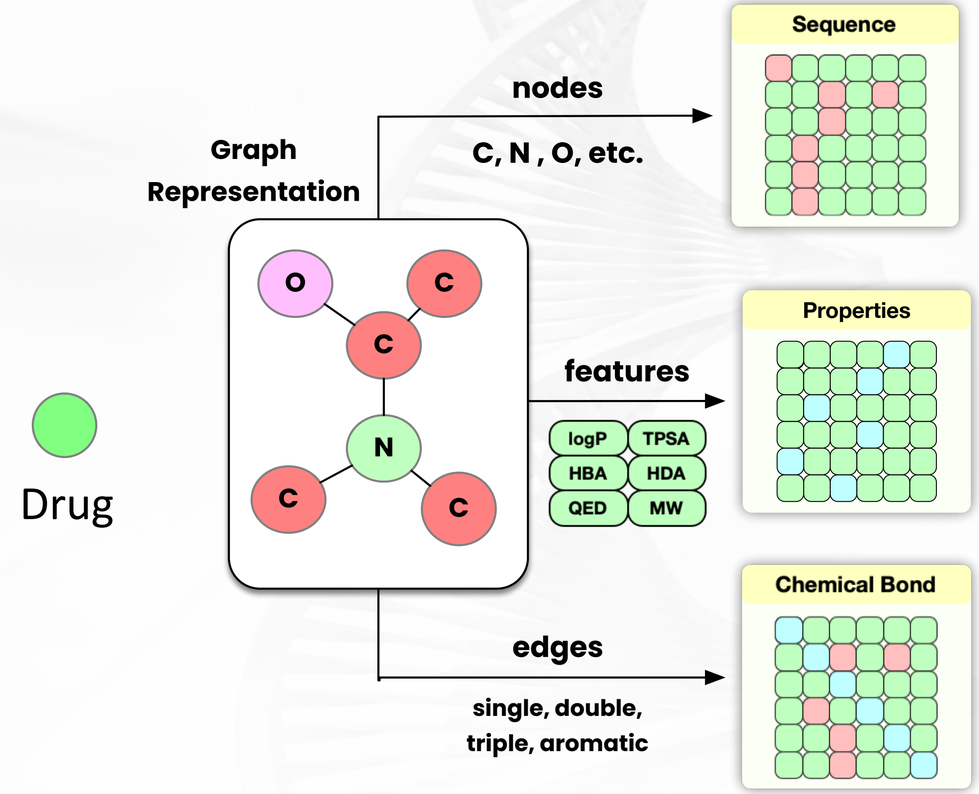
Systematic Representation



Target Identification

Drug Repositioning

Individual Representation



Drug Property Prediction

DTI, Drug Design

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Graphs in Drug Discovery

Data Scientist Perspective

Early Drug Discovery

TARGET IDENTIFICATION
- Drug-Target Identification
- Target Validation
- Protein and pathway screening in-vitro

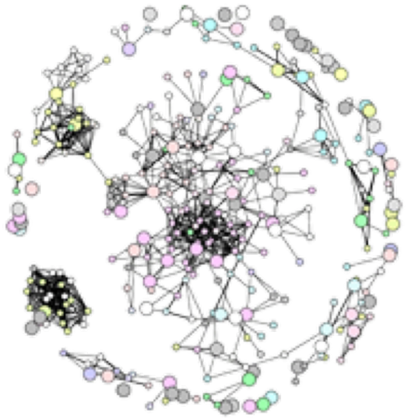
HIT IDENTIFICATION
- DRUG-TARGET INTERACTION
- QSAR
- Compound Screening

REPURPOSE EXISTING DRUGS
- Use Existing Drugs
- New Indication for Drugs

HIT TO LEAD
- Lead Identification
- Lead Optimization

Input
DISEASE, TARGET(Omics, Pathways, PPI)
COMPOUND, TARGET
DRUG, TARGET RELATIONSHIP
COMPOUND, TARGET

GRAPH
Representation and Graph Learning



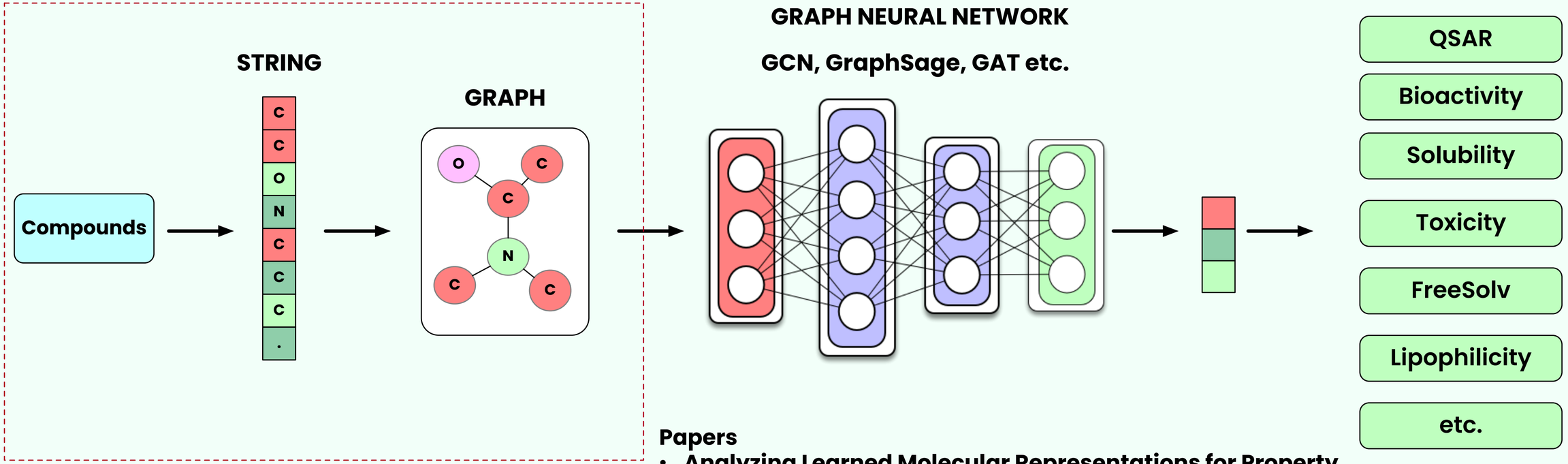
INDIVIDUAL REPRESENATION
SYSTEMATIC REPRESENATION

Output
NEW TARGET
COMPOUND (HITs)
New Indications, Drug Candidates
Drug Candidates (LEADs)

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

DRUG PROPERTY PREDICTION

Data Augmentation

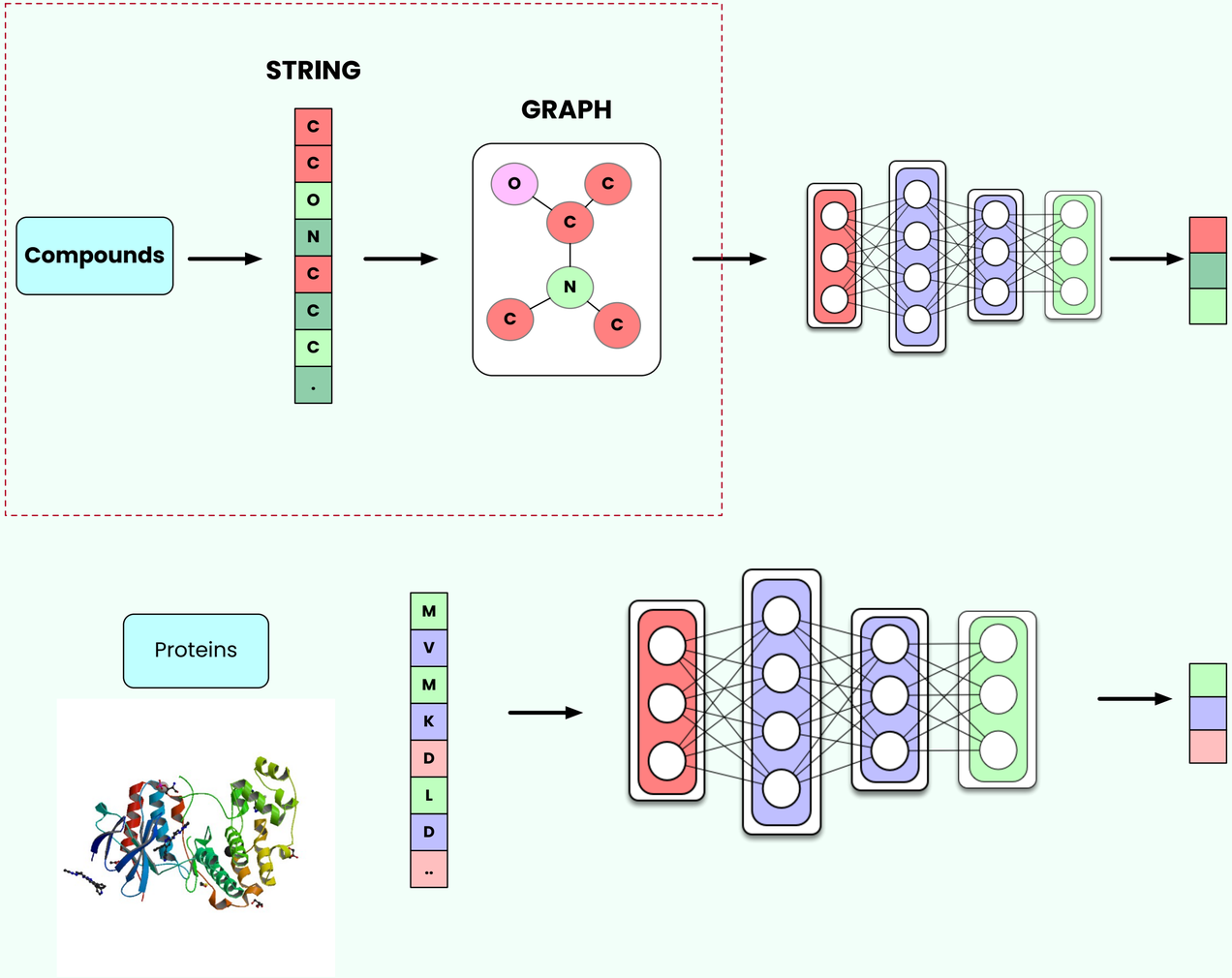


Papers

- **Analyzing Learned Molecular Representations for Property Prediction, 2019**
- **Molecule Attention Transformer, 2020**
- **Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism, 2020**

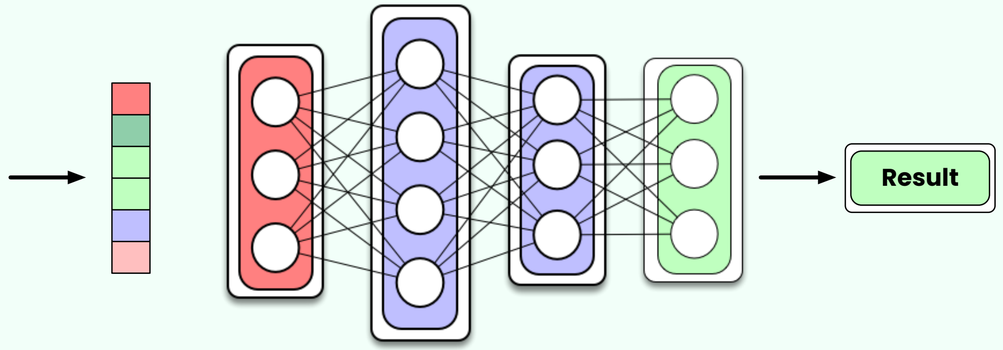
KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

DRUG-TARGET INTERACTION PREDICTION



Papers

- **GraphDTA 2019**
- **MONN: A Multi-objective Neural Network for Predicting Compound-Protein Interactions and Affinities 2020**

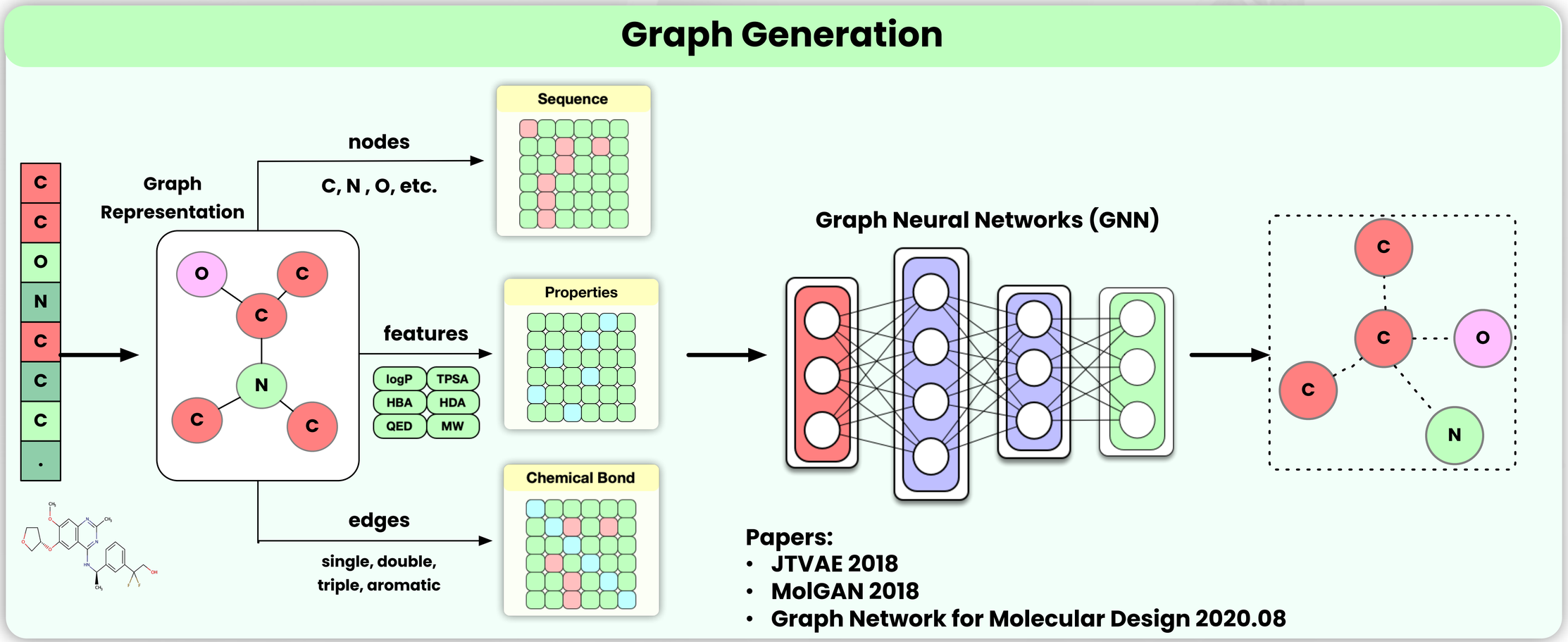


#	Compound	Protein	IC50/EC50	Kd/Ki
1	ChemicalID	ProteinID	0.99	0.99
2	...	...	...	...
3			...	...
4			...	...
N			...	...

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

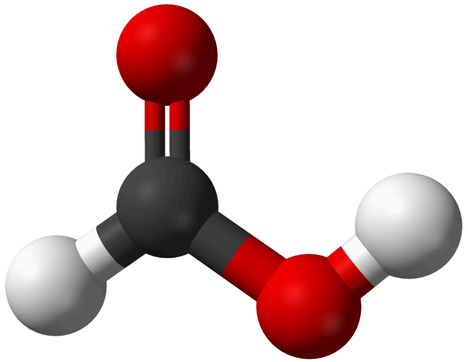
Graph Generation

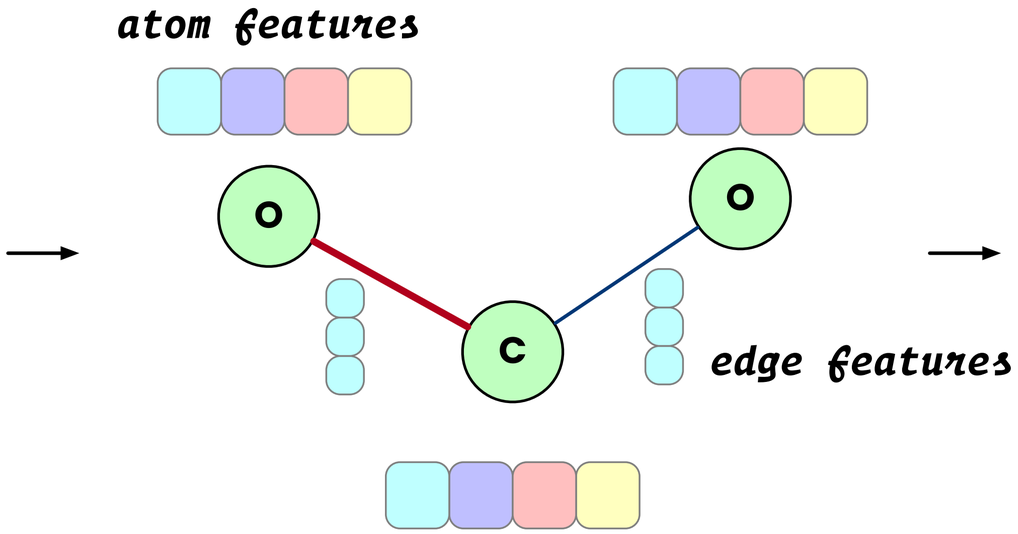
- Graph Neural Networks
  - Input: Graph; Output: Graph



KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

How Graph is Used In Chemical Space ?

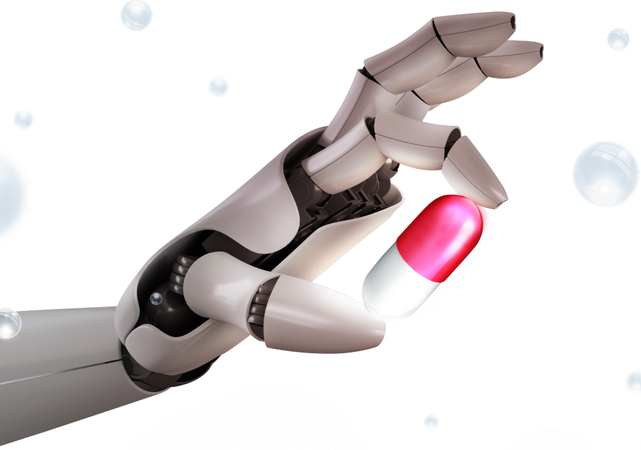


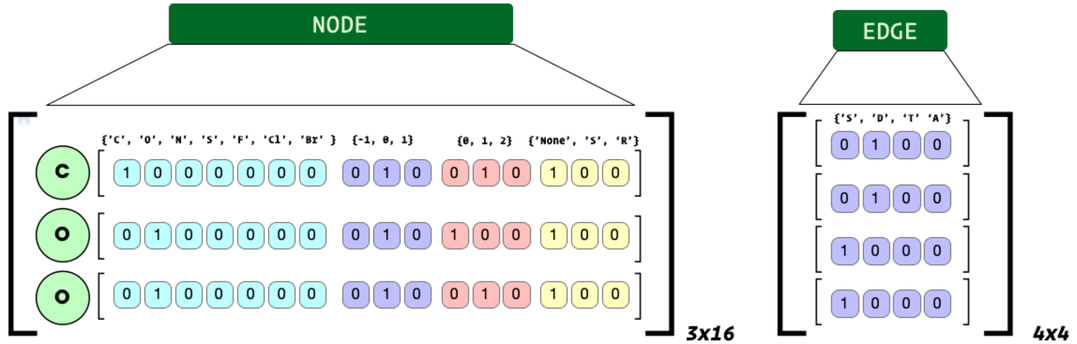


PROPERTY PREDICTION

DRUG-TARGET INTERACTION

DE NOVO DRUG DESIGN





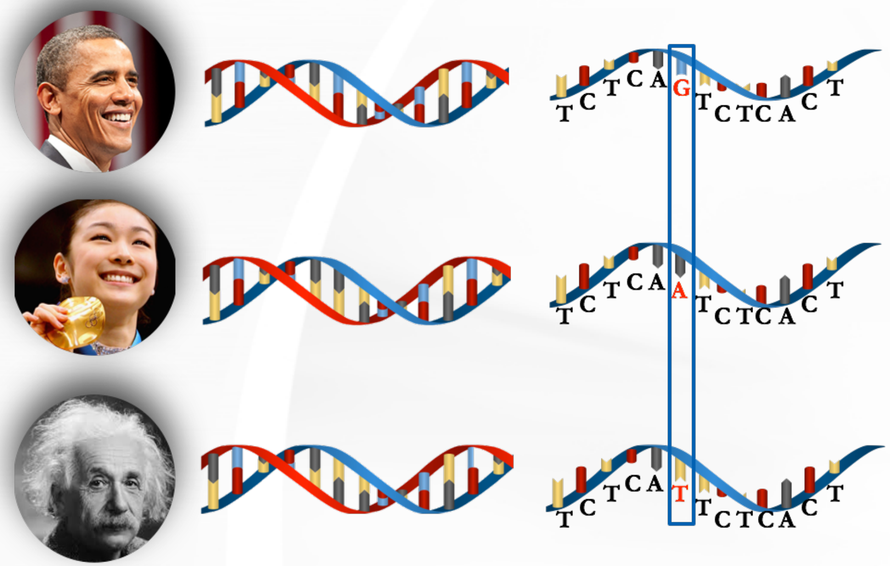
CODING SESSION: How to Make Graph Augmentation

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

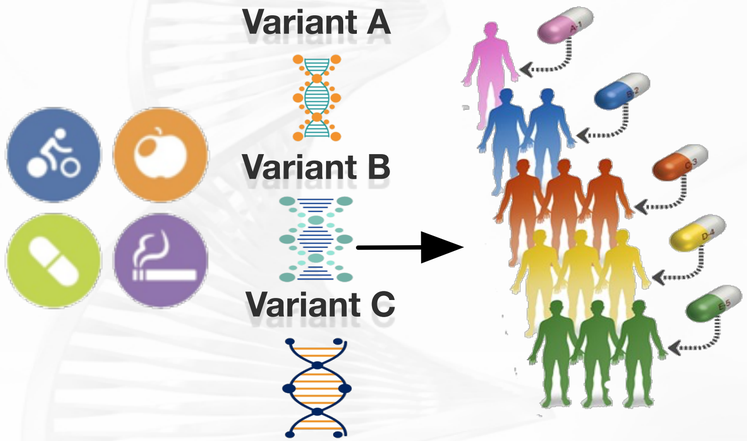
NATIONAL CANCER INSTITUTE
PRECISION MEDICINE IN CANCER TREATMENT

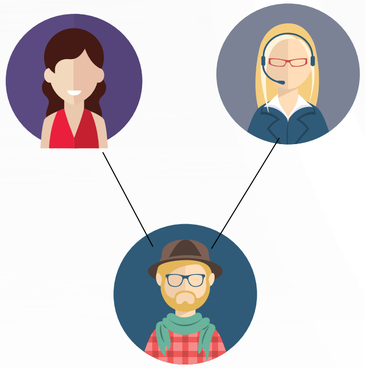
Discovering unique therapies that treat an individual's cancer based on the specific genetic abnormalities of that person's tumor.

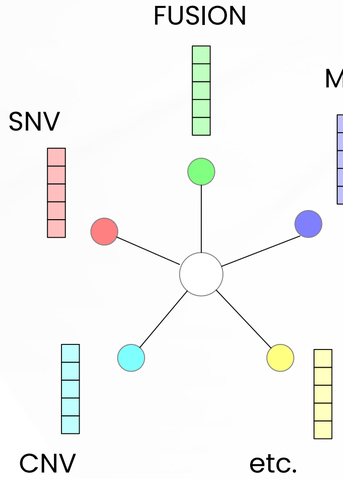
www.cancer.gov

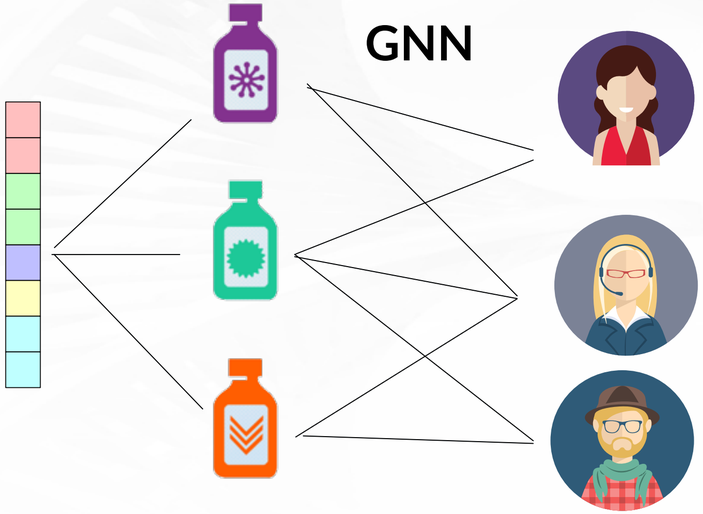


Personal Genomic Profile



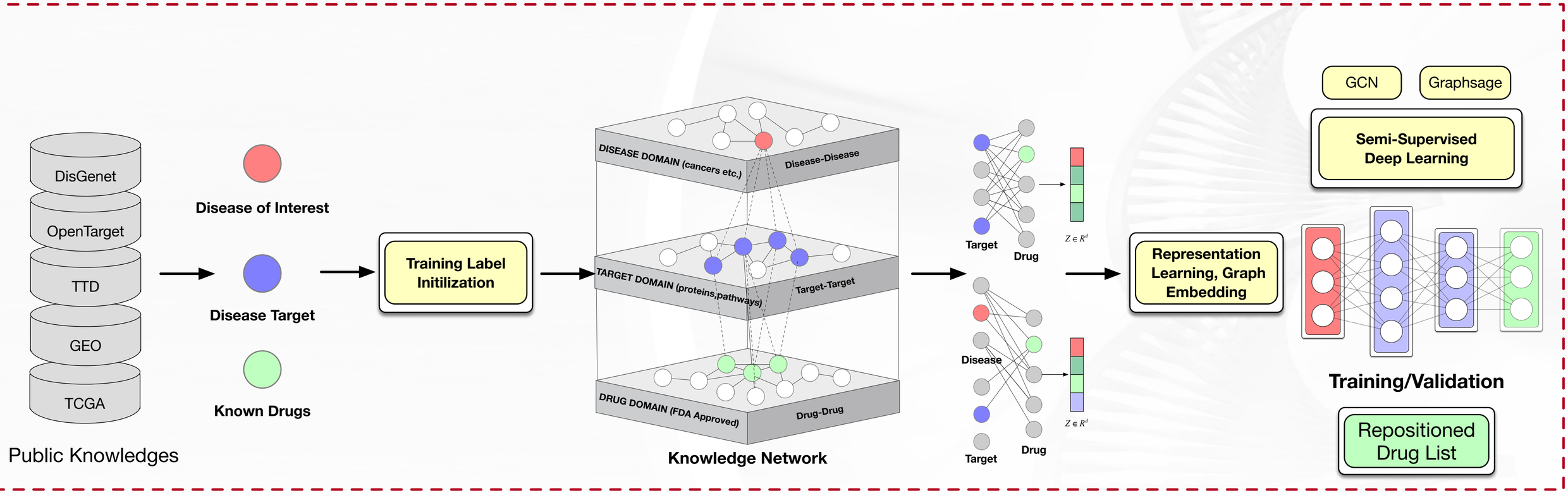






KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Drug Repositioning



DRUG REPOSITIONING

"The most fruitful basis for the discovery of a new drug is to start with an old drug"

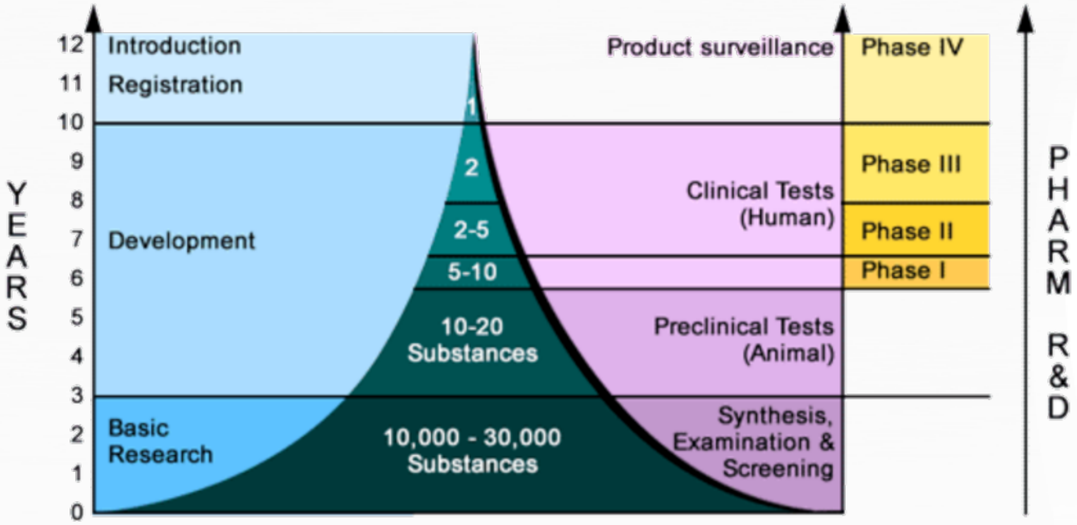
Nobel Prize-winning pharmacologist Sir James Black

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Drug Repositioning

- A method for identifying and discovering new uses for existing drugs

Drug Development Process

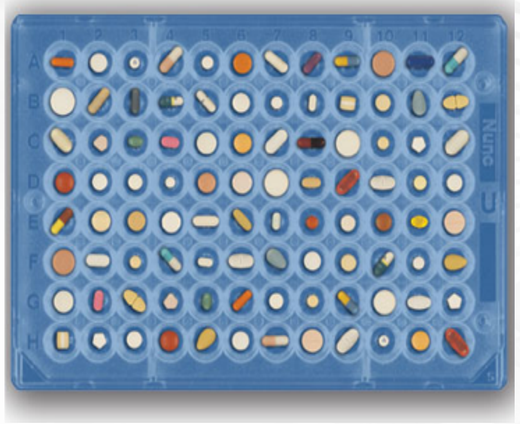


It takes too long and costs too much to bring new drugs to market

~ 15 years, US$ 1 billions for one drug

risk of drug failure

Existing Drugs



New Use

Let`s beep up efforts to screen existing drugs for new uses

nature

New uses for old drugs 2007

Curtis R. Chong[1] & David J. Sullivan, Jr[2]

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

nature biotechnology

Search E-alert

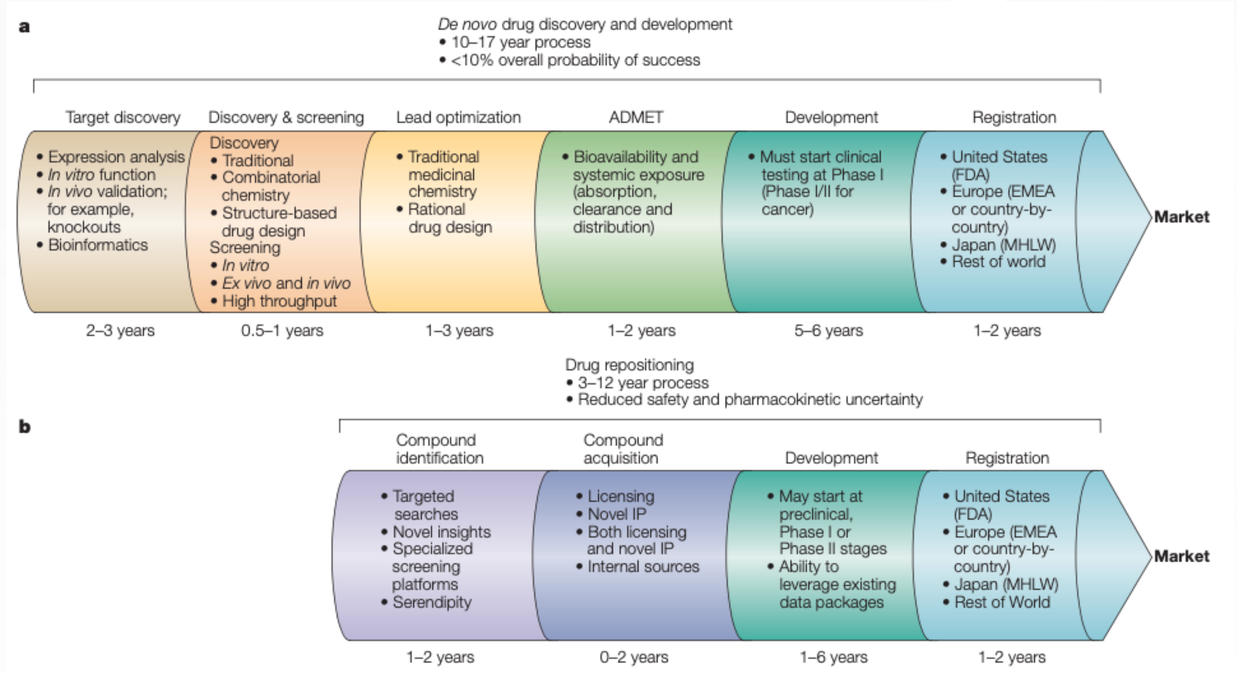
NEWS · 27 FEBRUARY 2020

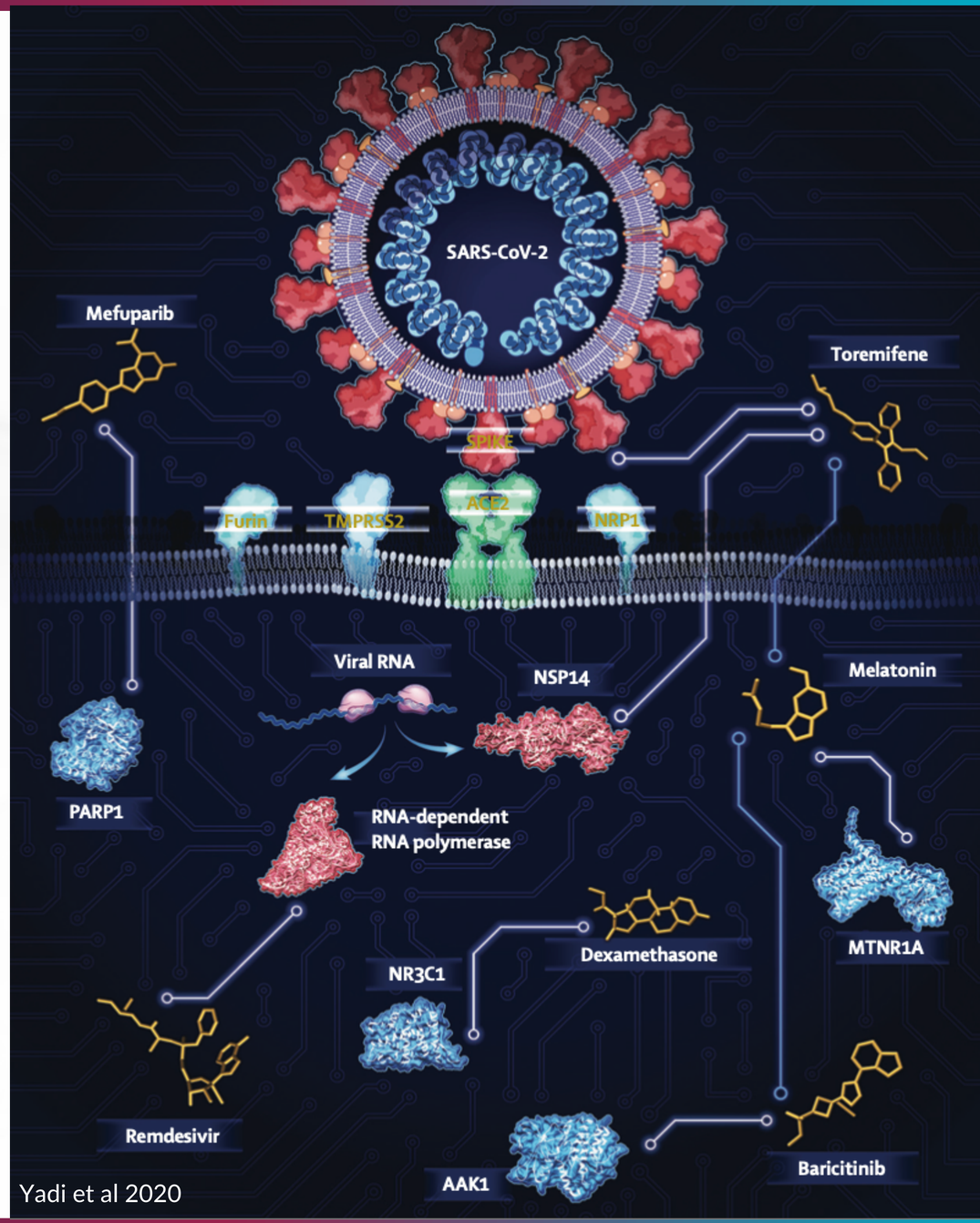
Coronavirus puts drug repurposing on the fast track

Existing antivirals and knowledge gained from the SARS and MERS outbreaks gain traction as the fastest route to fight the current coronavirus epidemic.

Advantages

- Opportunity:
  - Reduced development timelines and overall cost
  - **Fast response to pandemic/epidemic** (e.g: COVID19,)



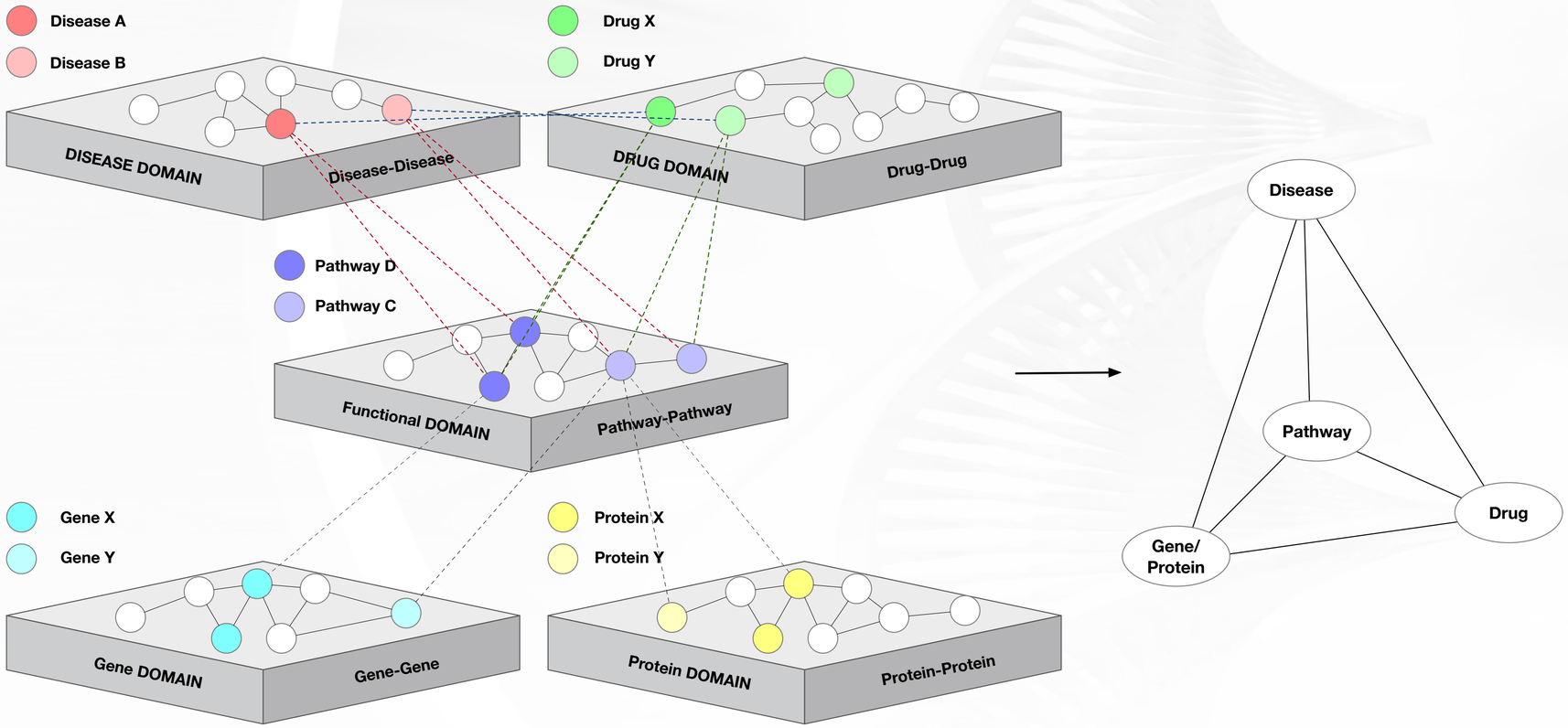


Yadi et al 2020

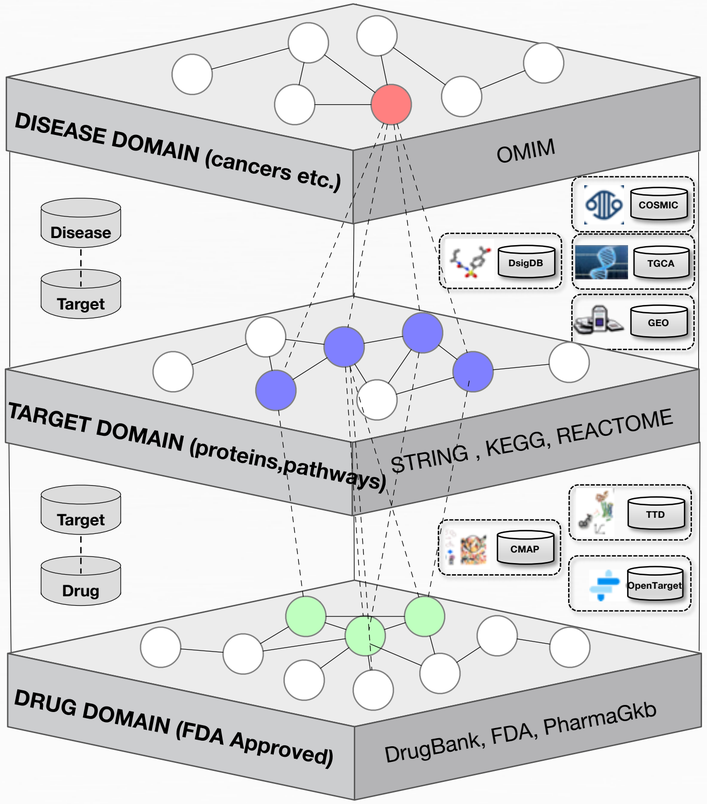
KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Traditional Method

- A classic way to repurpose drugs is through network medicine, which includes the construction of medical knowledge graphs
  - Target-centric
  - Disease-centric
  - Drug-centric



KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development



Knowledge Integration

$G_{1..n} = \{G_1, G_n\}$

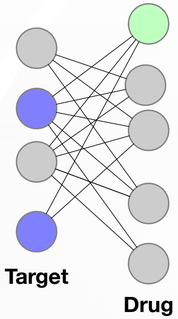
Feature Embedding

$Z \in R^d$

TRIPLETS

Centric

Targets



$Z \in R^d$

Graph based Semi-Supervised Learning

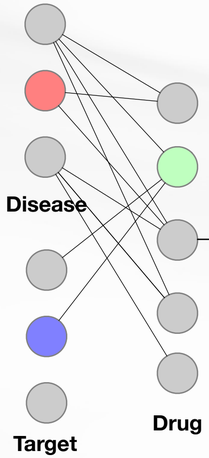
Output

DR List

#	Candidate Drugs	DR Score
1	Drug ID	0.99
2	…	…
3		…
4		…
N	….	….

Input

Disease Centric



$Z \in R^d$

Graph based Deep Learning

Output

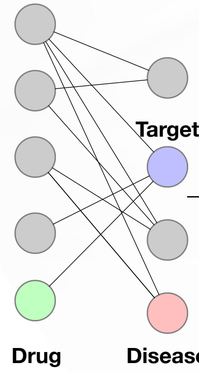
DR List

#	Disease	Indication Score
1	Disease	0.99
2	…	…
3		…
4		…
N	….	….

#	Drug	Indication Score
1	Drug ID	0.99
2	…	…
3		…
4		…
N	….	….

Input

Drug Centric



$Z \in R^d$

Graph based Deep Learning

Output

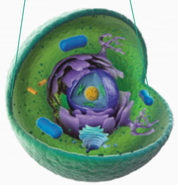
DR List

#	Disease	Indication Score
1	Disease	0.99
2	…	…
3		…
4		…
N	….	….

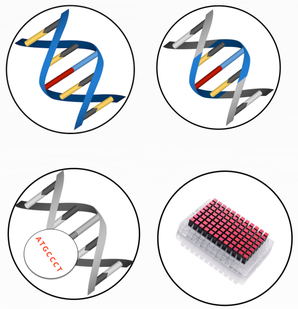
#	Drug	Indication Score
1	Drug ID	0.99
2	…	…
3		…
4		…
N	….	….

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

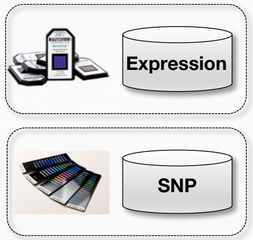
Patient

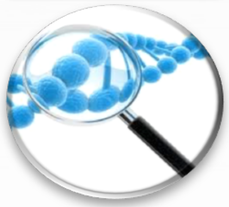


Disease Cell

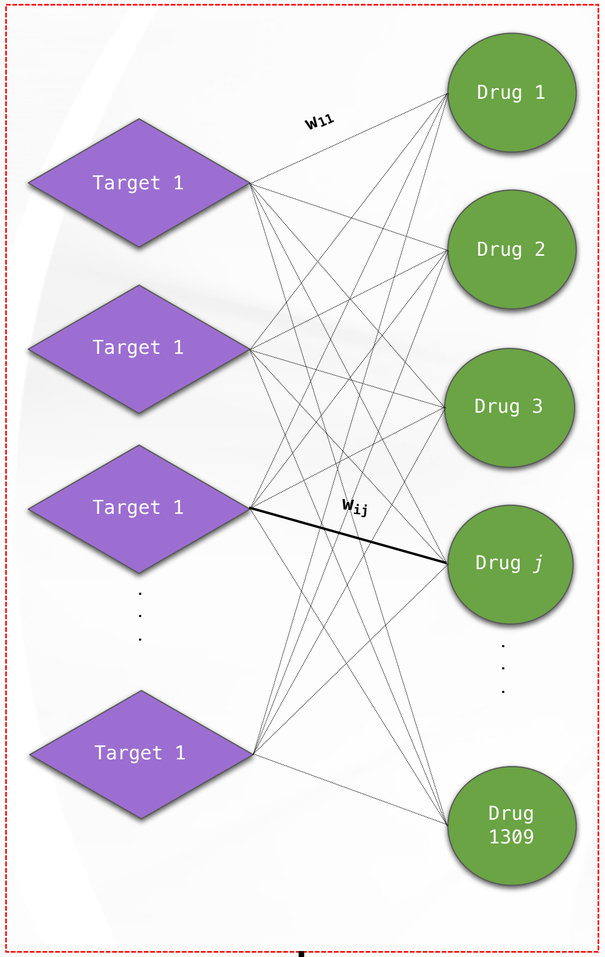


Target Source



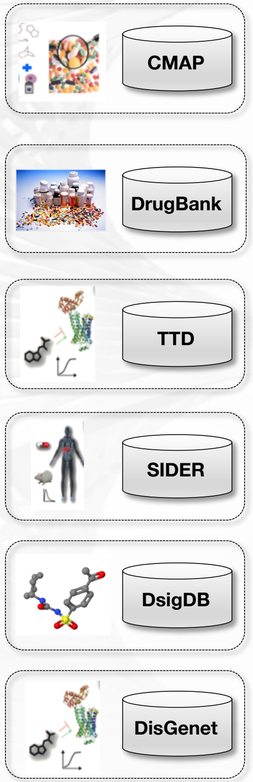


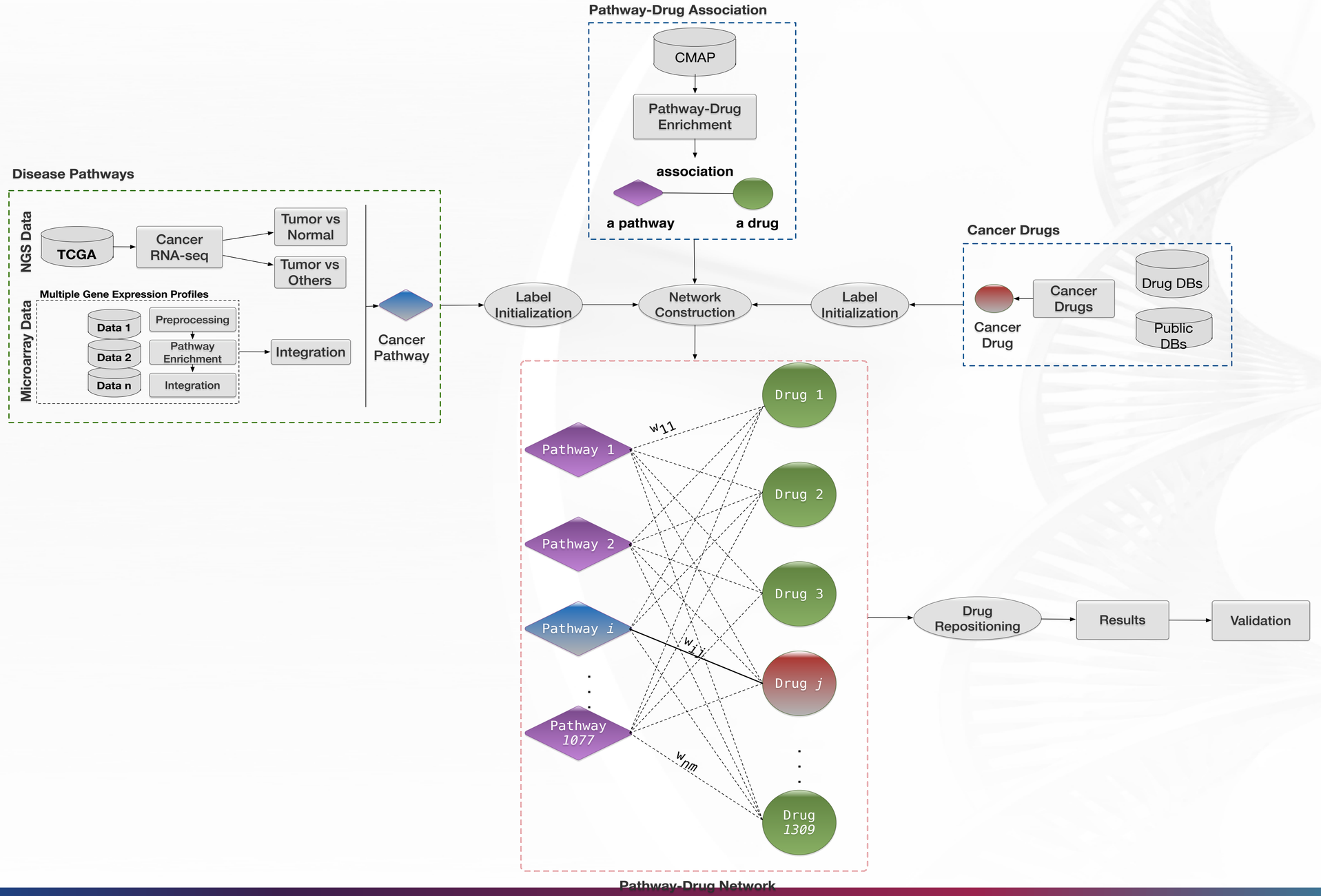
EMBEDDED TRIPLET



Drug Repositioning

Drug Resources



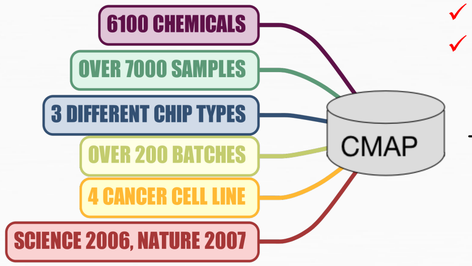

Pathway-Drug Association
CMAP
Pathway-Drug Enrichment
association
a pathway
a drug
Disease Pathways
NGS Data
TCGA
Cancer RNA-seq
Tumor vs Normal
Tumor vs Others
Microarray Data
Multiple Gene Expression Profiles
Data 1
Data 2
Data n
Preprocessing
Pathway Enrichment
Integration
Integration
Cancer Pathway
Label Initialization
Network Construction
Label Initialization
Cancer Drugs
Cancer Drug
Cancer Drugs
Drug DBs
Public DBs
Pathway 1
Pathway 2
Pathway i
Pathway 1077
w11
wij
wnm
Drug 1
Drug 2
Drug 3
Drug j
Drug 1309
Drug Repositioning
Results
Validation
Pathway-Drug Network

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

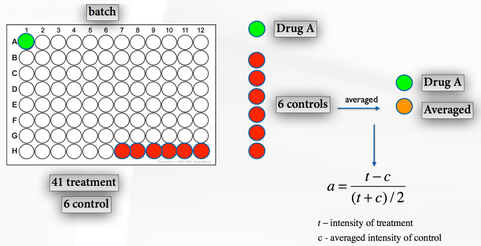
Drug Induced Omics

- CMAP was mostly used for DR, but now LINCS

	HT_HG-U133A	HG-U133A	HT_HG-U133A_EA	Total
Treated	5242	674	184	6100
Control	787	133	36	956
Total	6029	807	220	7056

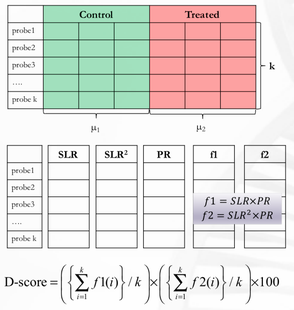


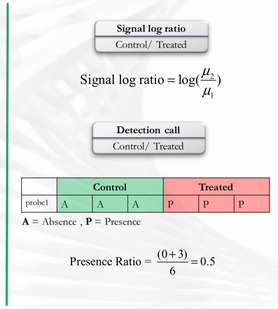
- ✓ HT-HG-U133A platform
- ✓ 106 batches
- ✓ 3906 samples
- ✓ **1309 drugs**

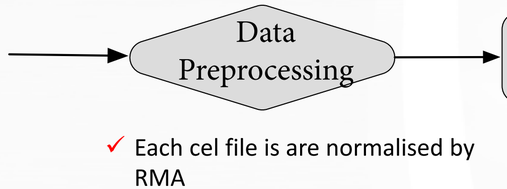


25

Pathway Enrichment







- ✓ Each cel file is are normalised by RMA
- ✓ All probe-sets less than 64 are set to 64
- ✓ Log2 transformed
- ✓ All controls for same batches averaged
- ✓ Batch effect removed by mean centering

- ✓ 222277 probes converted 19930 gene symbol

- ✓ D-score enrichment
- ✓ GSEA

- ✓ Each Pathway enrichment scores for each drug

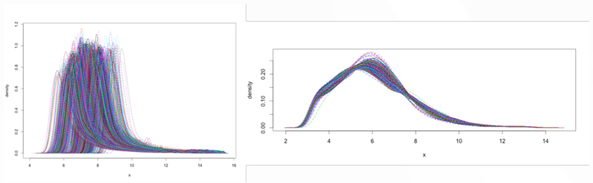
	batch 1			…		batch n	
	drugA	drug C	drugL	drugB	drugD	drugN	drugK
probe1	e_A	…	e_L		…	…	…
probe2	…	…		…	…	…	…
.							
.							
.							
probe22283				..	…	…	…

mean center within batch | mean center within batch | mean center within batch

$$z = \frac{probe(ij) - mean(within\_batch)}{std(probe_i)}$$

32

Batch effect removal



RMA preprocessing

	Drug 1	Drug 2	Drug 3	….	Drug 1309
Pathway 1	Enrichment score				
Pathway 2					
Pathway 3					
.					
Pathway 1077					

Drug-Pathway Matrix

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Nature / Cell Discovery, 16 March 2020

Zhou et al. *Cell Discovery* (2020)6:14
https://doi.org/10.1038/s41421-020-0153-3

Cell Discovery
www.nature.com/celldisc

ARTICLE Open Access

Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2

Yadi Zhou[1], Yuan Hou[1], Jiayu Shen[1], Yin Huang[1], William Martin[1] and Feixiong Cheng[1,2,3]

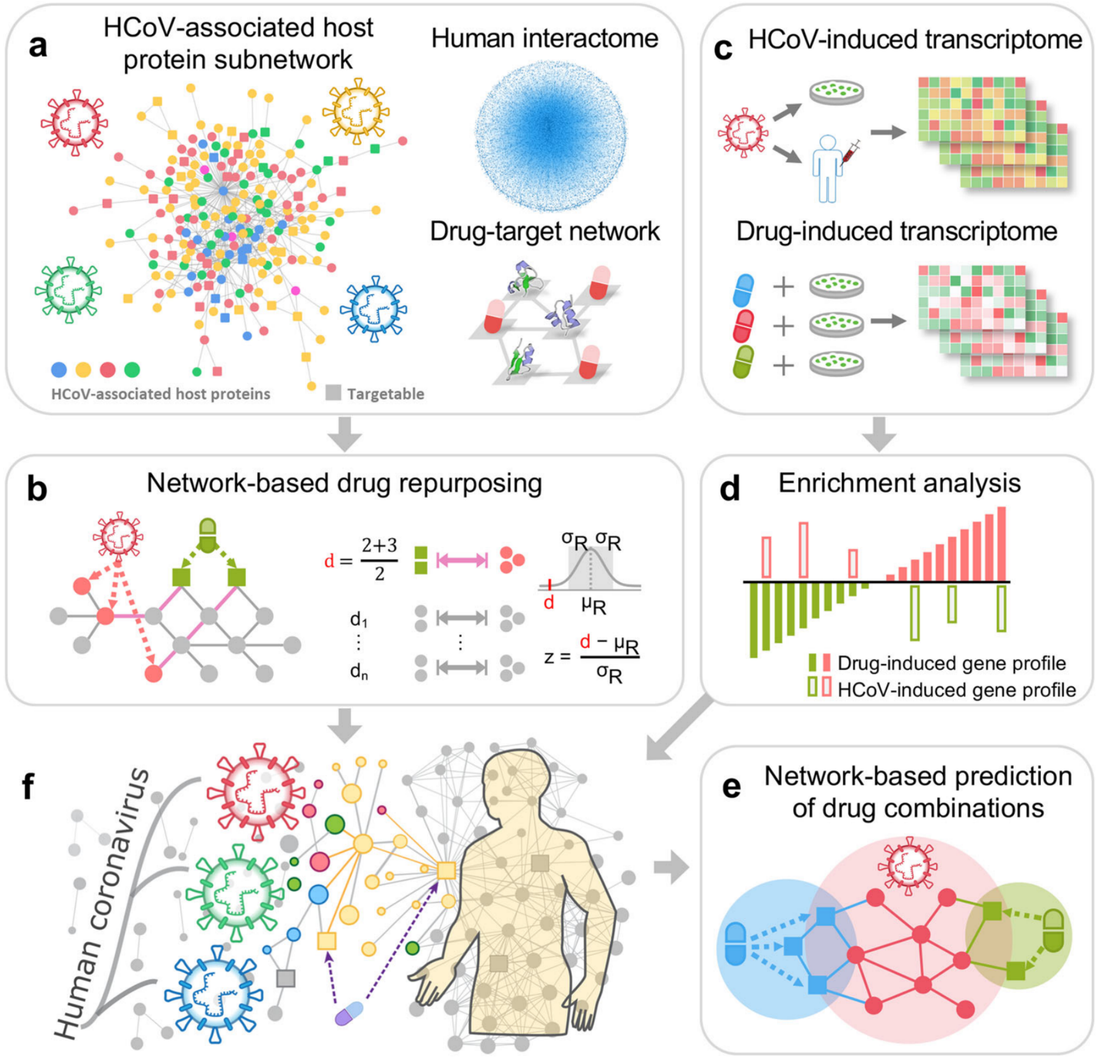
Abstract

Human coronaviruses (HCoVs), including severe acute respiratory syndrome coronavirus (SARS-CoV) and 2019 novel coronavirus (2019-nCoV, also known as SARS-CoV-2), lead global epidemics with high morbidity and mortality. However, there are currently no effective drugs targeting 2019-nCoV/SARS-CoV-2. Drug repurposing, representing as an effective drug discovery strategy from existing drugs, could shorten the time and reduce the cost compared to de novo drug discovery. In this study, we present an integrative, antiviral drug repurposing methodology implementing a systems pharmacology-based network medicine platform, quantifying the interplay between the HCoV–host interactome and drug targets in the human protein–protein interaction network. Phylogenetic analyses of 15 HCoV whole genomes reveal that 2019-nCoV/SARS-CoV-2 shares the highest nucleotide sequence identity with SARS-CoV (79.7%). Specifically, the envelope and nucleocapsid proteins of 2019-nCoV/SARS-CoV-2 are two evolutionarily conserved regions, having the sequence identities of 96% and 89.6%, respectively, compared to SARS-CoV. Using network proximity analyses of drug targets and HCoV–host interactions in the human interactome, we prioritize 16 potential anti-HCoV repurposable drugs (e.g., melatonin, mercaptopurine, and sirolimus) that are further validated by enrichment analyses of drug-gene signatures and HCoV-induced transcriptomics data in human cell lines. We further identify three potential drug combinations (e.g., sirolimus plus dactinomycin, mercaptopurine plus melatonin, and toremifene plus emodin) captured by the "*Complementary Exposure*" pattern: the targets of the drugs both hit the HCoV–host subnetwork, but target separate neighborhoods in the human interactome network. In summary, this study offers powerful network-based methodologies for rapid identification of candidate repurposable drugs and potential drug combinations targeting 2019-nCoV/SARS-CoV-2.

Introduction

Coronaviruses (CoVs) typically affect the respiratory tract of mammals, including humans, and lead to mild to severe respiratory tract infections[1]. In the past two decades, two highly pathogenic human CoVs (HCoVs), including severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), emerging from animal reservoirs, have led to global epidemics with high morbidity and mortality[2]. For example, 8098 individuals were infected and 774 died in the SARS-CoV pandemic, which cost the global economy with an estimated $30 to $100 billion[3,4]. According to the World Health Organization (WHO), as of November 2019, MERS-CoV has had a total of 2494 diagnosed cases causing 858 deaths, the majority in Saudi Arabia[2]. In December 2019, the third pathogenic HCoV, named 2019 novel coronavirus (2019-nCoV/SARS-CoV-2), as the cause of coronavirus disease 2019 (abbreviated as COVID-19)[5], was found in Wuhan, China. As of 24 February 2020, there have been over 79,000 cases with over 2600 deaths for the 2019-nCoV/SARS-CoV-2 outbreak worldwide; furthermore, human-to-human transmission has occurred among close contacts[6]. However, there are currently no effective medications against 2019-

Correspondence: Feixiong Cheng (chengf@ccf.org)
[1]Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA
[2]Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA
Full list of author information is available at the end of the article
These authors contributed equally: Yadi Zhou, Yuan Hou



A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19

Yiyue Ge[1,2,†], Tingzhong Tian[1,2,†], Suling Huang[3,†], Fangping Wan[1,†], Jingxin Li[2,†], Shuya Li[1], Hui Yang[11], Lixiang Hong[1], Nian Wu[1], Enming Yuan[1], Lili Cheng[4], Yipin Lei[11], Hantao Shu[1], Xiaolong Feng[6,7], Ziyuan Jiang[5], Ying Chi[2], Xiling Guo[2], Lunbiao Cui[2], Liang Xiao[10], Zeng Li[10], Chunhao Yang[3], Zehong Miao[3], Haidong Tang[4], Ligong Chen[4], Hainian Zeng[11], Dan Zhao[1,*], Fengcai Zhu[2,8,*], Xiaokun Shen[10,*], Jianyang Zeng[1,9,*]

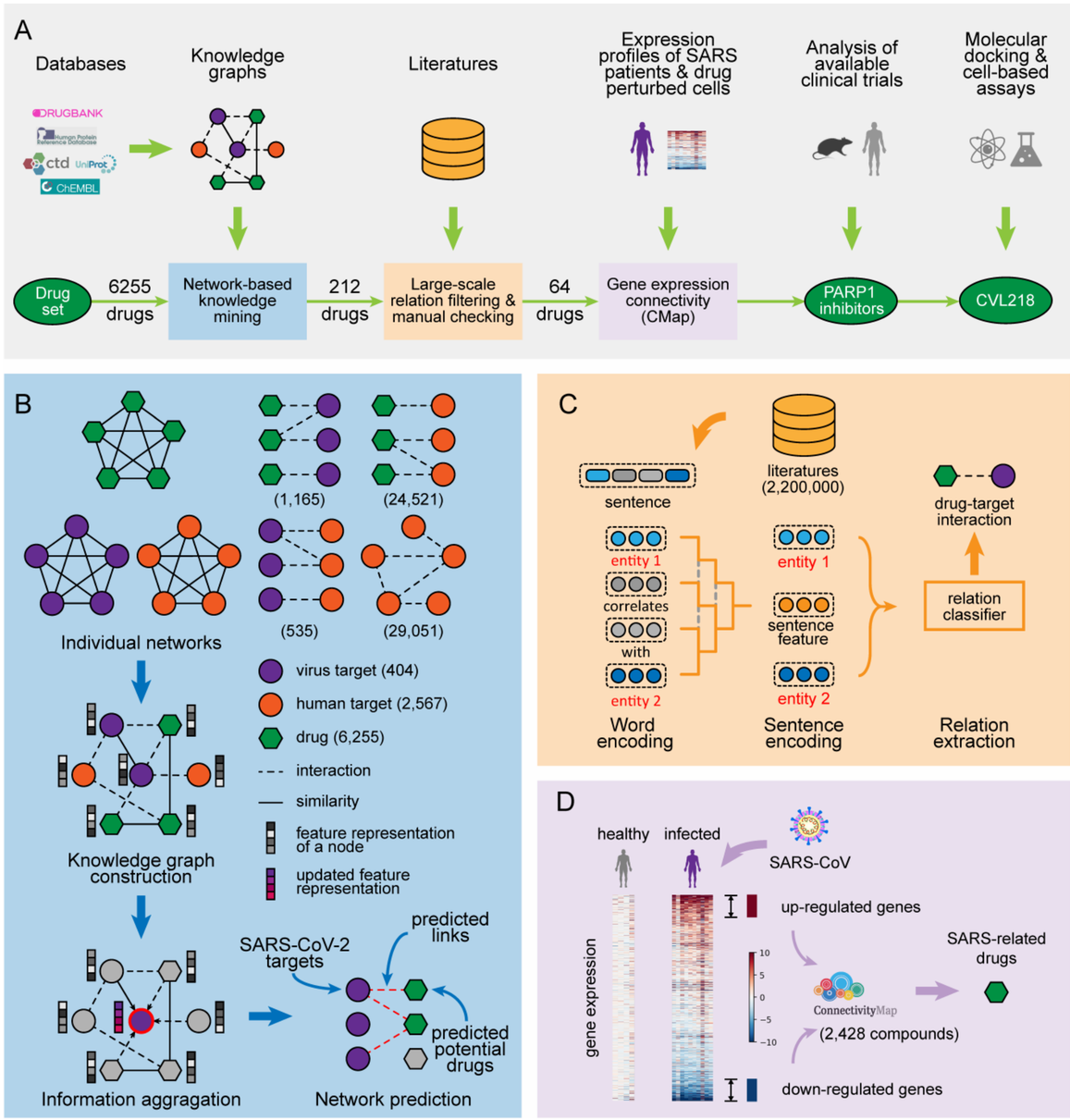
[1] *Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, 100084, China.*
[2] *NHC Key laboratory of Enteric Pathogenic Microbiology, Jiangsu Provincial Center for Diseases Control and Prevention, Nanjing, Jiangsu Province, 210009, China.*
[3] *Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, 201203, China.*
[4] *School of Pharmaceutical Sciences, Beijing Advanced Innovation Center for Structural Biology, Tsinghua University, Beijing, 100084, China.*
[5] *Department of Automation, Tsinghua University, Beijing, 100084, China.*
[6] *School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, Hubei Province, 430074, China.*
[7] *Institute of Pathology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei Province, 430030, China.*
[8] *Center for Global Health, Nanjing Medical University, Nanjing, Jiangsu Province, 210009, China.*
[9] *MOE Key Laboratory of Bioinformatics, Tsinghua University, Beijing, 100084, China.*
[10] *Convalife (Shanghai) Co., Ltd., Shanghai, 201203, China.*
[11] *Silexon AI Technology Co., Ltd., Nanjing, Jiangsu Province, 210033, China.*
[†] *These authors contributed equally to this work.*
[*] *Corresponding authors.*

Abstract

The global spread of SARS-CoV-2 requires an urgent need to find effective therapeutics for the treatment of COVID-19. We developed a data-driven drug repositioning framework, which applies both machine learning and statistical analysis approaches to systematically integrate and mine large-scale knowledge graph, literature and transcriptome data to discover the potential drug candidates against SARS-CoV-2. The retrospective study using the past SARS-CoV and MERS-CoV data demonstrated that our machine learning based method can successfully predict effective drug candidates

Email addresses: zhaodan2018@tsinghua.edu.cn (Dan Zhao), jszfc@vip.sina.com (Fengcai Zhu), steve.shen@convalife.com (Xiaokun Shen), zengjy321@tsinghua.edu.cn (Jianyang Zeng)

March 11, 2020



KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Knowledge-based DR

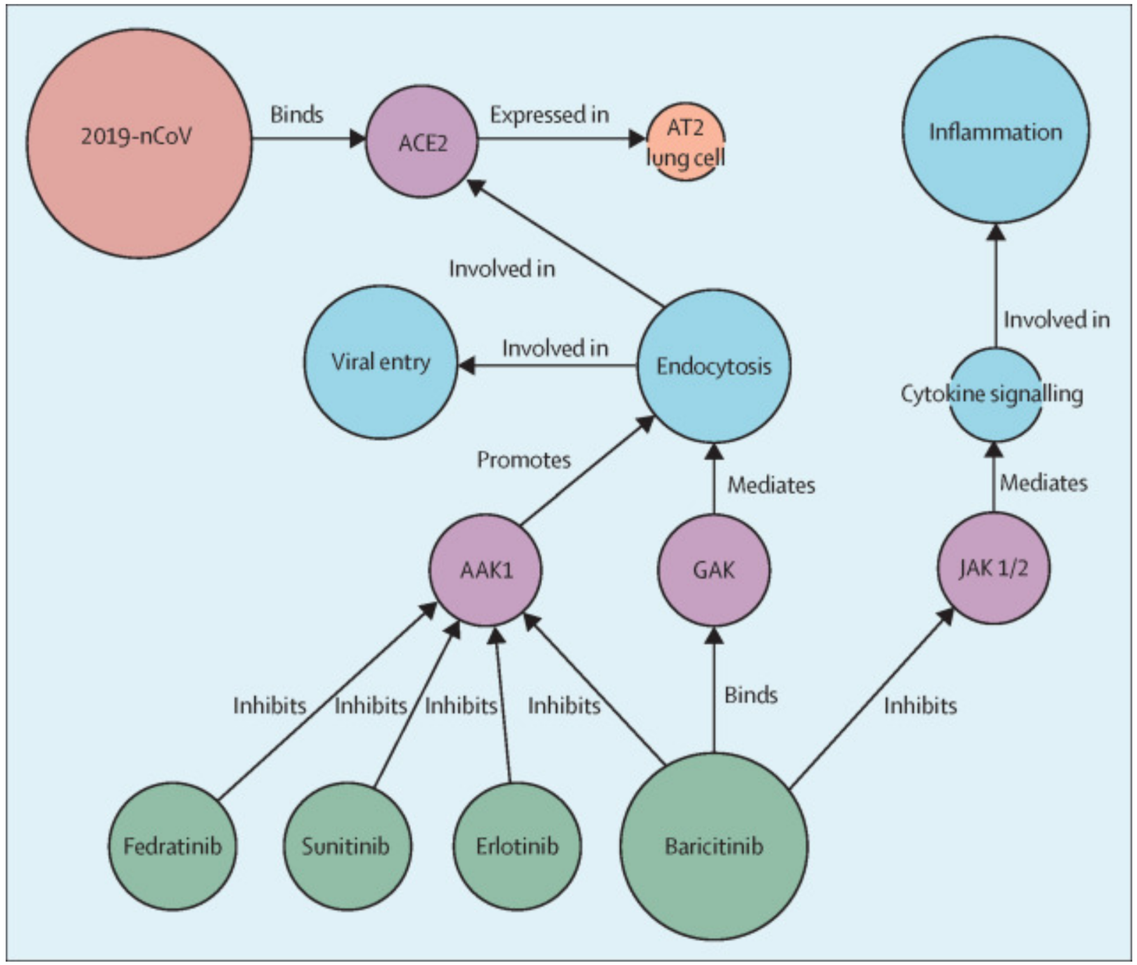
- **BenevolentAI**

CORRESPONDENCE | VOLUME 395, ISSUE 10223, PE30-E31, FEBRUARY 15, 2020

Baricitinib as potential treatment for 2019-nCoV acute respiratory disease

Peter Richardson • Ivan Griffin • Catherine Tucker • Dan Smith • Olly Oechsle • Anne Phelan • et al. Show all authors

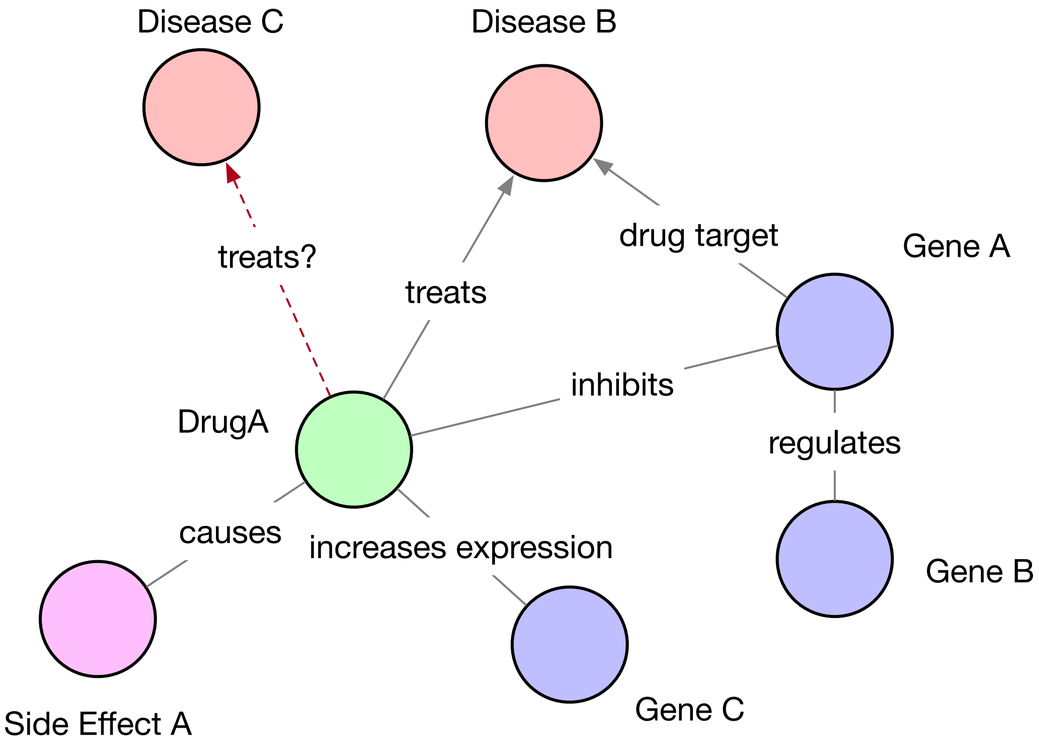
Published: February 04, 2020 • DOI: https://doi.org/10.1016/S0140-6736(20)30304-4

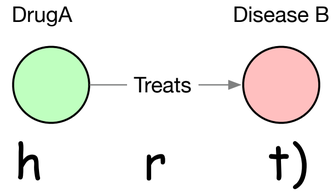


KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

What is Knowledge Graph?

- Specific type **directed** graph whose nodes and edges represent *entities* and *relationship* between entities
  - *The relationships organized in form of (head, relation, tail) -> (h, r, t) - TRIPLETs*



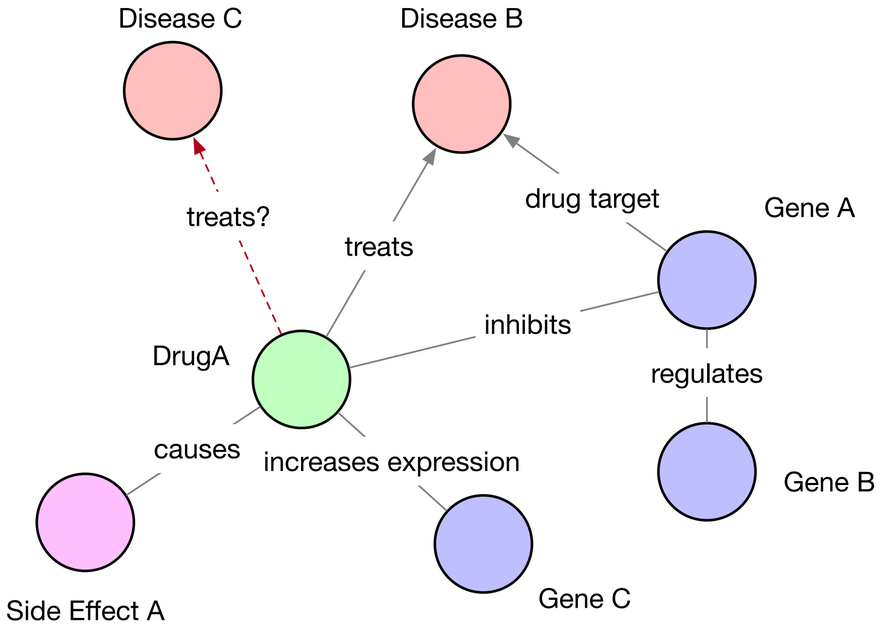


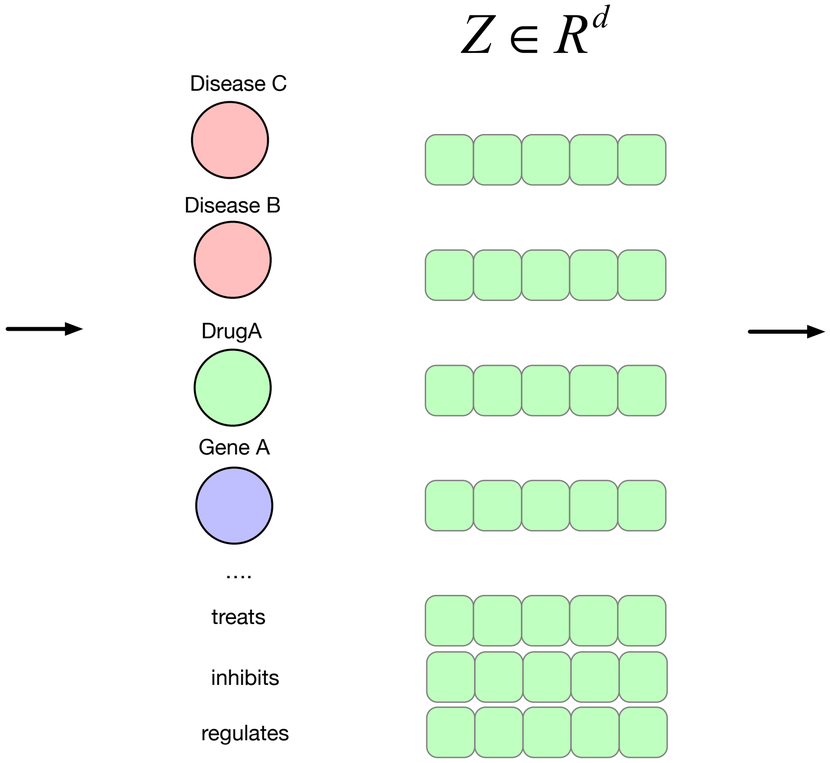
- Entity prediction:
  - (?, r, t)
- Relation prediction:
  - (h, ?, t) (e.g: DrugA-?-DiseaseB)
- Triplet classification
  - (h, r, t) is correct or not

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

KG Embedding

projections of entities and relations into a continuous low-dimensional space.





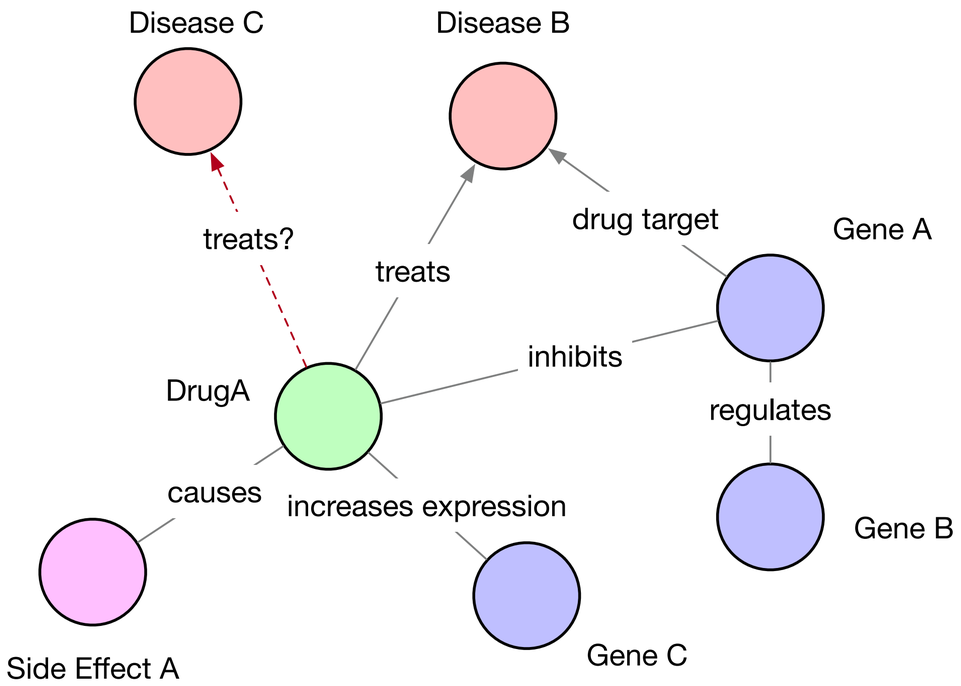
$$Z \in R^d$$

- Downstream Task
  - Entity Prediction
  - Relation Prediction
  - Triplet Classification

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

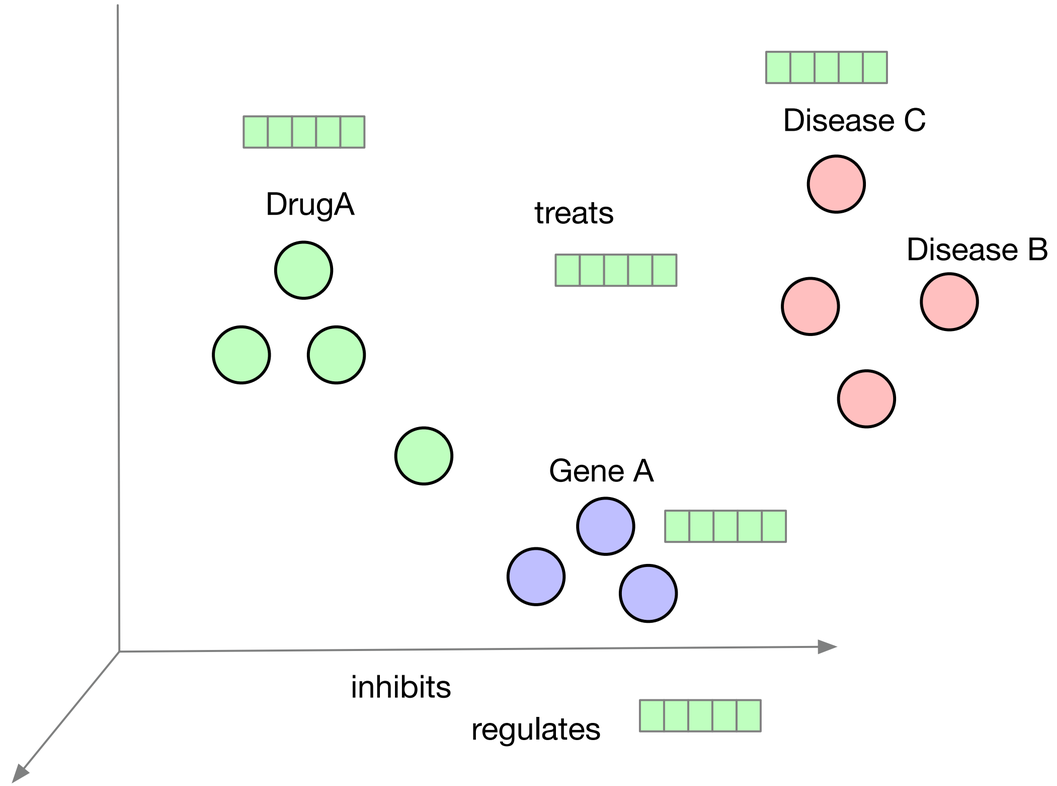
KG Embedding

projections of entities and relations into a continuous low-dimensional space.



f

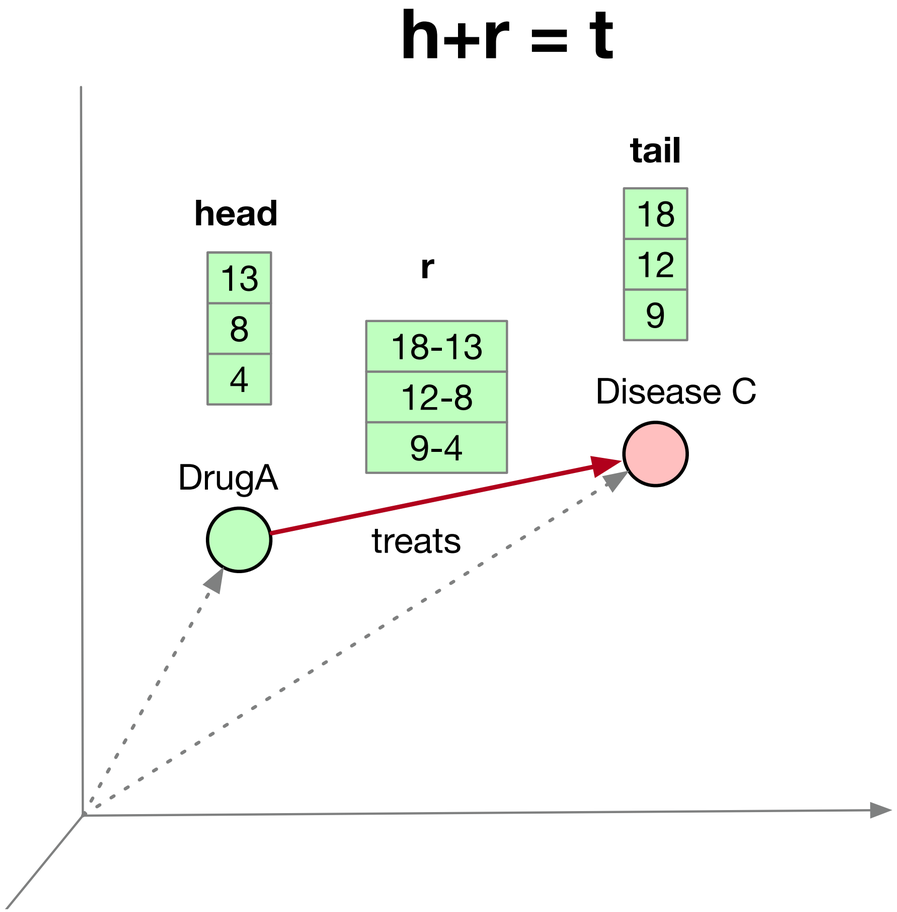
Scoring function



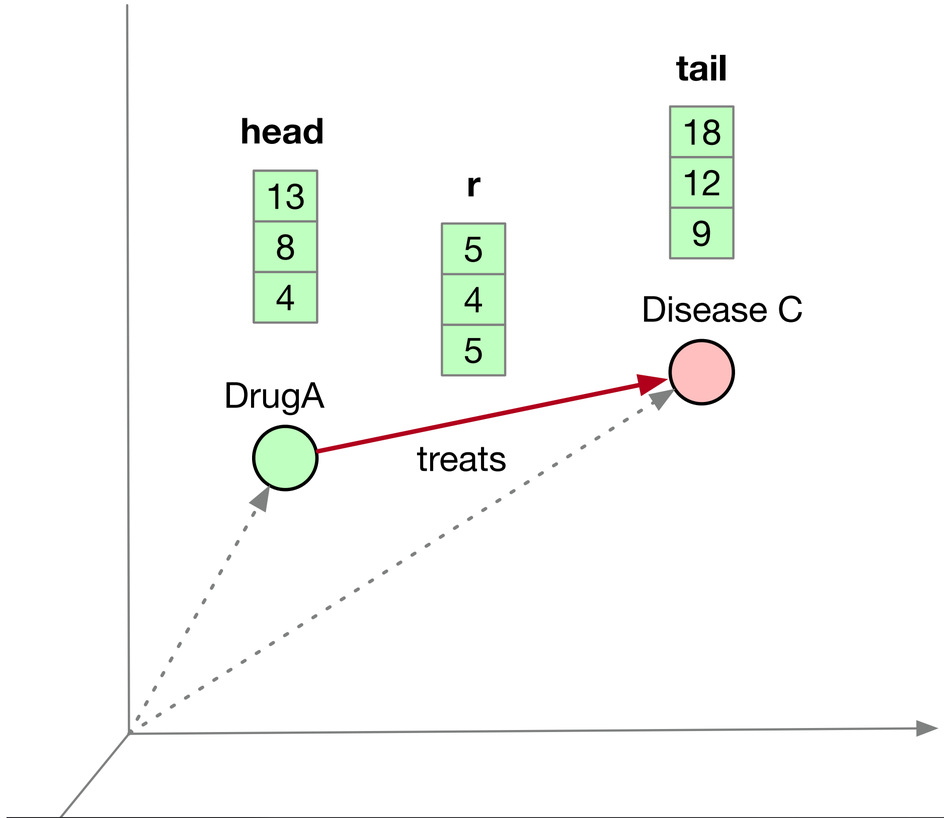
Semantically Meaningful Vector Representations

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Scoring Rule



$$\mathcal{L} = \sum_{(h,r,t \in S)} \sum_{(h',r',t')} [\gamma + d(h+r,t] - d(h'+r',t')]$$




```
~/Projects on ☁️erkhembayar@gmail.com
› DGLBACKEND=pytorch dglke_train --dataset DRKG --data_path ./train \
                                 --data_files drkg_train.tsv drkg_valid.tsv drkg_test.tsv \
                                 --format 'raw_udd_hrt' --model_name TransE_l2 --batch_size 2048 \
                                 --neg_sample_size 256 --hidden_dim 400 --gamma 12.0 --lr 0.1 \
                                 --max_step 100000 --log_interval 1000 --batch_size_eval 16 -adv \
                                 --regularization_coef 1.00E-07 --test --num_thread 1 \
                                 --gpu 0 1 2 3 4 5 6 7 --num_proc 8 \
                                 --neg_sample_size_eval 10000 --async_update
```

Embedding Methods

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Relation Properties

Symmetry `<Alice marriedTo Bob>`

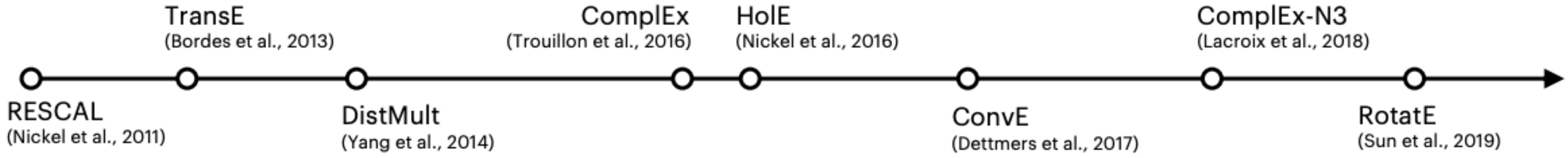
Asymmetry `<Alice childOf Jack>`

Inversion
`<Alice childOf Jack>`
`<Jack fatherOf Alice>`

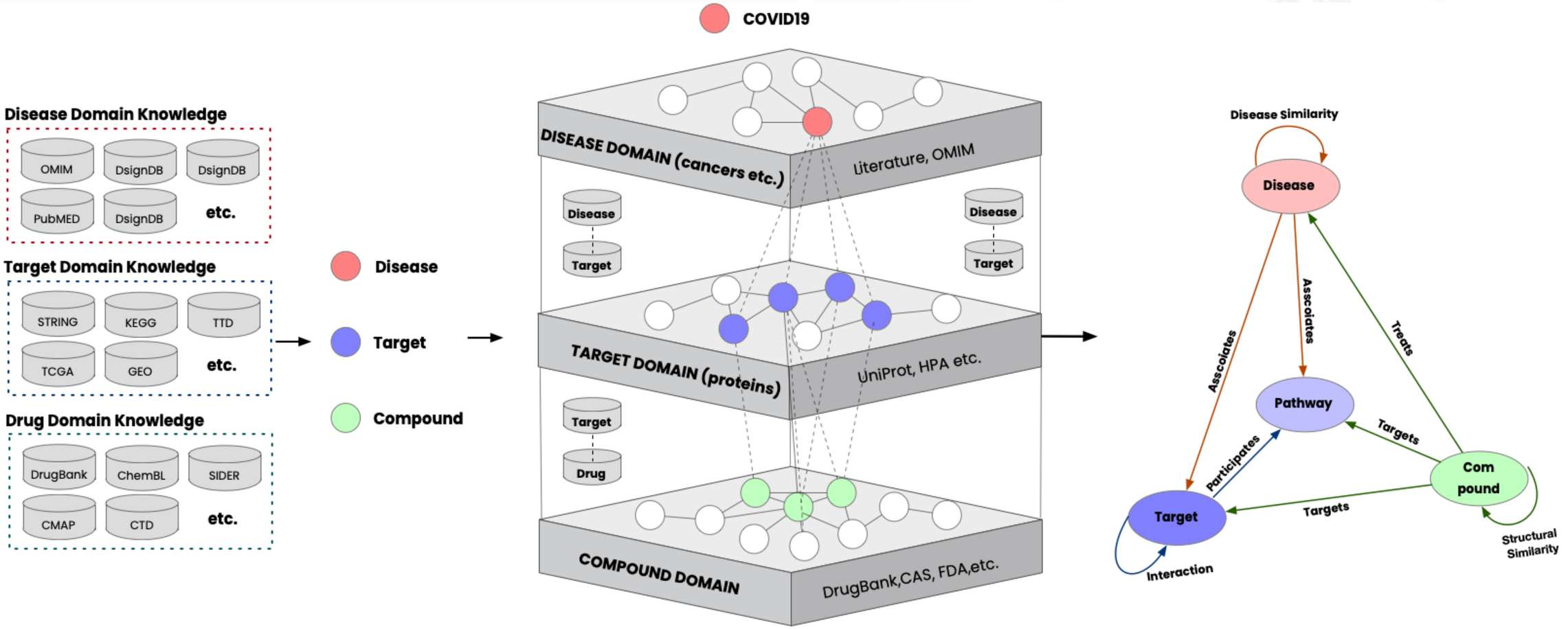
Composition
`<Alice childOf Jack>`
`<Jack siblingOf Mary>`
`<Alice nieceOf Mary>`

Model	Symmetry	Antisymmetry	Inversion	Composition
SE	✗	✗	✗	✗
TransE	✗	✓	✓	✓
TransX	✓	✓	✗	✗
DistMult	✓	✗	✗	✗
ComplEx	✓	✓	✓	✗
RotatE	✓	✓	✓	✓

Models	score function $f(\mathbf{h}, \mathbf{r}, \mathbf{t})$
TransE [2]	$-\|\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|\|_{1/2}$
TransR [10]	$-\|\|M_r\mathbf{h} + \mathbf{r} - \mathbf{M}_r\mathbf{t}\|\|_2^2$
DistMult [20]	$\mathbf{h}^\top \mathrm{diag}(\mathbf{r})\mathbf{t}$
ComplEx [16]	$\mathrm{Real}(\mathbf{h}^\top \mathrm{diag}(\mathbf{r})\bar{\mathbf{t}})$
RESCAL [12]	$\mathbf{h}^\top \mathbf{M}_r \mathbf{t}$
RotatE [15]	$-\|\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|\|^2$



KAICD 발지원센터 and Development



HEAD	RELATION	TAIL
Target	⟷	Target
Disease	⟷	Disease
Drug	⟷	Drug

HEAD	RELATION	TAIL
Target	⟷	Disease
Target	⟷	Drug
Drug	⟷	Disease

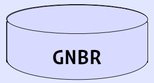
RELATIONSHIP TYPES
PPI, literature , coexpressed etc.
similarity, shares same target etc.
structural (Tanimoto), functional similarity etc.

RELATIONSHIP TYPES
target , association
bioactivity, binding, association
clinical, literatural, experimental

READING RELATIONSHIP TYPES (ONTOLOGY)

33.3GB

LITERATURE (TEXT)



Biomedical Relationships (PUBCHEM)

~ 24 million research articles

COMPOUND ⟷ DISEASE

COMPOUND-DISEASE

Compound-To-Disease →

(T) treatment/therapy (including investigatory)
(C) inhibits cell growth (esp. cancers)
(Sa) side effect/adverse event
(Pr) prevents, suppresses
(Pa) alleviates, reduces
(J) role in disease pathogenesis

Disease-To-Compound ←

(Mp) biomarkers (of disease progression)

Compound:1234 - GNBR::(T) Compound::Disease - Disease:MESH:DO1234

GENE ⟷ DISEASE

GENE-DISEASE

Gene-To-Disease →

(U) causal mutations
(Ud) mutations affecting disease course
(D) drug targets
(J) role in pathogenesis
(Te) possible therapeutic effect
(Y) polymorphisms alter risk
(G) promotes progression

Disease-To-Gene ←

(Md) biomarkers (diagnostic)
(X) overexpression in disease
(L) improper regulation linked to disease

Gene:1234 - GNBR::(D) Gene::Disease - Disease:MESH:DO1234

COMPOUND ⟷ GENE

COMPOUND-GENE

Compound-To-Gene →

(A+) agonism, activation
(A-) antagonism, blocking
(B) binding, ligand (esp. receptors)
(E+) increases expression/production
(E-) decreases expression/production
(E) affects expression/production (neutral)
(N) inhibits

Gene-To-Compound ←

(O) transport, channels
(K) metabolism, pharmacokinetics
(Z) enzyme activity

Compound:1234 - GNBR::(B) Compound::Gene - Gene:4321

GENE ⟷ GENE

GENE-GENE

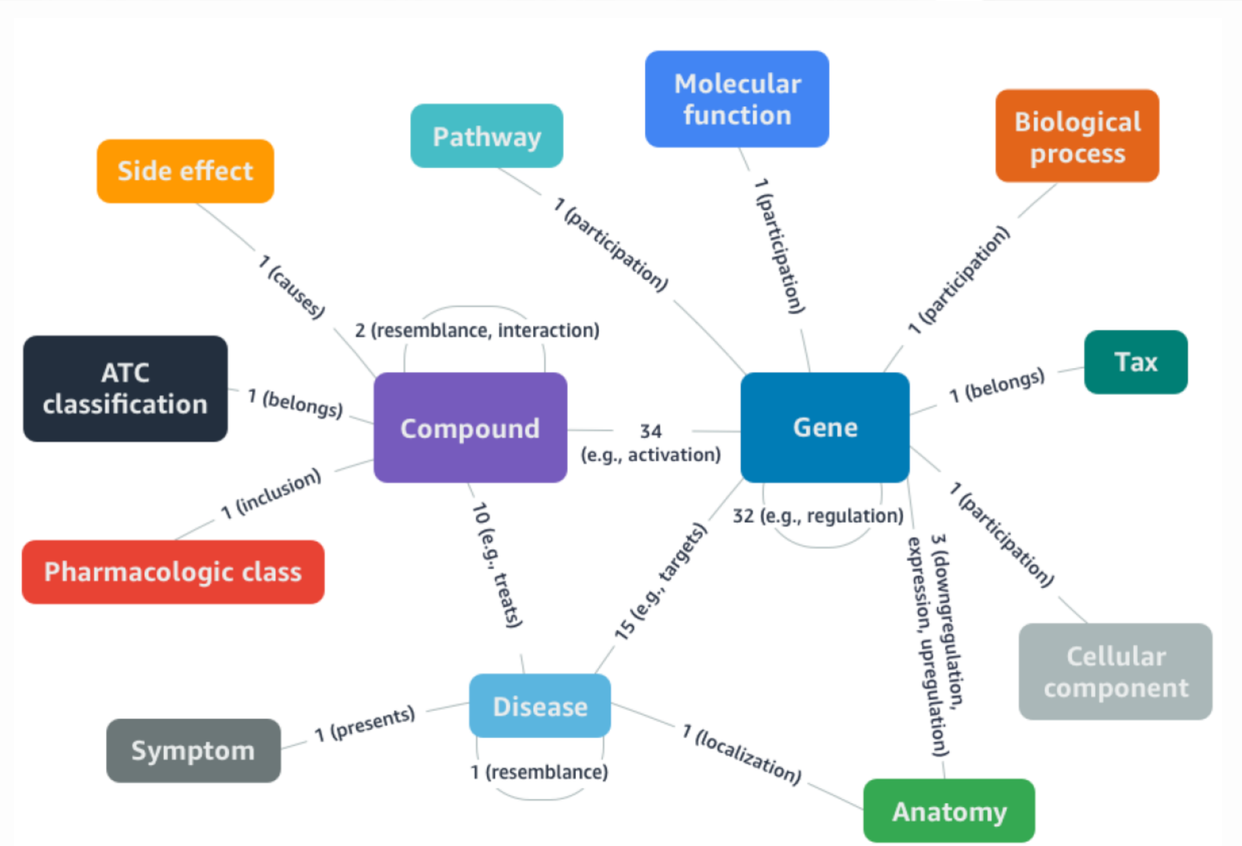
(B) binding, ligand (esp. receptors)
(W) enhances response
(V+) activates, stimulates
(E+) increases expression/production
(E) affects expression/production (neutral)
(I) signaling pathway
(H) same protein or complex
(Rg) regulation
(Q) production by cell population

Gene:1234 - GNBR::(Rg) Gene::Gene - Gene:4321

KAICD
한국인공지능신약개발지원센터
Korea AI Center for Drug Discovery and Development

Knowledge Graph Database

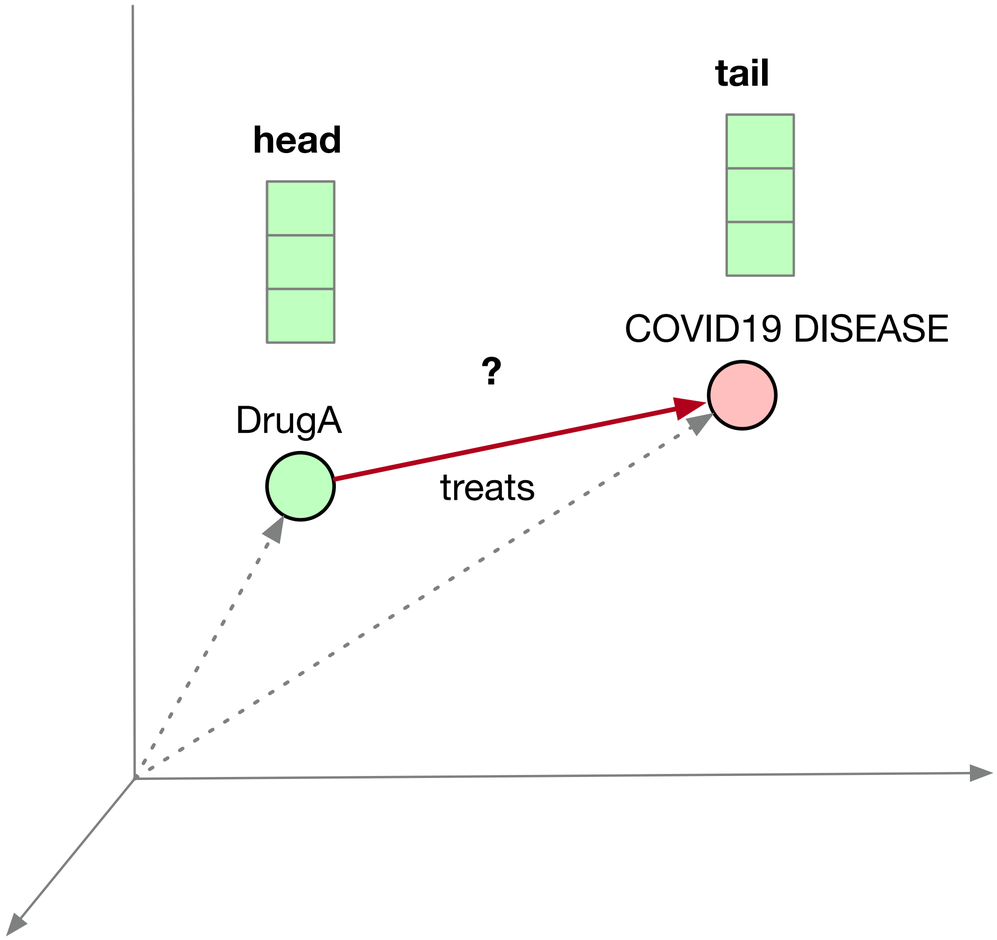
- Amazon Deep Engine Science Team
  - University of Minnesota, The Ohio State University and
  - **Drug Repurposing Knowledge Graph (DRKG)**



- 6 DBs
  - DrugBank, GNBR,Hetionet (Knowledge Database),STRING (PPI), IntAct (PPI), Bibliography
- 97,328 entities
  - 13 types of entities
- 5,874,461 triplets
  - 107 types of relations
- Embedded vectors

Knowledge-based DR for COVID19

Link Prediction:
[h, ?, t]



$$\mathbf{d} = \gamma - ||\mathbf{h} + \mathbf{r} - \mathbf{t}||_2$$

$$\mathbf{score} = \log\left(\frac{1}{1+\exp(-\mathbf{d})}\right)$$

Drugs will be repositioned (ranked) according to this score

Summary

- **Graph natural** representive power which helps to find patterns in data that scientist might not see
- The solution to many applications in biomedical field can be formulated as **GRAPH LEARNING PROBLEMs**
- **Graphs** are used to
  - Detect new target by taking advantages of complenterieness in variety of biomedical resources
  - Represent chemical compounds and showed promising performance
- **KNOWING GRAPH REPRESENTATION** is **IMPORTANT in DRUG DISCOVERY**

LET`S DO SOME CODING

Thank You