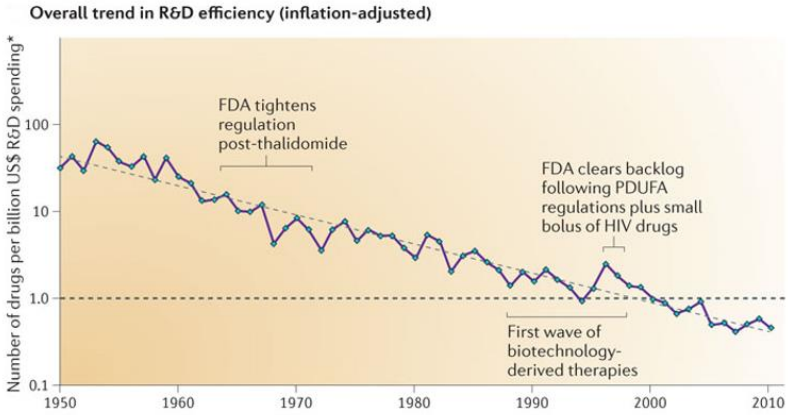
Application of Deep Learning to Drug Discovery

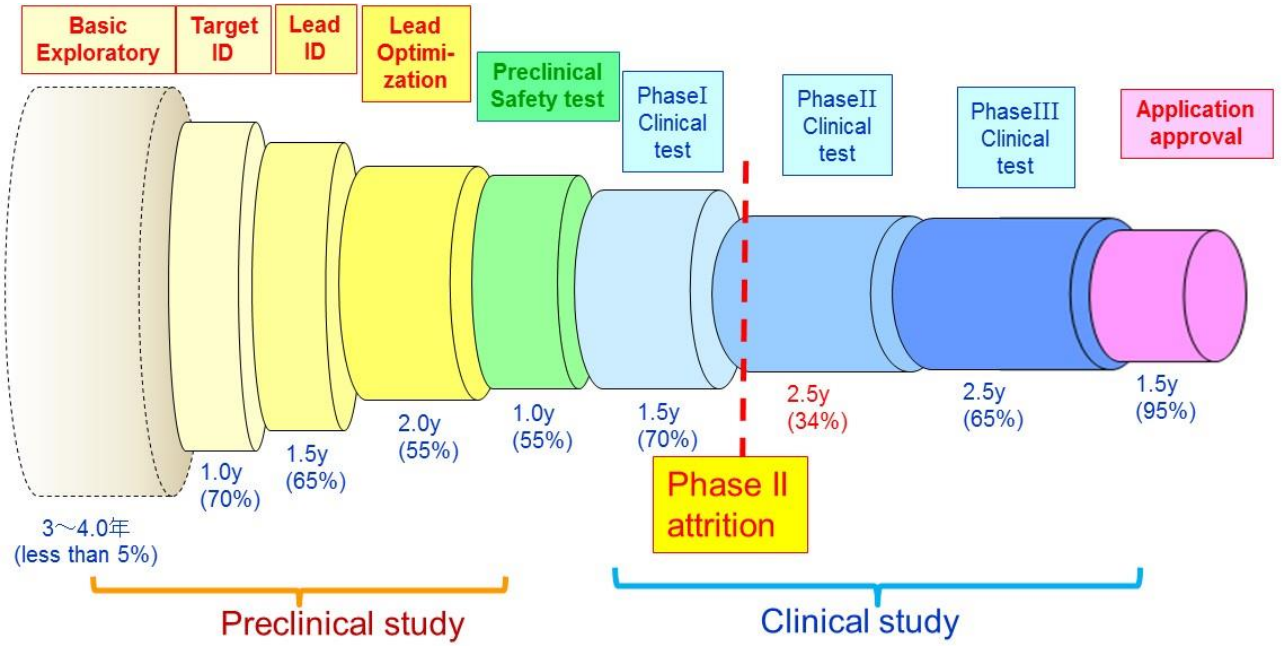
Hiroshi Tanaka Tohoku Medical Megabank Organization,
Tohoku University

Current Situation of Drug Discovery

- Rapid increase of R&D expenditure
  - More than 1B $ for one marketed drug
- Decrease of success rate
  now about 1/20,000～1/30,000
  - Remarkable Drop Between non-clinical and clinical test **(phase II attrition)**
- **Clinical Predictability**
  - At as early as possible stage,
    Estimation of clinical efficacy and toxicity
- **Efficient measures**
  - **Use Disease-specific iPS cell**
  - **Use of Human Bio Big Data in early stage**



Nature Reviews Drug Discovery (2012)



Biomolecular Profiling DrugDiscovery/DR

TMDU

Relation among Drug, Disease and Target

Main causes of Disease

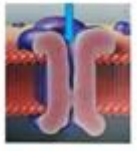
Disease-related molecules

Drug: acting through Drug **Target** to exert influences on **Disease-related genes**

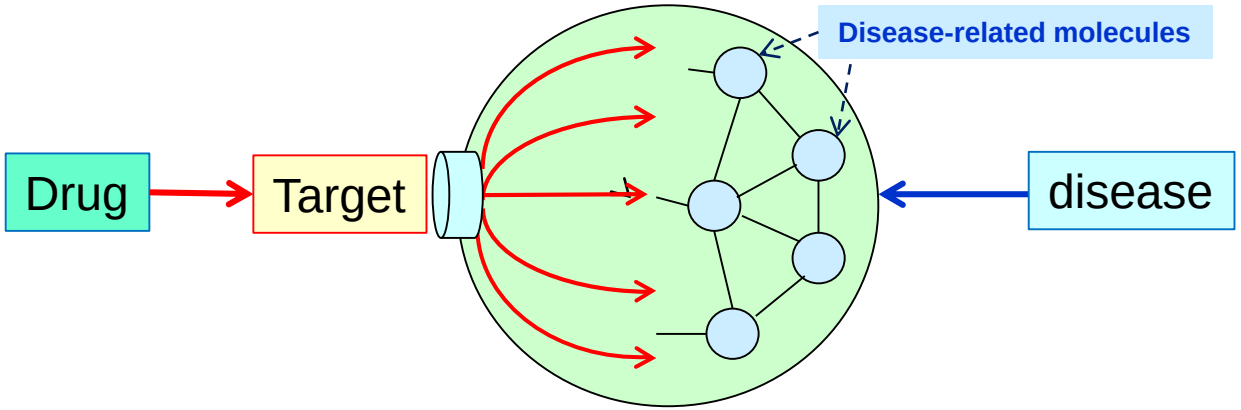
Target Molecule of Drug
Receptor, Enzyme, Channel

receptor

enzyme



channel



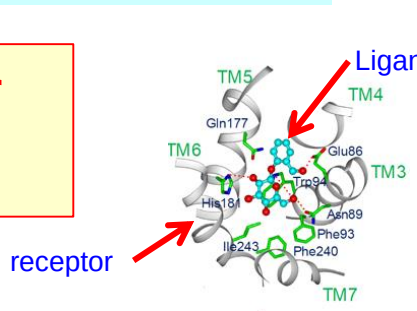
Human biological system(network)

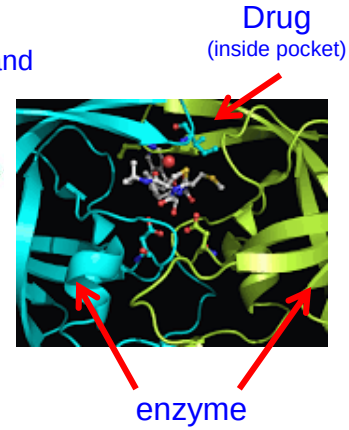
TMDU

Biomolecular profiling DrugDiscovery

New approach to computational drug discovery

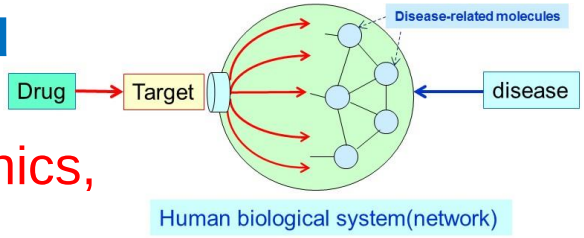
Computational drug discovery so far
in silico drug discovery





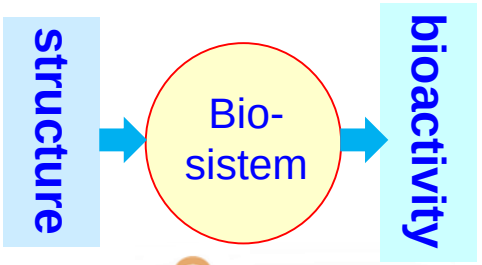
Molecule Centric

- **Structure-based rational drug design**
- **Computational Molecular Docking Design** between
  Target M (receptor, enzyme, channel) **and**
  Drug (ligand)
- Molecular Orbital Method, Molecular Mechanics, Molecular Dynamics (MD method)
  ex. influenza drug; zanamivir (relenza)
- **Optimization of lead compounds**



Quantitative structure-activity relationship

- QSAR: bioactivity and molecular structure
- Between **drug** and **response, biosystem exists**



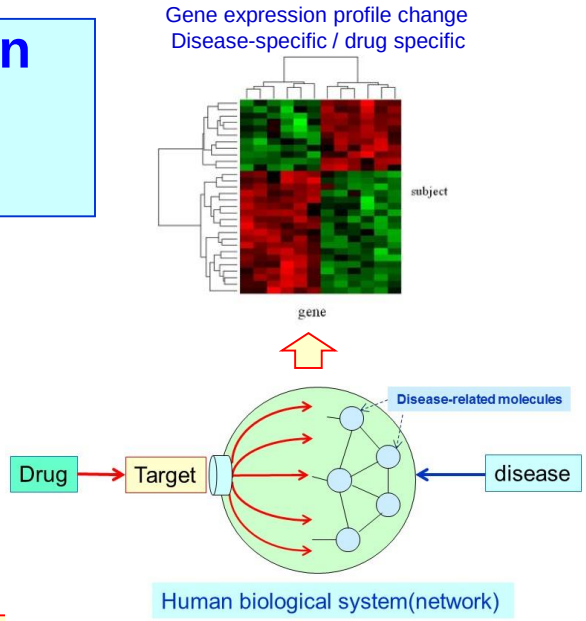
TMDU

Biomolecular profiling DrugDiscovery/DR

New computational Drug Discovery/DR（Biomolecular Profiling）

Gene Expression Profile at Disease Contraction
Induced by Disease-related gene's activity
Disease-specific gene expression profile change

Gene Expression Profile at Drug Prescription
Genome-wide change of gene expression
Induced by junction between drug and target
Drug-specific gene expression profile change



Comphrehensive molecular profile change
⇒molecular network change of total biosystem

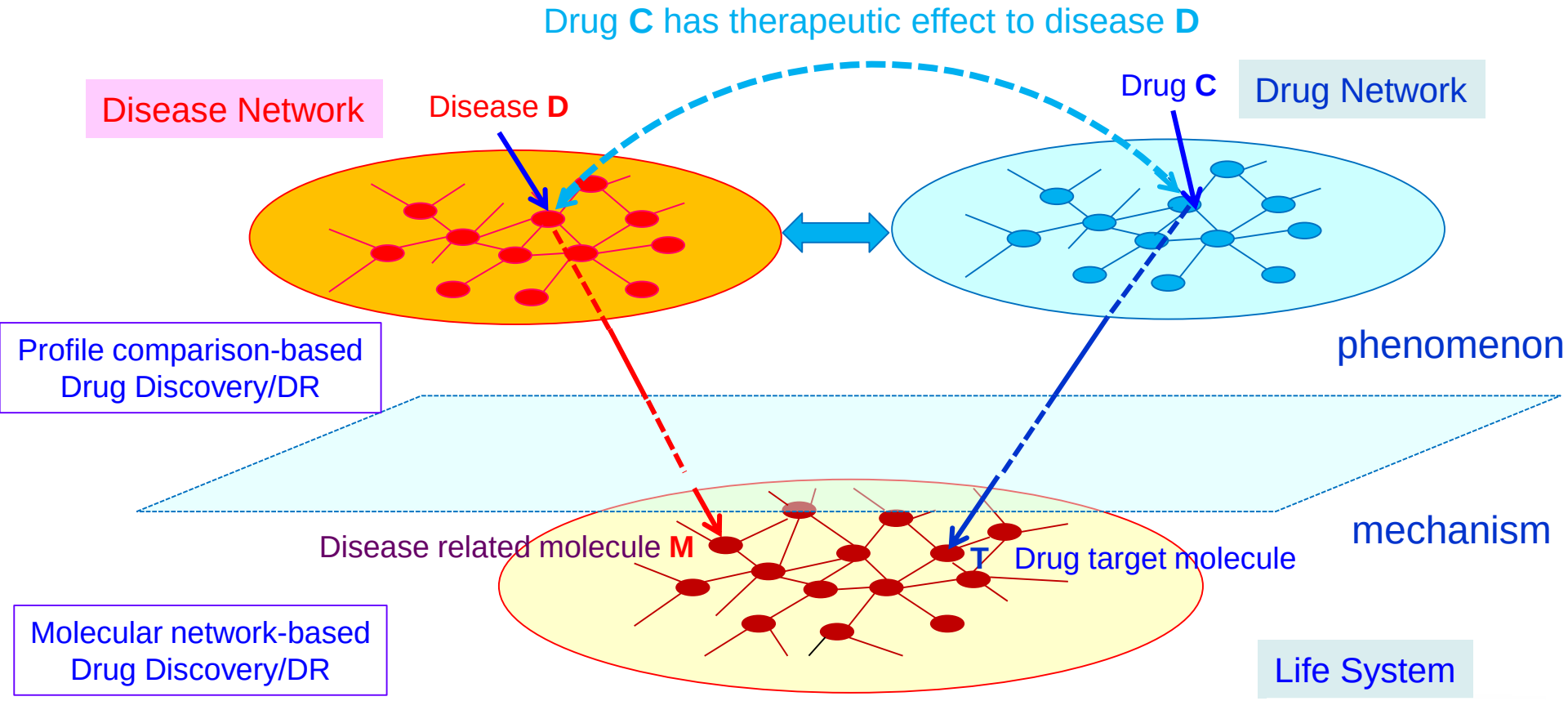
＜drug-target molecule docking＞exerts an influence to＜diseased state＞
From the genome-wide viewpoint of total biosystem

Utilize big data DB connecting compounds, target molecules, disease

TMDU

Basic structure of profile-based computational drug discovery

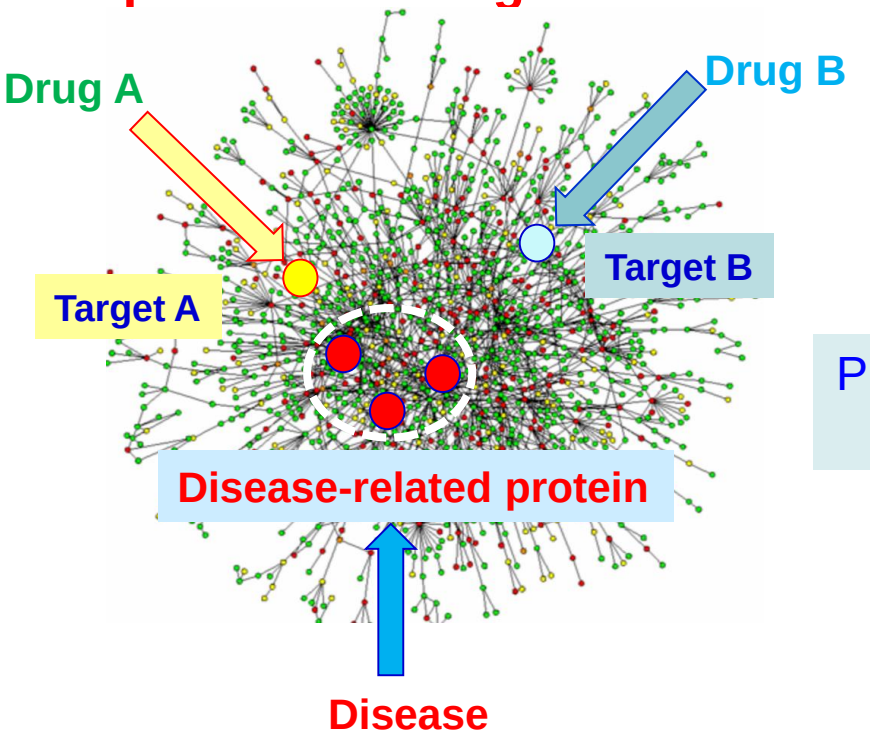
Framework of Triple-layer disease and drug network



DR: Drug Repositioning: is the application of known drugs (compounds) to treat new indications (i.e., new diseases)

Common Platform of DrugDiscovery/DR
Protein-Protein interaction network（PPIN）

- Common Platform bionetwork: mediating disease and drug action
- **Protein-protein interaction network（PPIN）** as common platform
- **Disease:** Scaffolding in PPIN: **Disease-related protein** (gene)
- **Drug** : Scaffolding in PPIN**: Drug Target protein**
- Based on **the distance (proximity)** between **Disease-related protein** and **target protein**,
  the impact of the drug is measured

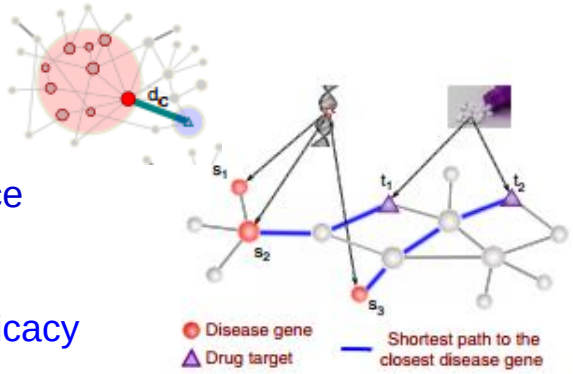


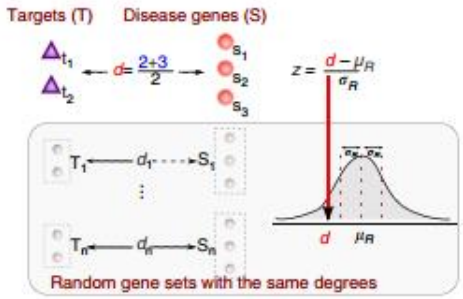
Protein-protein Interaction Network（PPIN）

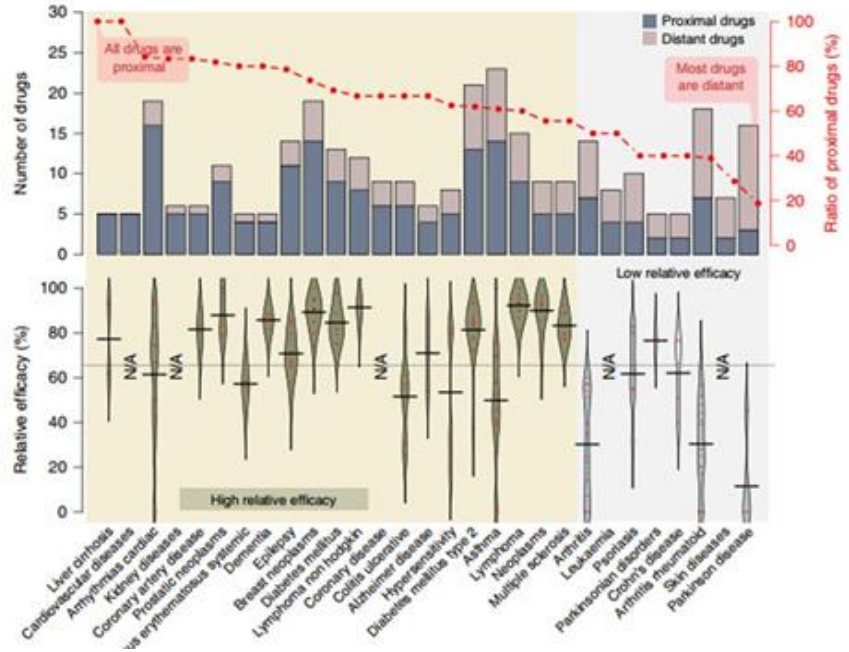
TMDU

Proximity between Drug and Disease at PPIN

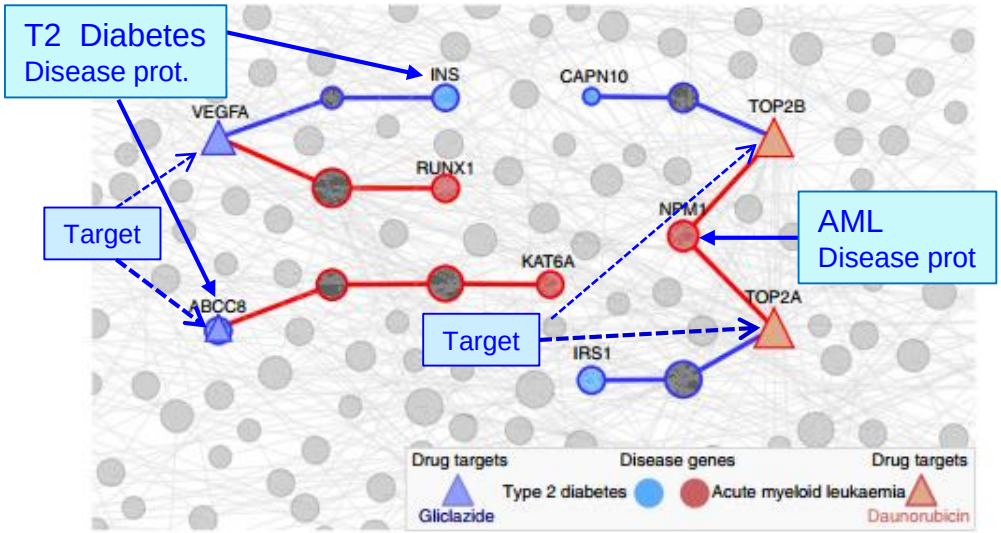
Relaltive Proximity Index d_c :
①Distance between Target and the nearest protein among disease-related protein module
②The distance is normalized among the distance of the molecules in same context
$z < -0.15 \Rightarrow$ proximate
③closest measure d_c : best index to measure efficacy

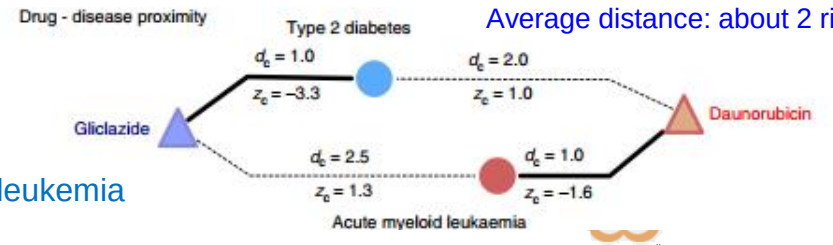






(Guney, Barabasi,2016, Nat. Com)





Average distance: about 2 rinks

AML: acute myeloid leukemia

Need for Learning

- We are **still missing in understanding** of the necessary conditions for molecule to be effective
- We should find these conditions by **learning from the succeeded <disease-drug-target molecule> combinations**
- **Artificial Intelligence** (AI), specially **Deep Learning** is now the most powerful method

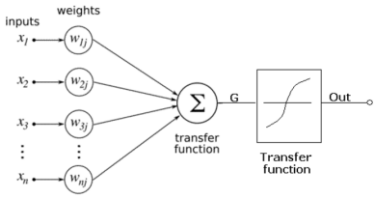
TMDU

Artificial Intelligence based DrugDiscovery/DR

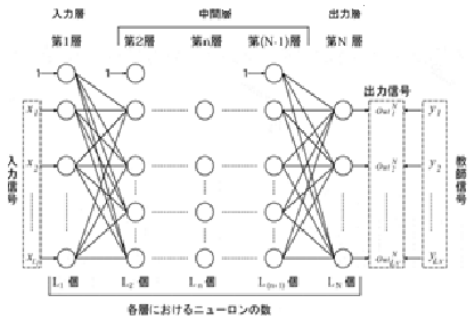
TMDU

Artificial Intelligence revolution by Deep Learning

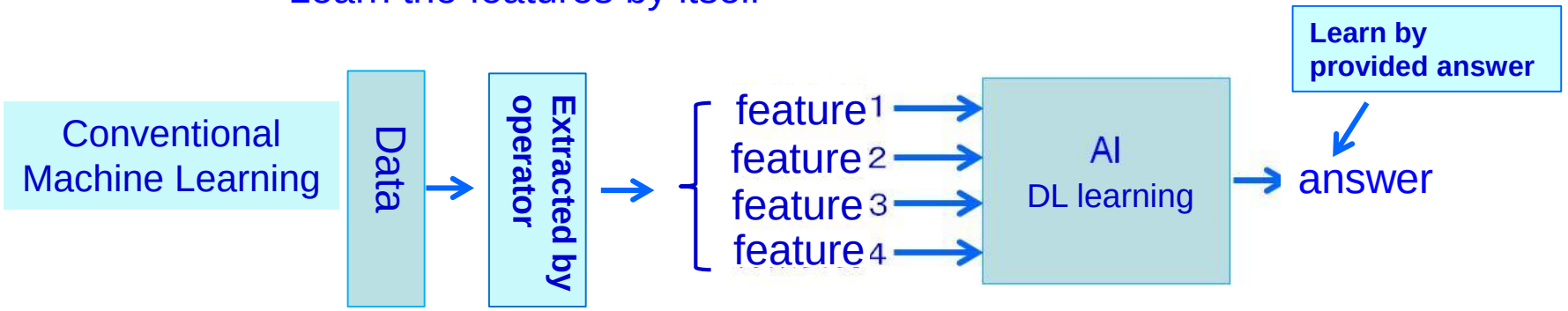
- Limitation of Machine Learning
  - "Supervised learning"
    - Construct AI by providing the feature and answer
- Deep Learning revolution
  - "Unsupervised learning"
    - Learn the features by itself

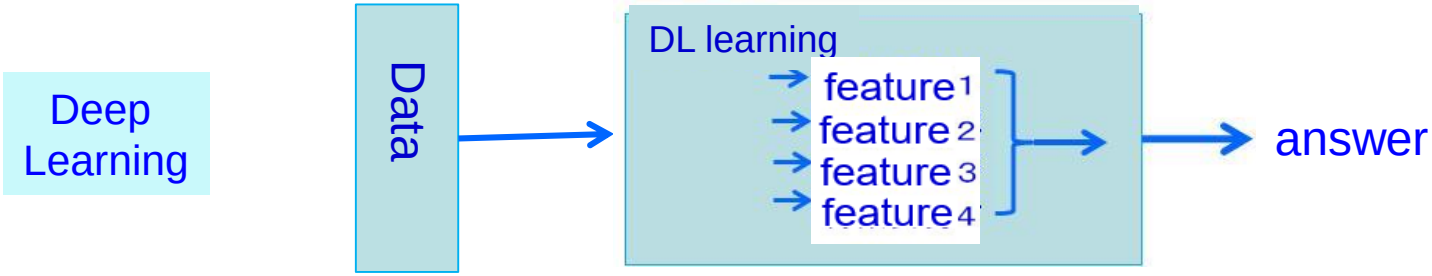


Neural information element



Multiple Layer neuro-network

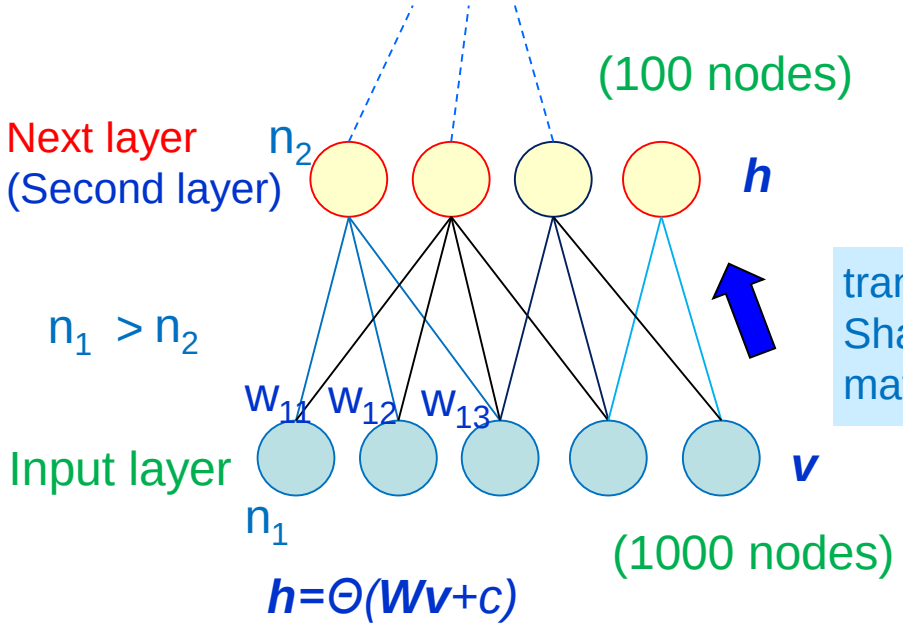




Revolutionary point of DL Autoencoder 1

- Principle of **autoencoder**: Learn specific intrinsic features of the big data
- Restore the node values of input layer from the node values of next layer where the number of nodes is decreasing compared with input layer.
  → Intrinsic features should be explored so that the input layer to be recovered as same as possible

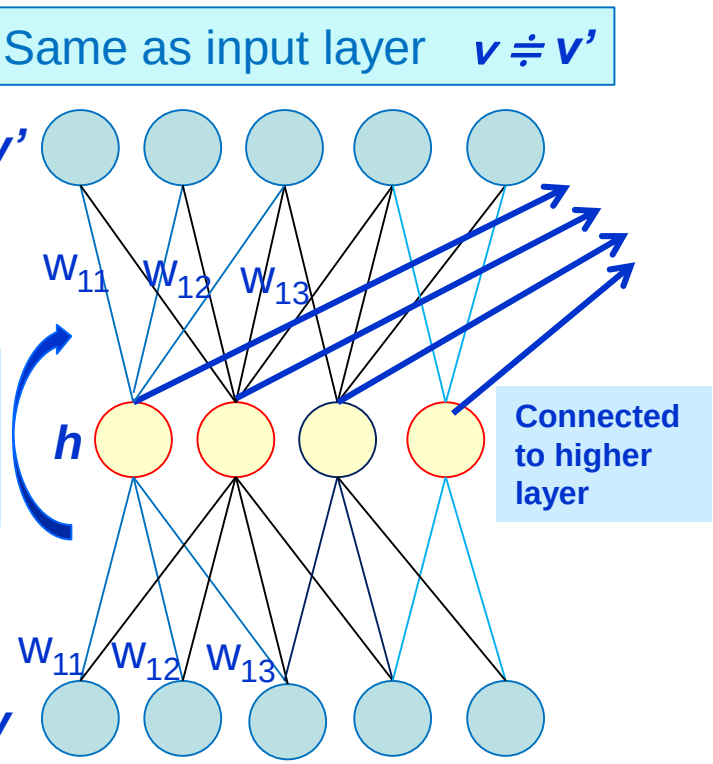
→ discover intrinsic features



$v'=\Theta(W^t h+c)$

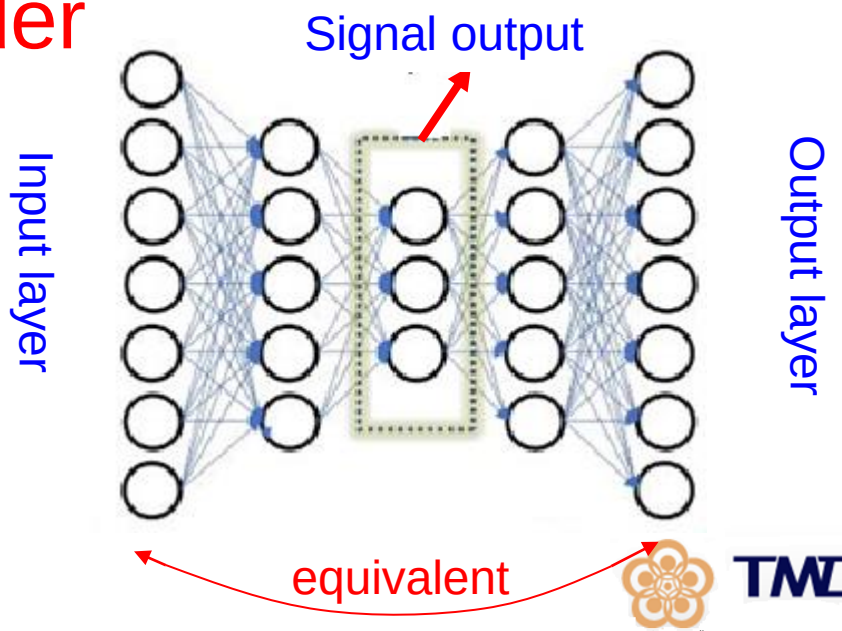
$W'=W^t$

transposed matrix
Share the weight matrix



Revolutionary point of DL Autoencoder 2

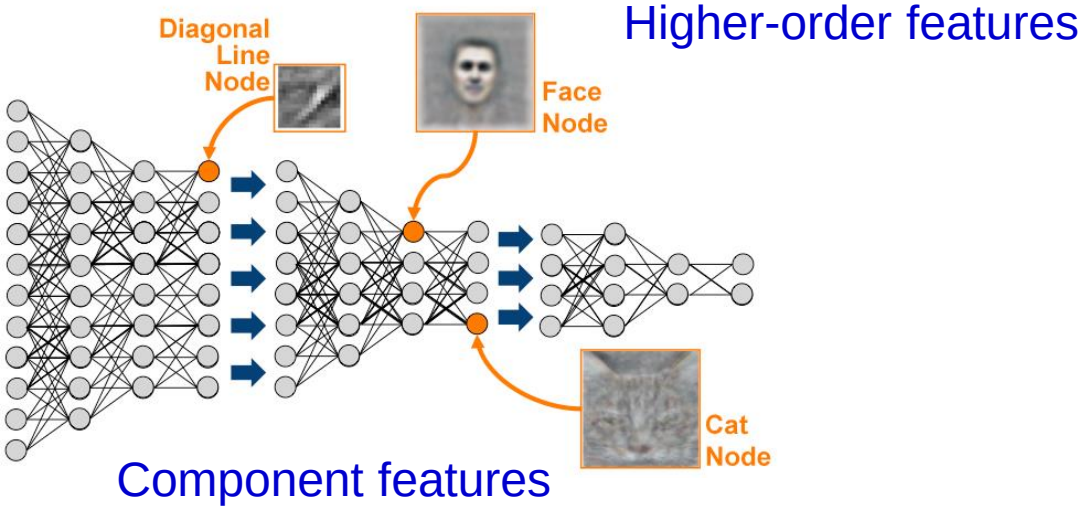
- Consisting of multiple layers of autoencoders
  - stacked autoencoder
- Consisting of input layers(encoder) and output layers(decorder) which are contrasted
  - deep autoencoder



TMDU

Revolutionary point of DL Autoencoder 3

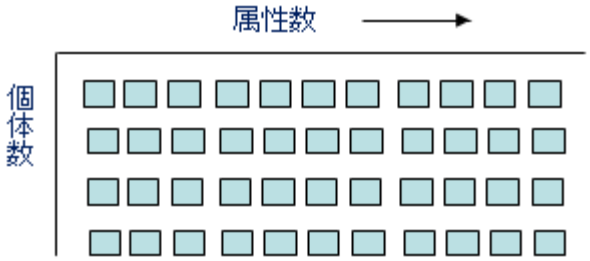
- Any number of layers can be used to construct the DL neuronetork by applying autoencoder to each layer → **“autoencoder” for each layer**
  （stacked autoencoder ）
- Since the next layer is created using feature quantities learned in the previous layer, higher-order features are created through the hierarchy
- "Supervised learning" is finally necessary to combine feature expressions and human concepts/intention
- Overcome the limits of previous learning methods by automatic feature extraction
  → Structural understanding by intrinsic features
  → **e.g. AlphaGo which exceeed human ability**



Date principle of “Big Data”

Problem : number of attribute value (p)≫ number of sample (n)

p: may be several hundred million　n: At most tens of thousands, normally thousands



If all these huge nuber of attribute variables are independent, structural analysis of big data is impossible.

Big Data・Sparse theory

Big data is composed of a large number of independent components but less than the number of attribute values as its basis.

Principle of compositionality

TMDU

Contracting method for multidimensional network by Deep Learning

- Application to big data of medicine and drug discovery
- Increase of “multidimensional network information structure”
  - Genome medicine : <comprehensive molecular information – clinical phenotype information>
  - Genome cohort : <gene – environment information> (lifestyle)
- Deep Learning-based Network Contraction

Multidimensional network information structure
⇒ Contact to be composed of a few network bases

- Not linear decomposition.
- Project to be composed of intrinsic bases by nonlinear decomposition.

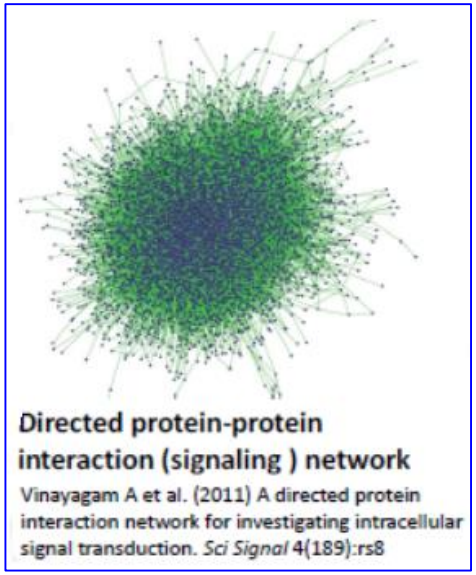
TMDU

Decomposition to intrinsic network basis

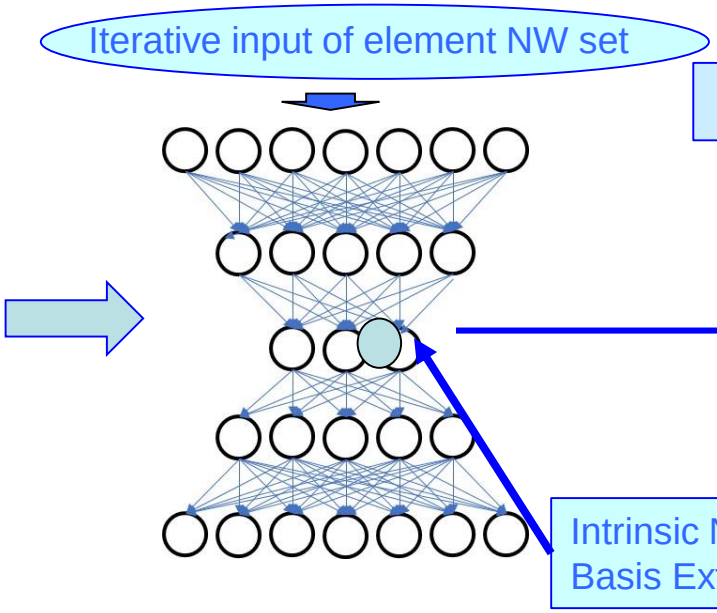
Contracting to sum of intrinsic network bases

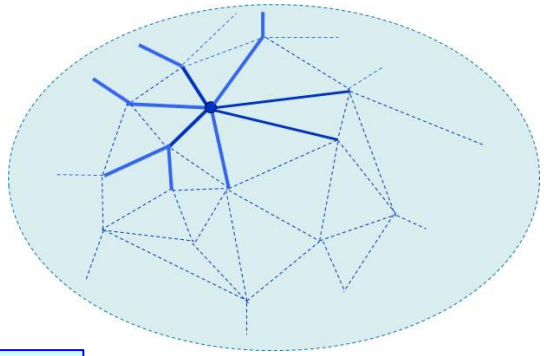
DL is applied to iterative presentation of element NW, where all network links are viewed as a sum of element NW (partial NW)

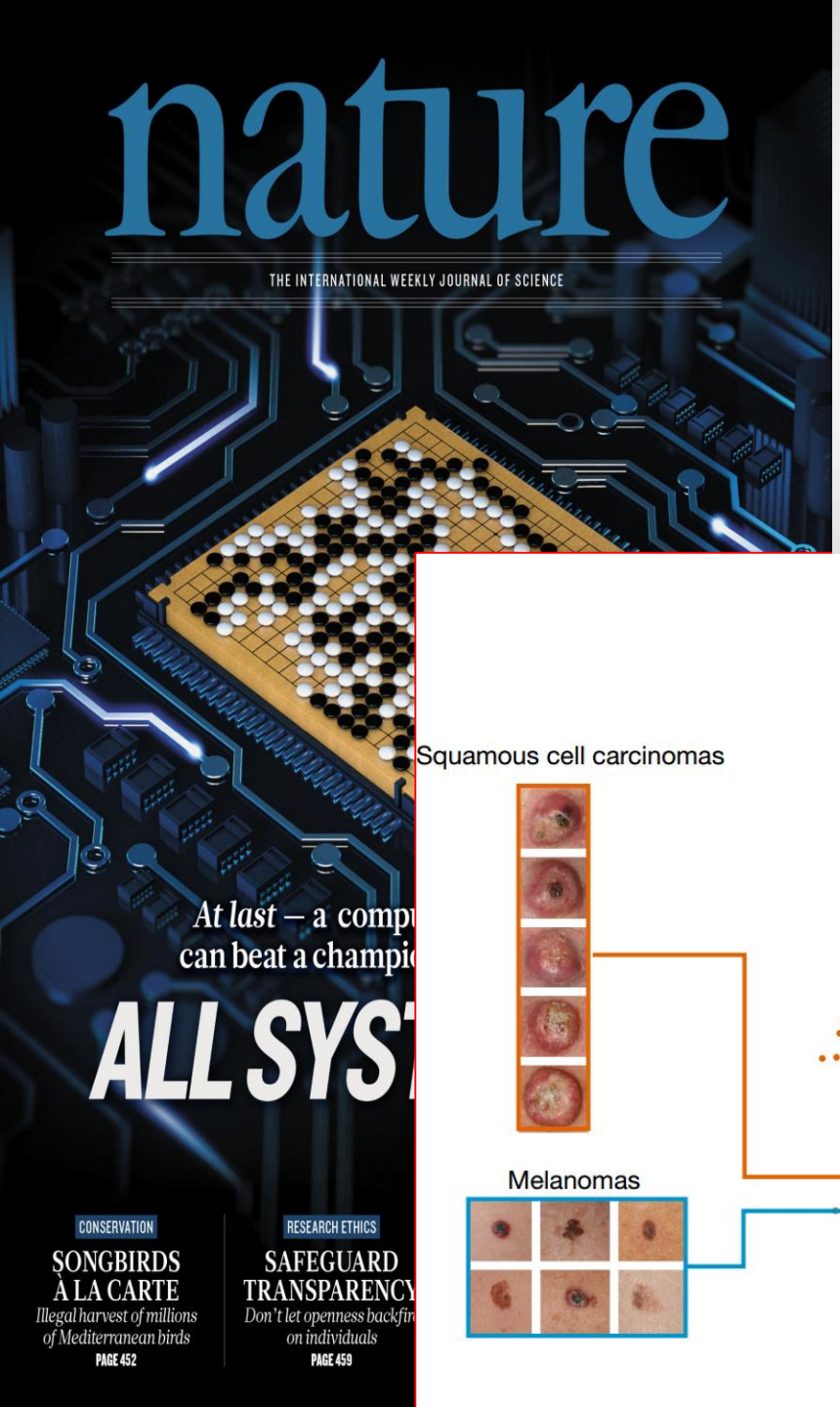
Sum of intrinsic network basis that reproduces all networks

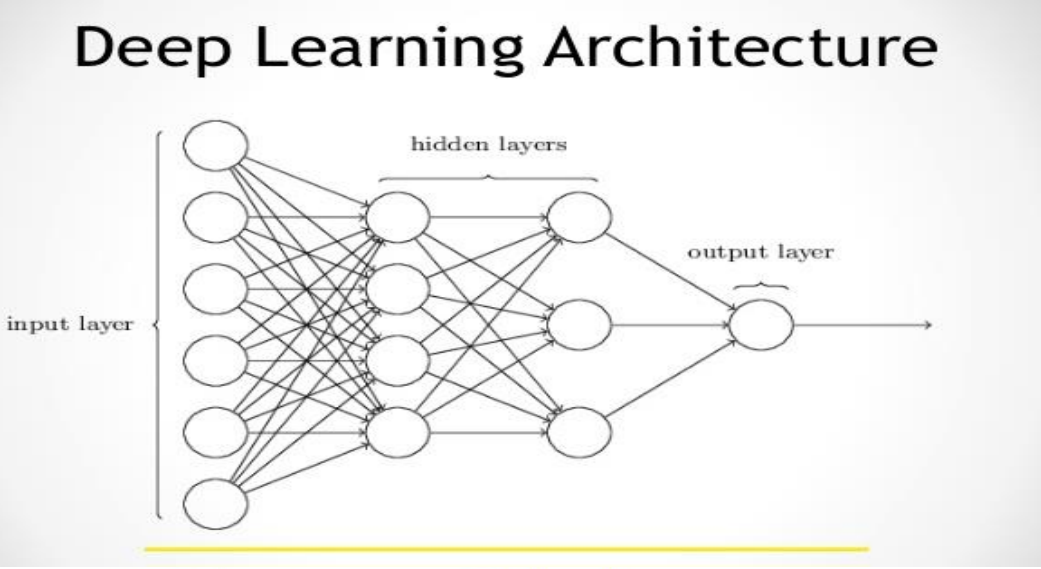


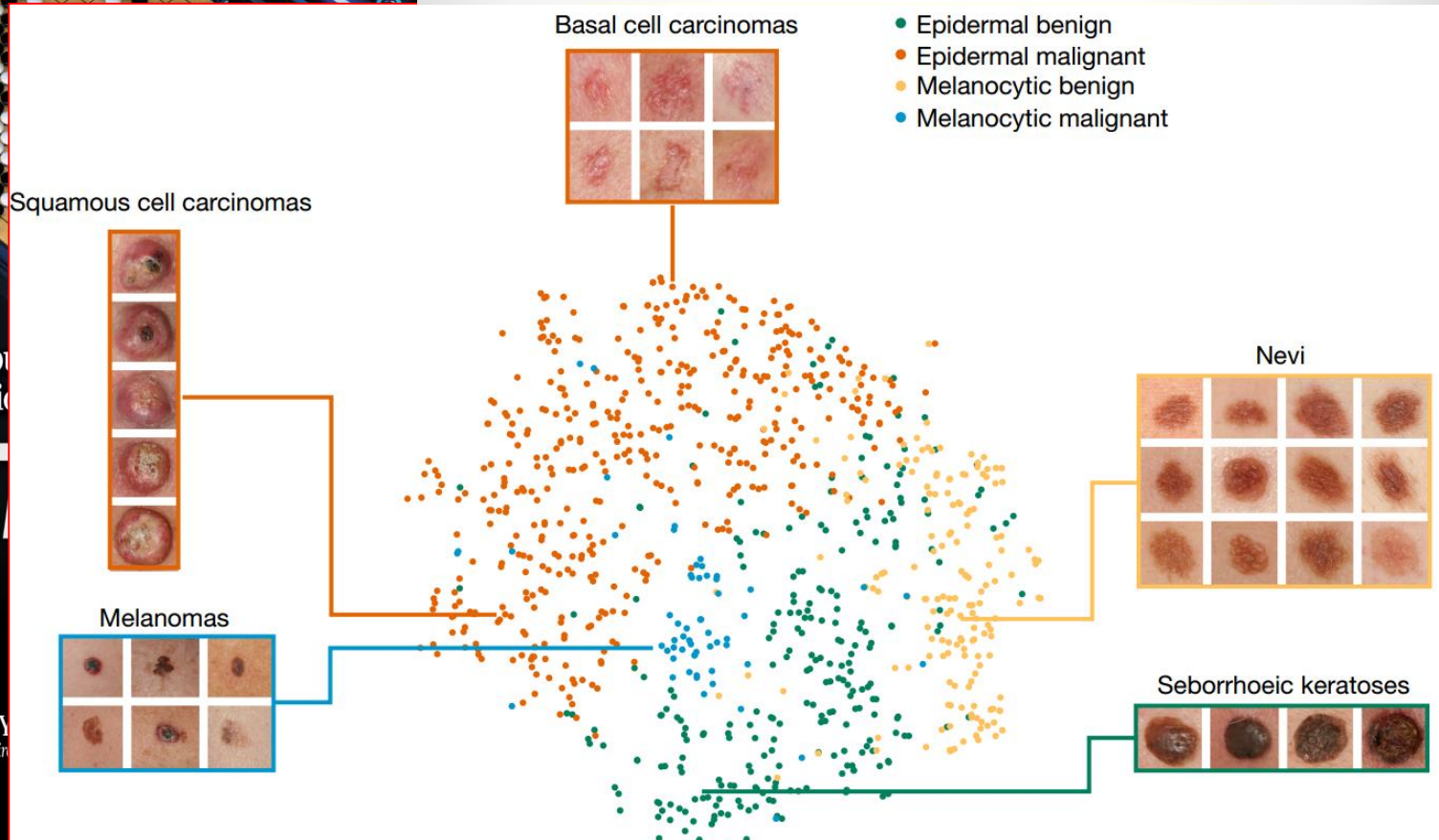
PPI Network






nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last – a compu
can beat a champi
ALL SYST
CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfir on individuals
PAGE 459

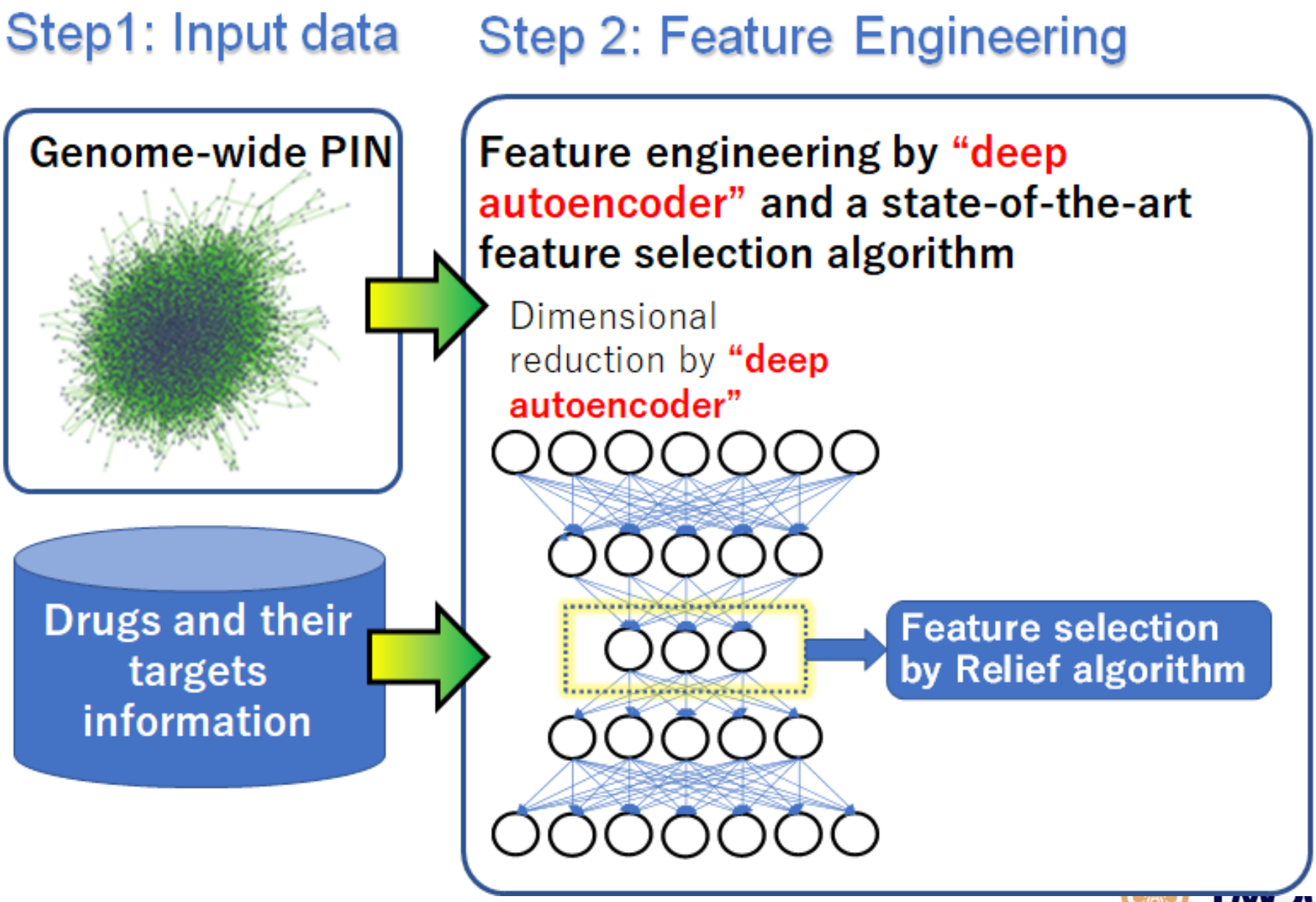

Deep Learning Architecture
hidden layers
input layer
output layer


Basal cell carcinomas
Epidermal benign
Epidermal malignant
Melanocytic benign
Melanocytic malignant
Squamous cell carcinomas
Nevi
Melanomas
Seborrhoeic keratoses

Our Approach

- By using deep learning and genome-wide protein interaction network,
- We build a computational framework to predict potential Drug Target genes and
- Repositionable drugs for Alzheimer's disease.

Our computational workflow

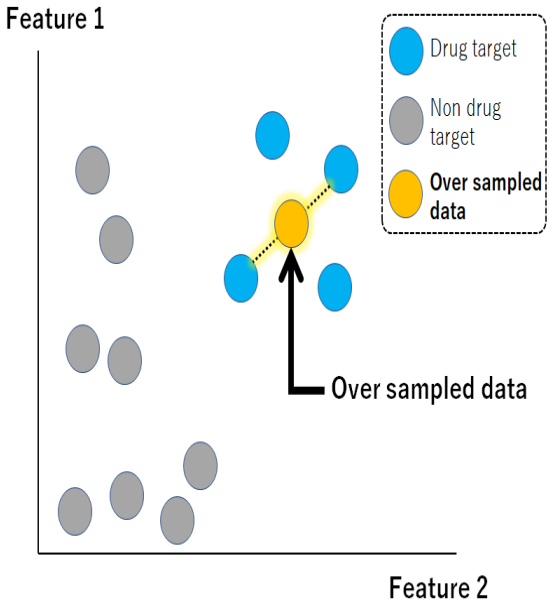


Step 3: Classifier model

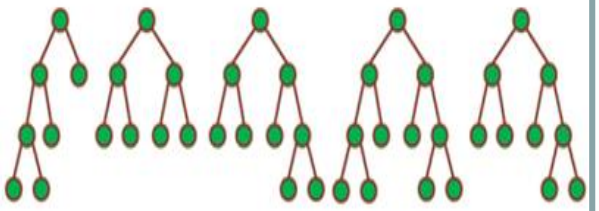
Step 4: Target prioritization

A binary classifier model to target prioritization by state-of-the-art machine learning algorithms

SMOTE algorithm to build a training data



Xgboost algorithm to build a binary classfier



Scores for potential targets

Gene	Score (mean probability)
GRASP	0.982971499
PGRMC1	0.98234516
GPM6A	0.98234516
NRP2	0.975193546
PFKM	0.972127568
DLGAP2	0.953659343
CD81	0.941095327
IQGAP1	0.926867425
TROVE2	0.916886333
TOP3B	0.915745595
TJP1	0.914564961
PDGFB	0.914082375
SETD2	0.905462331
CFLAR	0.900456515
PROS1	0.883435477
SIT1	0.879989294
SIGLEC7	0.879989294
SHC2	0.879989294

TMDU

GRASP	PIK3C2B	PKIA
PGRMC1	NEU3	PFKP
GPM6A	SLC25A38	PAN2
NRP2	TNFSF12	GLUD1
PFKM	ADRA1B	DNM3
DLGAP2	DPM2	ITGA5
CD81	NLRP12	RILPL2
IQGAP1	NLRC4	MAEA
TROVE2	UIMC1	NCDN
TOP3B	IL8	DGCR14
TJP1	VAV1	PACSIN3
PDGFB	ARHGEF1	CD46
SETD2	WISP2	NIT1
CFLAR	PRKCE	ICAM4
PROS1	TBXA2R	GNA13
SIT1	TSPAN4	STK40
SIGLEC7	EPHB4	ROGDI
SHC2	LOC63920	CDH10
SH2D1A	PSEN1	WSB2
	SPOCK3	PHPT1
	TSPO	
	SLC4A1	

By using the AI-based method, we successfully predict potential drug targets (more than 100 genes) for Alzheimer's disease.

TMDU

Example,

SLC25A38 (APPOPTOSIN)

SLC25A3 increases in the brain from Alzheimer's disease patients as well as from infarct patients. Further, SLC25A38 downregulation is likely to inhibit apoptosis induced by Bax/BH3I and neuronal death induced by Aβ/glutamate.

Previous Next

Featured Article | Articles, Cellular/Molecular

Appoptosin is a Novel Pro-Apoptotic Protein and Mediates Cell Death in Neurodegeneration

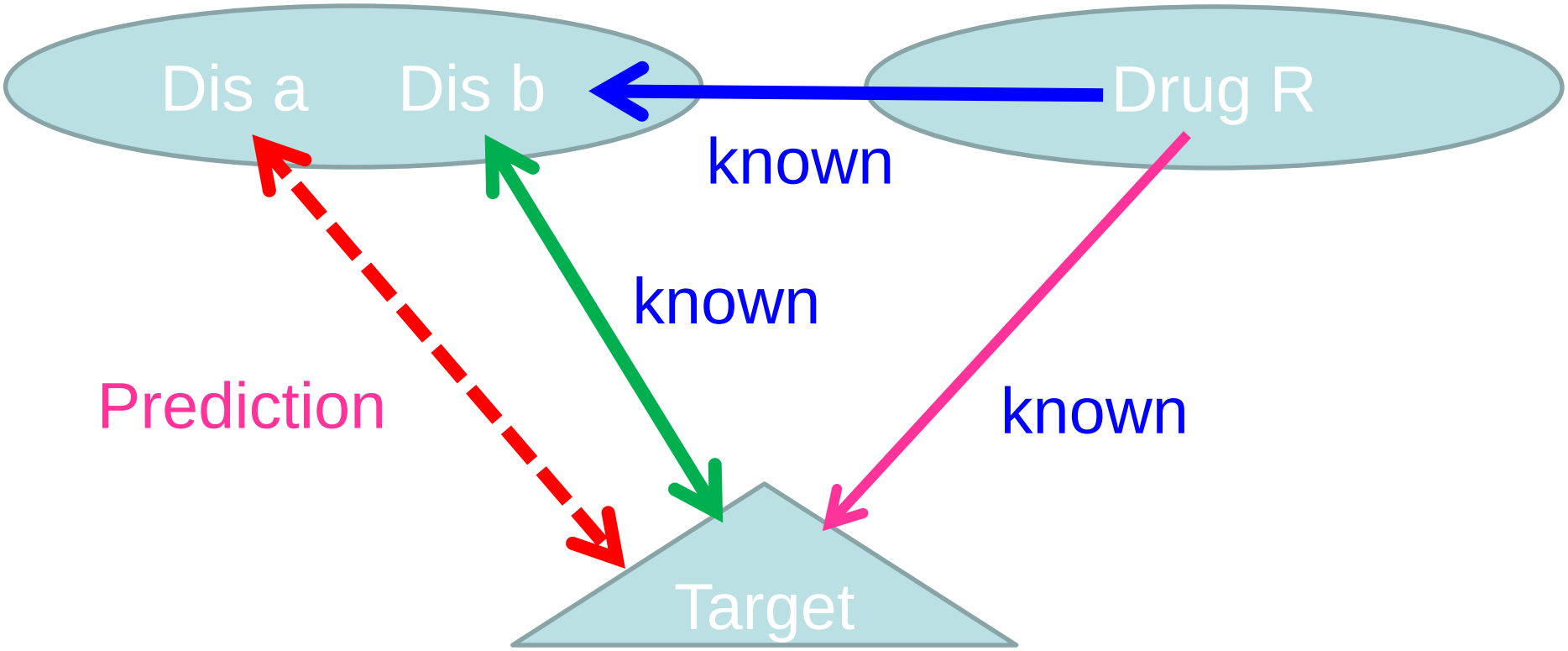
Han Zhang, Yun-wu Zhang, Yaomin Chen, Xiumei Huang, Fangfang Zhou, Weiwei Wang, Bo Xian, Xian Zhang, Eliezer Masliah, Quan Chen, Jing-Dong J. Han, Guojun Bu, John C. Reed, Francesca-Fang Liao, Ye-Guang Chen, and Huaxi Xu

Journal of Neuroscience 31 October 2012, 32 (44) 15565-15576; DOI: https://doi.org/10.1523/JNEUROSCI.3668-12.2012

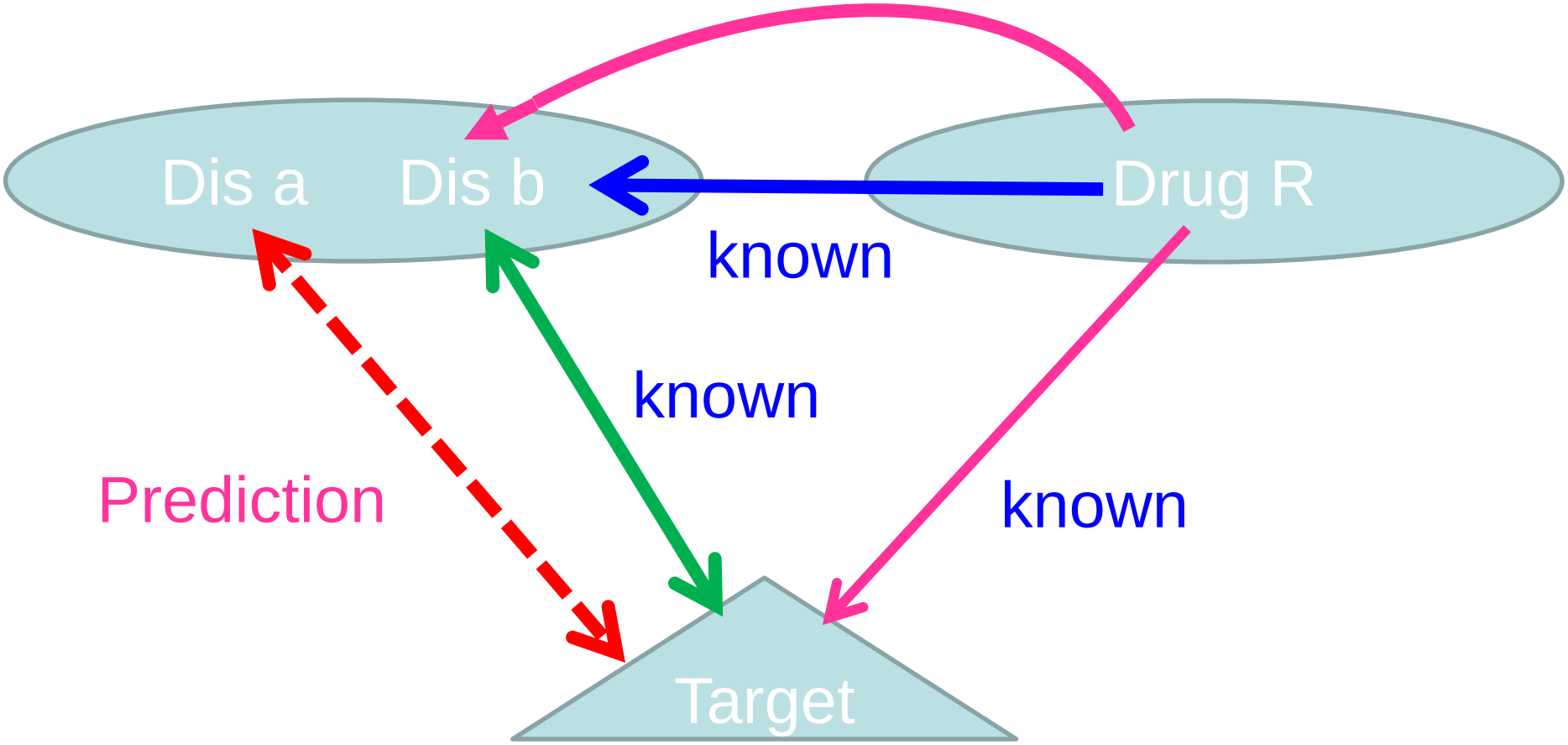
Possible DR for Alzheimer's Disease

- Computational network based methods may be among the most promising approaches for **computational drug repositioning.**
- Especially, drug-disease-target network would be useful resources to investigate novel indications for existing drugs.
- We mapped **predicted targets** on the drug-disease-target network and analysed the network to investigate **novel indications for existing drugs**

If predicted target for disease A is known drug-target of drug R for disease B, the drug R may be repositionable drug for disease A.



If predicted target for disease A is known drug-target of drug R for disease B, the drug R may be repositionable drug for disease A.



Potential (predicted) repositionable drugs for Alzheimer's disease

repositonable drug	taregt	# of target	category
Tamoxifen	PRKCB PRKCE PRKCG ESRRG	4	Anti-Estrogens; Antineoplastic Agents; Antineoplasti
Mianserin	SLC6A4 DRD3 OPRK1 ADRA1B	4	Adrenergic Agents; Adrenergic alpha-Antagonists; A
Amitriptyline	SLC6A4 OPRK1 ADRA1B OPRM1	4	
Dextromethorphan	SLC6A4 PGRMC1 OPRM1 OPRK1	4	Alkaloids; Antitussive Agents; Central Nervous Syste
Mirtazapine	OPRK1 ADRA1B DRD3 SLC6A4	4	Adrenergic Agents; Adrenergic alpha-Antagonists; A
Tramadol	OPRM1 OPRK1 SLC6A4	3	Alcohols; Amines; Analgesics; Analgesics, Opioid; C
Zinc	MPG SERPINA1 SERPIND1	3	Acetates; Acetic Acid; Acids; Acids, Acyclic; Acids, N
Amoxapine	SLC6A4 DRD3 ADRA1B	3	Adrenergic Agents; Adrenergic Uptake Inhibitors; Al
Etorphine	OPRM1 OPRK1 OPRL1	3	Alkaloids; Analgesics; Analgesics, Opioid; Central N
Tapentadol	OPRM1 OPRK1 SLC6A4	3	Analgesics; Analgesics, Opioid; Benzene Derivatives
Loxapine	ADRA1B DRD3 SLC6A4	3	Antipsychotic Agents; Antipsychotic Agents (First Ge
Pethidine	OPRK1 OPRM1 SLC6A4	3	Acids, Heterocyclic; Adjuvants; Adjuvants, Anesthesi
Talampanel	GRIA1	1	Benzazepines; Heterocyclic Compounds; Heterocycli
Etanercept	FCGR3B	1	Amino Acids, Peptides, and Proteins; Analgesics; A
Vitamin E	PRKCB	1	Antioxidants; Benzopyrans; Chemical Actions and Us
N-[(2R)-2-benzyl-4-(hydroxyamino)-4-	LTA4H	1	
Adalimumab	FCGR3B	1	Amino Acids, Peptides, and Proteins; Anti-Inflamm
ALPHA-HYDROXYFARNESYLPHOSPH	FNTB	1	Alcohols; Fatty Alcohols; Hydrocarbons; Lipids; Orga

Example,

The two FDA-approved drugs, **adalimumab and etanercept**, may be most promising candidates, because they are inhibitors of TNF-alpha (a key cytokine to regulate immune response) and overexpression of TNF-alpha cause inflammation in various organs, especially in central nerve system.

MedGenMed
Medscape General Medicine

MedGenMed. 2006; 8(2): 25.
Published online 2006 Apr 26.
PMCID: PMC1785182

TNF-alpha Modulation for Treatment of Alzheimer's Disease: A 6-Month Pilot Study

Edward Tobinick, MD, Assistant Clinical Professor of Medicine, Hyman Gross, MD, Clinical Professor of Neurology, Alan Weinberger, MD, Associate Clinical Professor of Medicine/Rheumatology, and Hart Cohen, MD, FRCPC, Associate Clinical Professor of Medicine/Neurology

CNS Drugs
November 2016, Volume 30, Issue 11, pp 1111–1120

Treatment for Rheumatoid Arthritis and Risk of Alzheimer's Disease: A Nested Case-Control Analysis

Authors Authors and affiliations

Richard C. Chou, Michael Kane, Sanjay Ghimire, Shiva Gautam, Jiang Gui

TMDU

Our computational method would be promising approach to investigate novel drug targets and new indications for FDA-approved drugs.

We are now applying the method to investigate novel targets and repositionable drugs for various diseases including various types of cancers, rheumatoid, diabetes and etc.

Future strategies and trends

- Second stage of **genomic medicine and drug discovery**
- Contracting method for multidimensional network by Deep Learning
  - Apply to big data medicine
  - Correlative network structure of comprehensive molecular information – clinical phenotype in genome medicine
  - Disease onset and genetic – environment factor in biobank
- Framework of AI drug discovery has possibility to achieve
- Undertake the implementation of AI drug discovery at the end of this year. Otherwise all will be taken to the United States.
  - Establish the "Big data medicine/AI drug discovery consortium" to promote the project, coordinated by pharmaceutical company, IT company and medical institution

TMDU