Syntekabio

ResiScan & CDRscan: Deep learning based Cancer drug response prediction

2017.12.02. 신테카바이오 용인 인실리코 의학센터

Paradigm Shift in Medicine

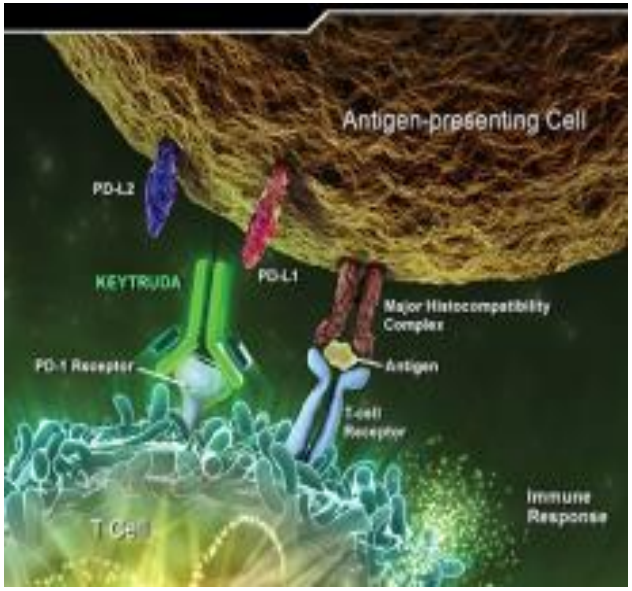
	Yesterday	Today	Tomorrow
Data	Symptom-based	Cohort-based	Algorithm-based
Actions	Intuitional medicine	Evidence-based medicine	Precision medicine
Hospital			
	• Larger hospitals (More Patients) • Better Intuitional Clinical Practices	• Specialized Hospitals • More Patients with the Similar Diseases • Cohort & Evidence based Practice	• Big Data & AI Hospitals • Application of Precision Medical Practice

First Accelerated FDA Approval

May 23 2017

Keytruda (Pembrolizumab) PD-L1 receptor existing list of approved indications:
Melanoma, lung cancer, Head & neck cancer, Lymphoma, Bladder cancer. More studies are currently underway

FDA granted accelerated approval for the first time to use for solid tumor with a **specific genetic feature as indication**.



"**PD-L1 발현율이 50% 이상이며, EGFR 및 ALK변이가 없는 진행형 비소세포폐암 환자**"의 1차 치료제로 적응 증을 확대승인

FDA News Release

FDA approves first cancer treatment for any solid tumor with a specific genetic feature

SHARE | TWEET | LINKEDIN | PIN IT | EMAIL | PRINT

For Immediate Release May 23, 2017

Release

The U.S. Food and Drug Administration today granted accelerated approval to a treatment for patients whose cancers have a specific genetic feature (biomarker). This is the first time the agency has approved a cancer treatment based on a common biomarker rather than the location in the body where the tumor originated.

Keytruda (pembrolizumab) is indicated for the treatment of adult and pediatric patients with unresectable or metastatic solid tumors that have been identified as having a biomarker referred to as microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR). This indication covers patients with solid tumors that have progressed following prior treatment and who have no satisfactory alternative treatment options and patients with colorectal cancer that has progressed following treatment with certain chemotherapy drugs.

Coming of Age

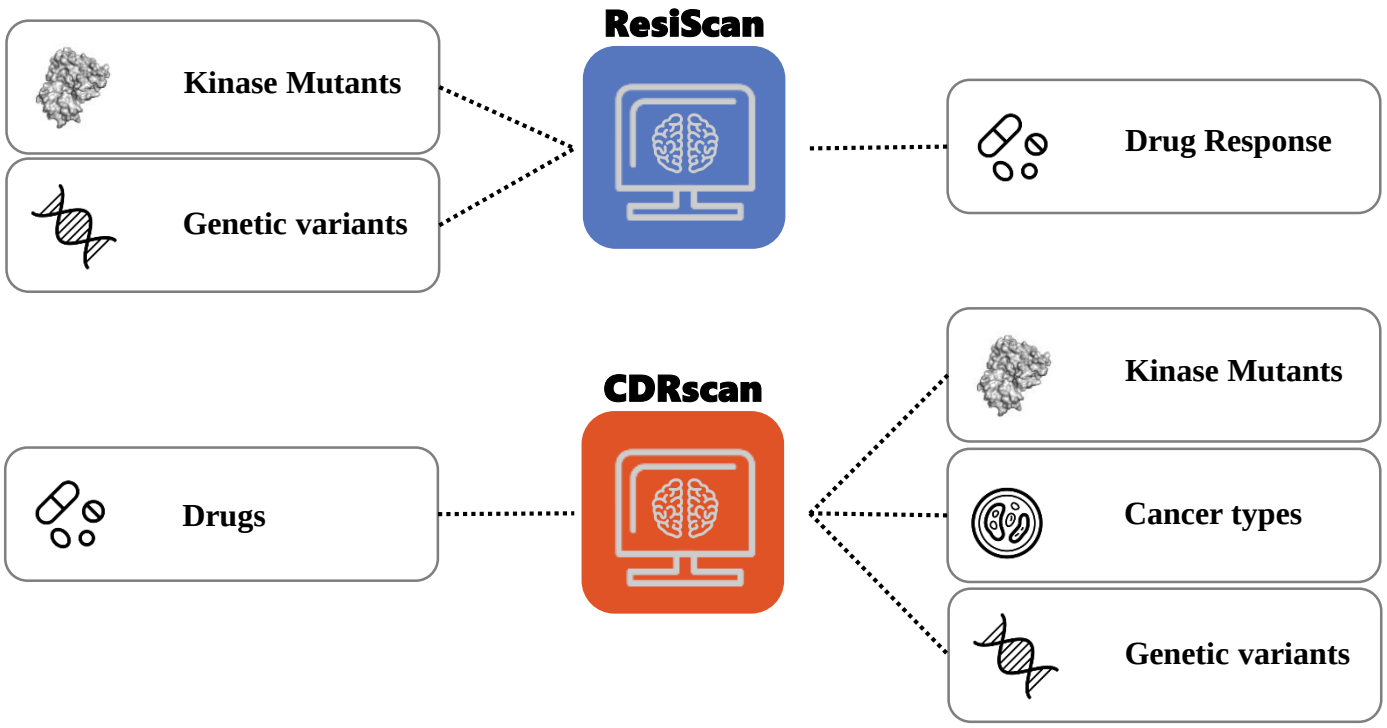
Global Pharma Investment Trend in Precision Medicine (USD$ investment amount)

National Human Genome Research Institute. The Cost of Sequencing a Human Genome. Accessed September 13, 2016 at http://www.genome.gov/sequencingcosts.
Data provided by: Concert Genetics. Available at concertgenetics.com. *Methodological notes: Concert Genetics began publishing the first reliable data on the number of genetic testing products available in January of 2016. PMC has published a list of 127 genetic tests commonly associated with the 132 personalized medicines listed in the Appendix of this document at http:// www.personalizedmedicinecoalition.org/Education/Tests.

AI를 통한 신약개발 업체

업체	데이터	질환	국가
베네볼런트	논문, 임상데이터	난치성 질환	영국
투사(twoXAR)	단백질 상호작용, 임상 데이터	희귀 피부질환, 간암 등	미국
인실리코 메디슨	논문,임상데이터	노화	미국
아톰와이즈	단백질-약물 결합구조	에볼라	미국
버그(Berg)	유전자, in vitro 실험	암	미국
이화학연구소	논문, 임상데이터	암, 치매	일본
스탠다임	유전자 발현-패스웨이	Drug repositioning	한국

Strategy

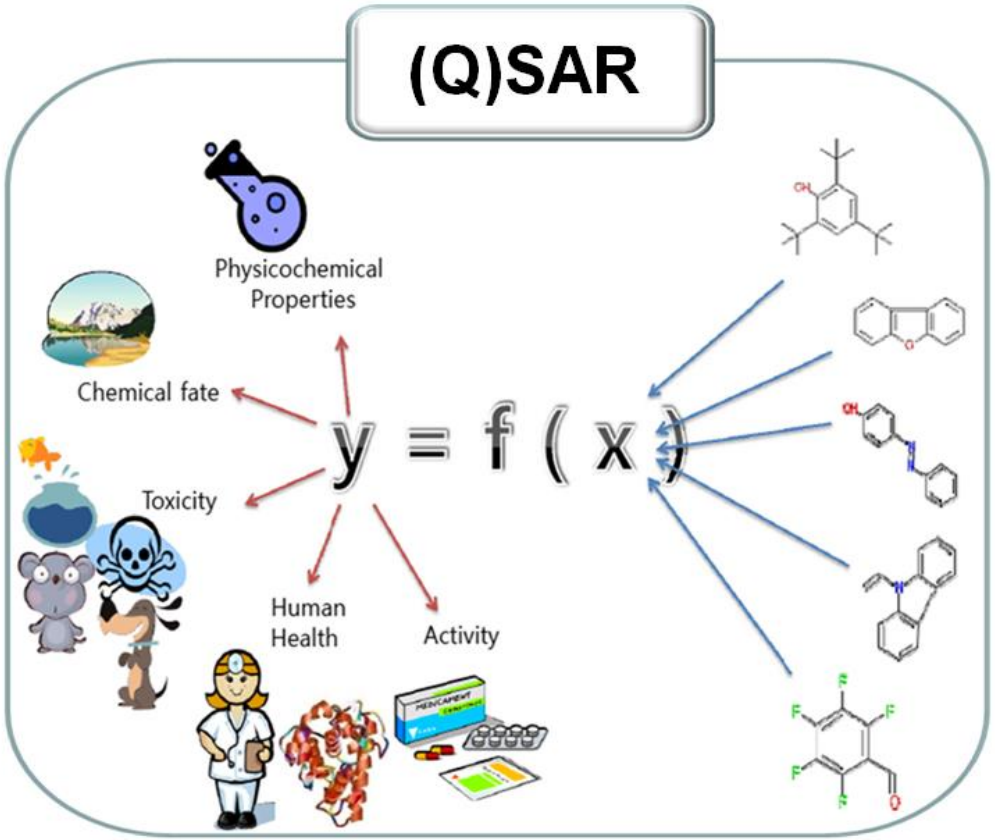

Kinase Mutants
Genetic variants
ResiScan
Drug Response
Drugs
CDRscan
Kinase Mutants
Cancer types
Genetic variants

ResiScan: 딥러닝 기반 약물 내성변이 예측

Traditional *In silico* Model (SAR)

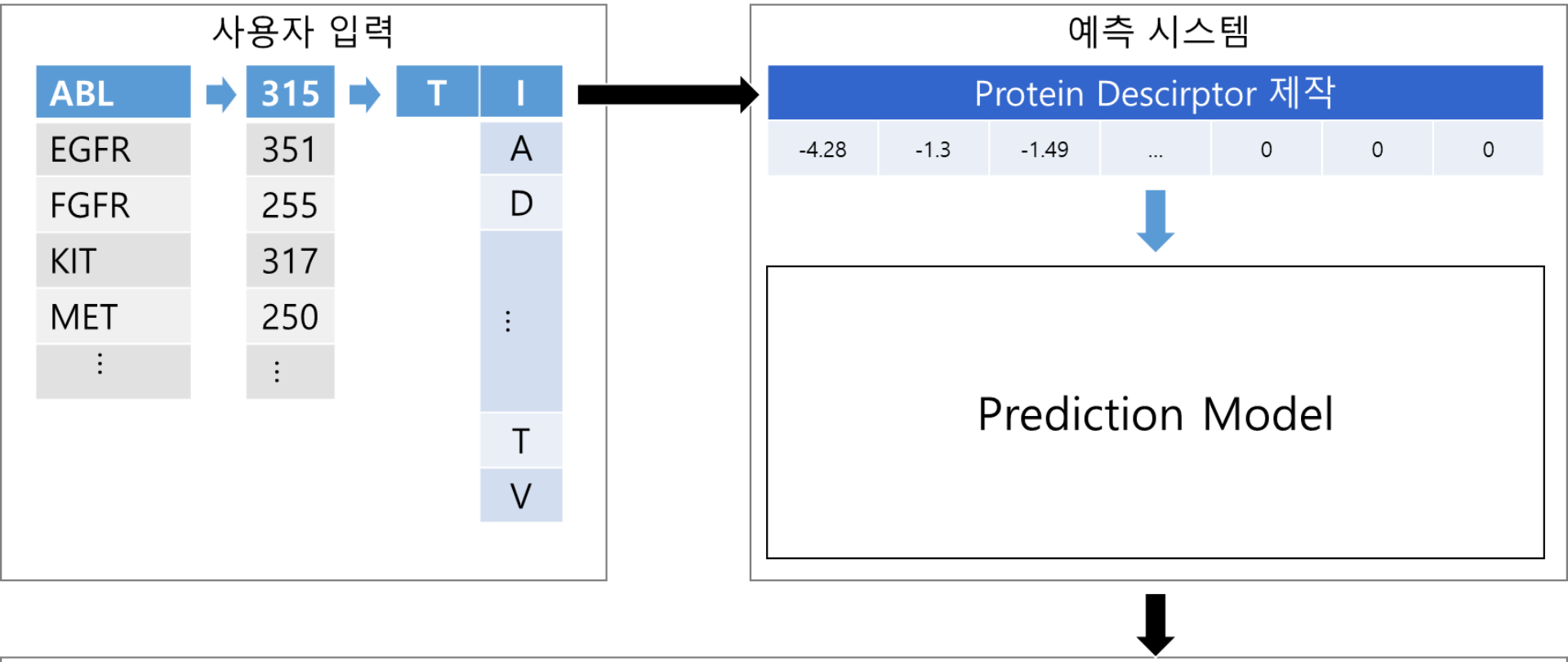
SAR (Structure-Activity-Relationship) ➔ Finding F(X)

Data→Model→Knowledge



End-Points **Data→Descriptors**

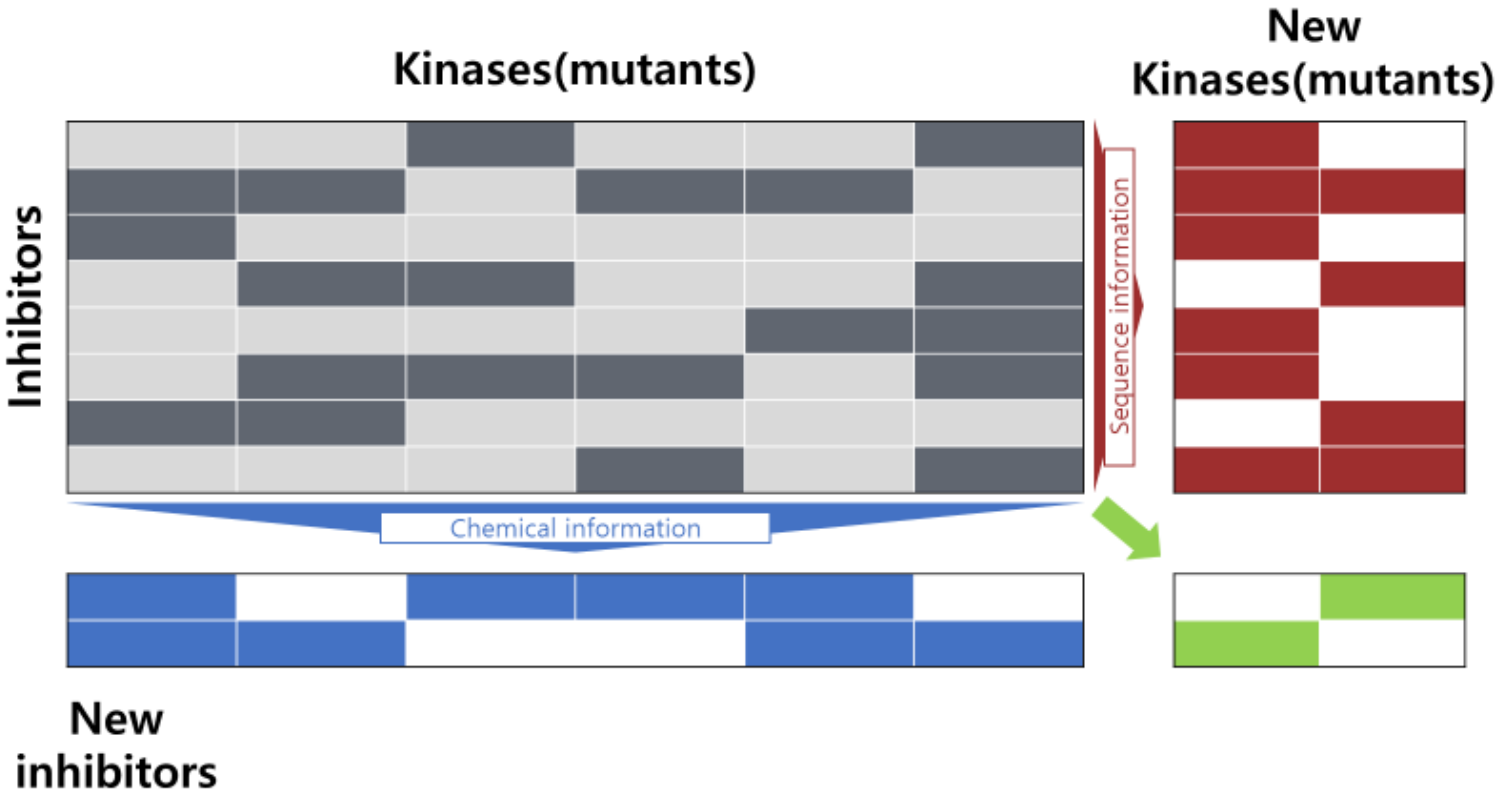
System 개요



예측 결과

Drug Inhibition Prediction Results											
Afatinib	Bosutinib isomer	Dasatinib	Erlotinib	Gefitinib	Imatinib	Lapatinib	Nilotinib	Pazopanib	Sorafenib	Sunitinib	Vandetanib
X	O	O	X	X	X	X	O	X	X	X	X

Proteochemometrics(PCM) model



Drug Data Information

- **Duong-Ly KC, *et.al.* Kinase Inhibitor Profiling Reveals Unexpected Opportunities to Inhibit Disease-Associated Mutant Kinases. *Cell Rep.* 2016 Feb 2;14(4):772-81.**

	Kinase mutant	Drug
Number of Data	96 (wildtype:21개)	183
Descriptor	Amino acid one-hot encoding	CDK fingerprint
Number of Feature	9900 (495*20)	3072
total	15702 (Data) * 12972 (Feature)	

Data structure

Length : 12972

Length : 15702

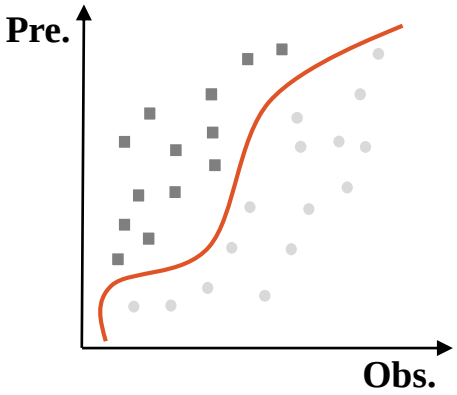
	Protein Descirptor(9900)							Drug Descirptor(3072)					Percent Remainig Inhibition Activity
C-KIT_A829P +Afatinib	0	1	0	...	0	0	0	0	1	...	0	0	97.71
LRRK2_G2019S +AG 1024	0	0	0	...	0	1	0	0	0	...	1	0	95.87
LRRK2 +AG 112	0	0	0	...	0	1	0	0	1	...	1	0	97.09
MEK1 +AG 1295	0	0	0	...	0	0	0	0	0	...	0	0	97.98
C-MET +AG 1296	0	0	1	...	1	0	0	0	0	...	0	1	92.59
C-MET_F1200I +AG 1478	0	0	1	...	1	0	0	0	1	...	1	0	77.95
TIE2_R849W +AG 490	0	1	0	...	0	1	0	0	0	...	0	0	98.85
p38-alpha +AG 9	0	0	0	...	1	0	0	0	0	...	0	1	100
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Virtual docking

기존 Model

Binding site in proteins **Drugs**

Machine learning



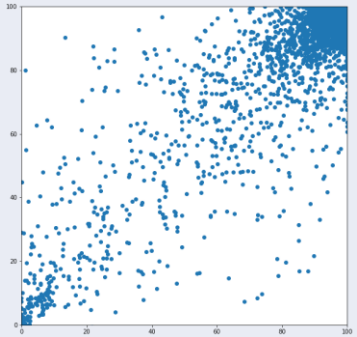
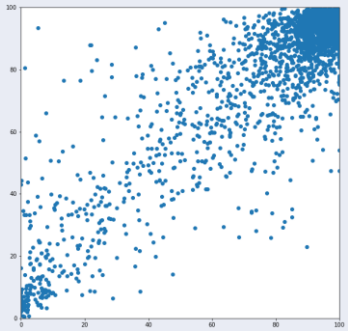
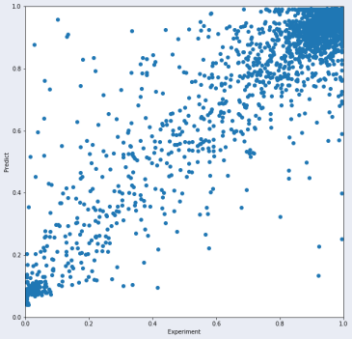
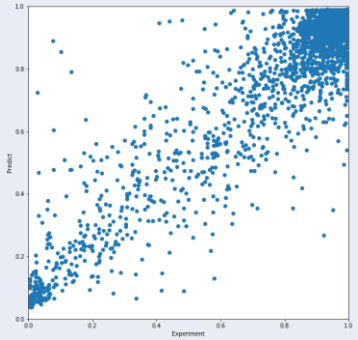
Virtual Docking Model

Binding site in proteins **Drugs**

Virtual Docking

Result: Regression

- 기존 머신러닝(SVM, Random Forest) 방법과 비교

	SVM	RF	CNN	Virtual docking
R2	0.7799	0.8179	0.8354	0.8741
Scatter plot				

- SVM과 RF의 경우, Feature Selection 과정을 거침

Result: Classification

- 기존 머신러닝(SVM, Random Forest) 방법과 비교

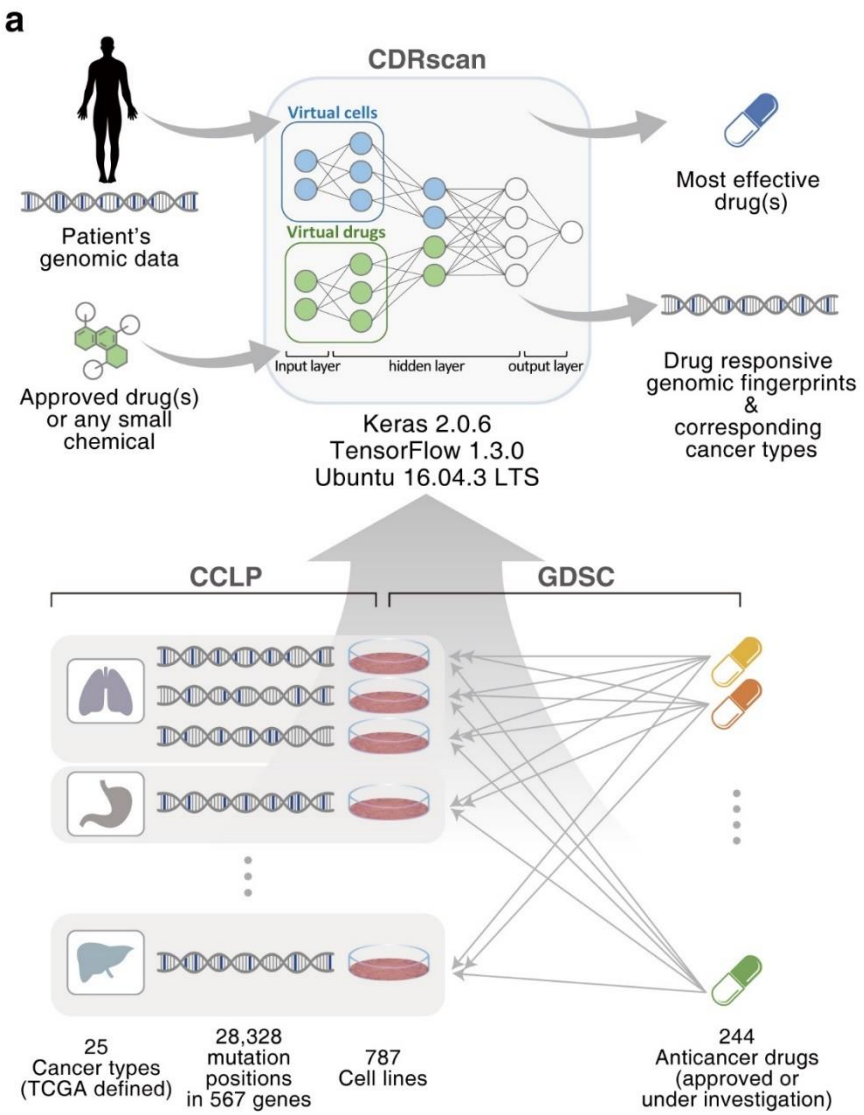
	Feature		Accuracy	Specificity	AUC
	Protein	Drug			
DNN (1D, earlystop)	9900	3072	0.901	0.967	0.940
SVM	499	2355	0.893	0.929	0.919
RF	499	2355	0.895	0.942	0.903

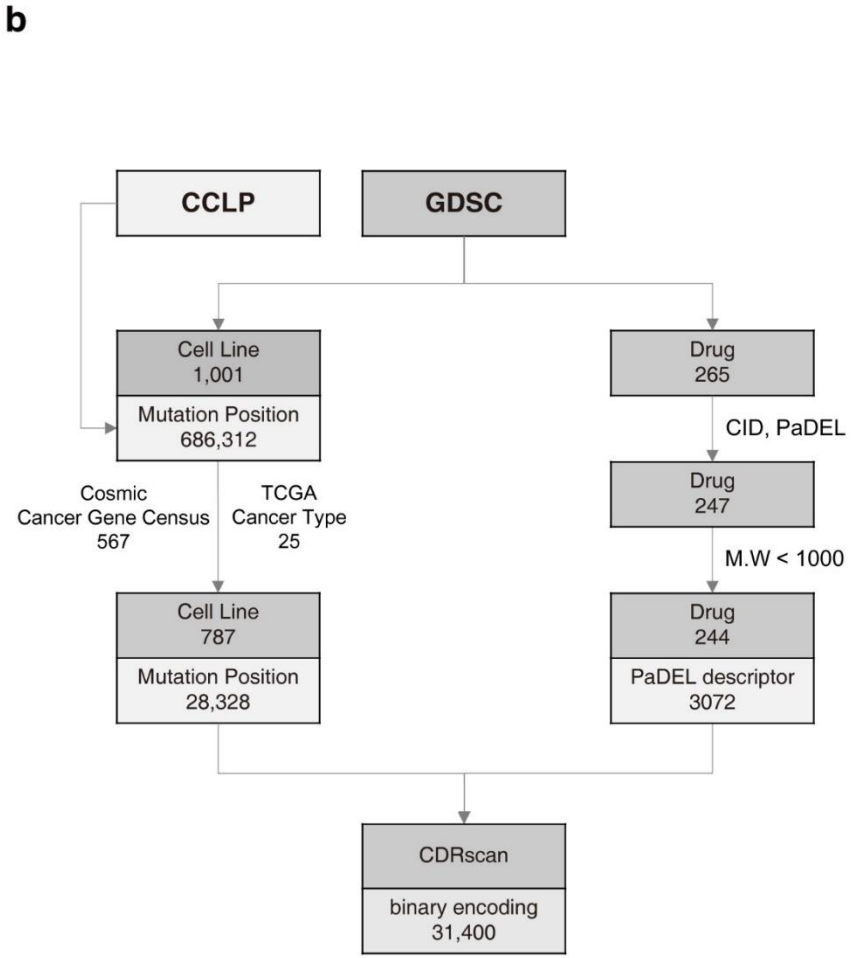
- SVM과 RF의 경우, Feature Selection 과정을 거침

CDRscan: 딥러닝 기반 항암 약물 반응 예측

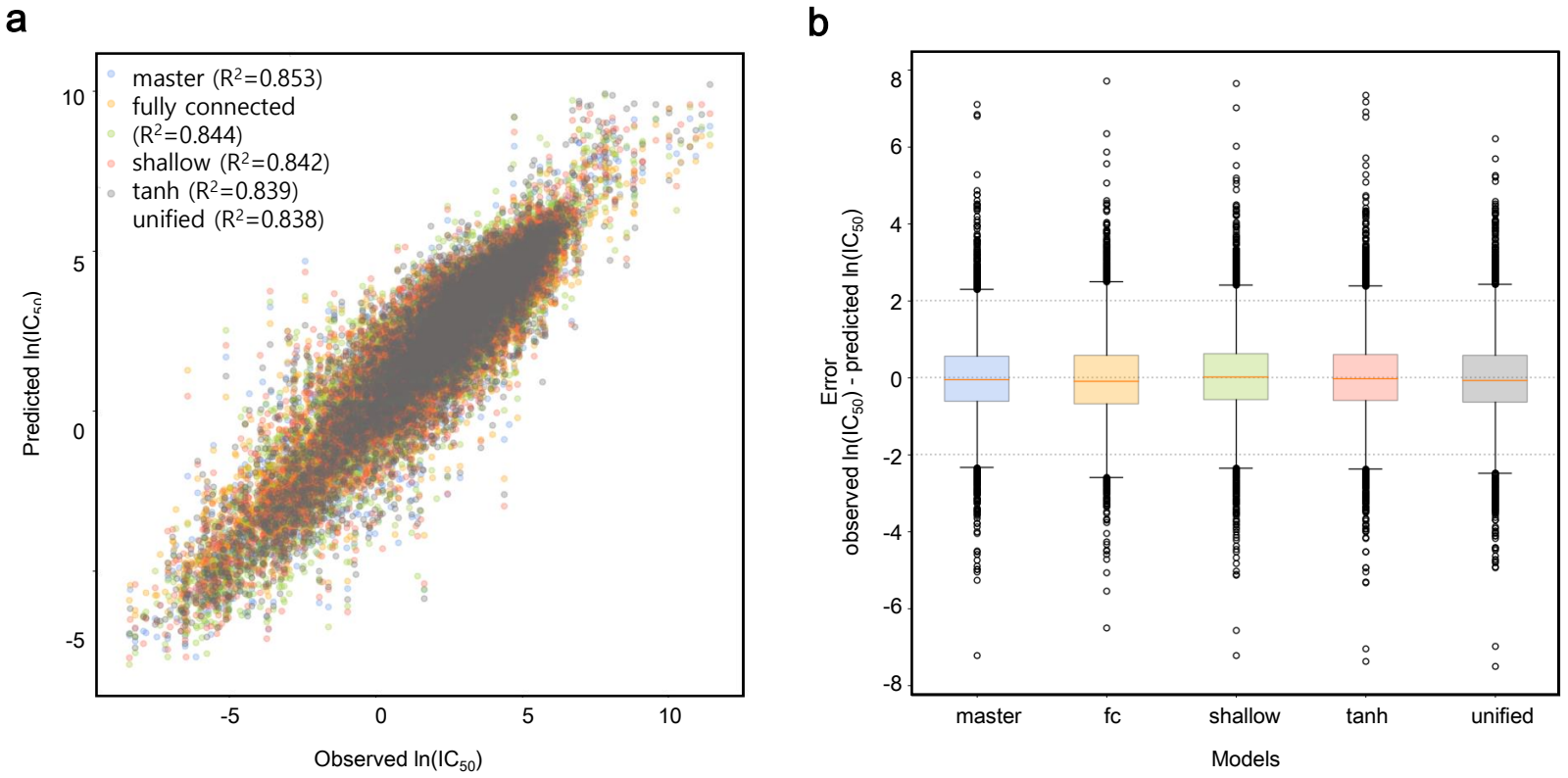
개요

Figure 1

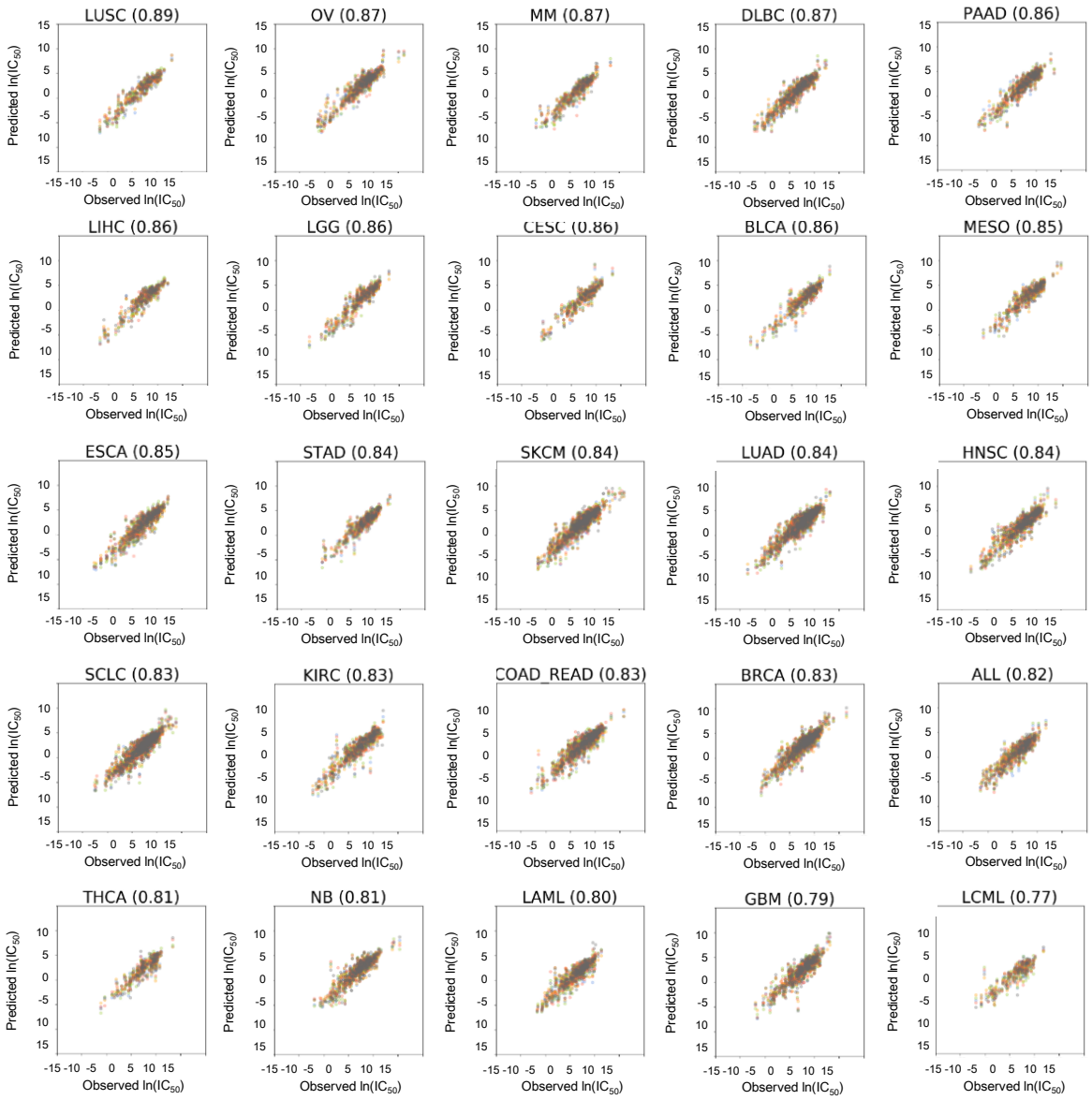

a
CDRscan
Virtual cells
Virtual drugs
Input layer
hidden layer
output layer
Patient's genomic data
Approved drug(s) or any small chemical
Most effective drug(s)
Drug responsive genomic fingerprints & corresponding cancer types
Keras 2.0.6
TensorFlow 1.3.0
Ubuntu 16.04.3 LTS
CCLP
GDSC
25 Cancer types (TCGA defined)
28,328 mutation positions in 567 genes
787 Cell lines
244 Anticancer drugs (approved or under investigation)


b
CCLP
GDSC
Cell Line 1,001
Mutation Position 686,312
Cosmic Cancer Gene Census 567
TCGA Cancer Type 25
Cell Line 787
Mutation Position 28,328
Drug 265
CID, PaDEL
Drug 247
M.W < 1000
Drug 244
PaDEL descriptor 3072
CDRscan
binary encoding 31,400

예측 결과

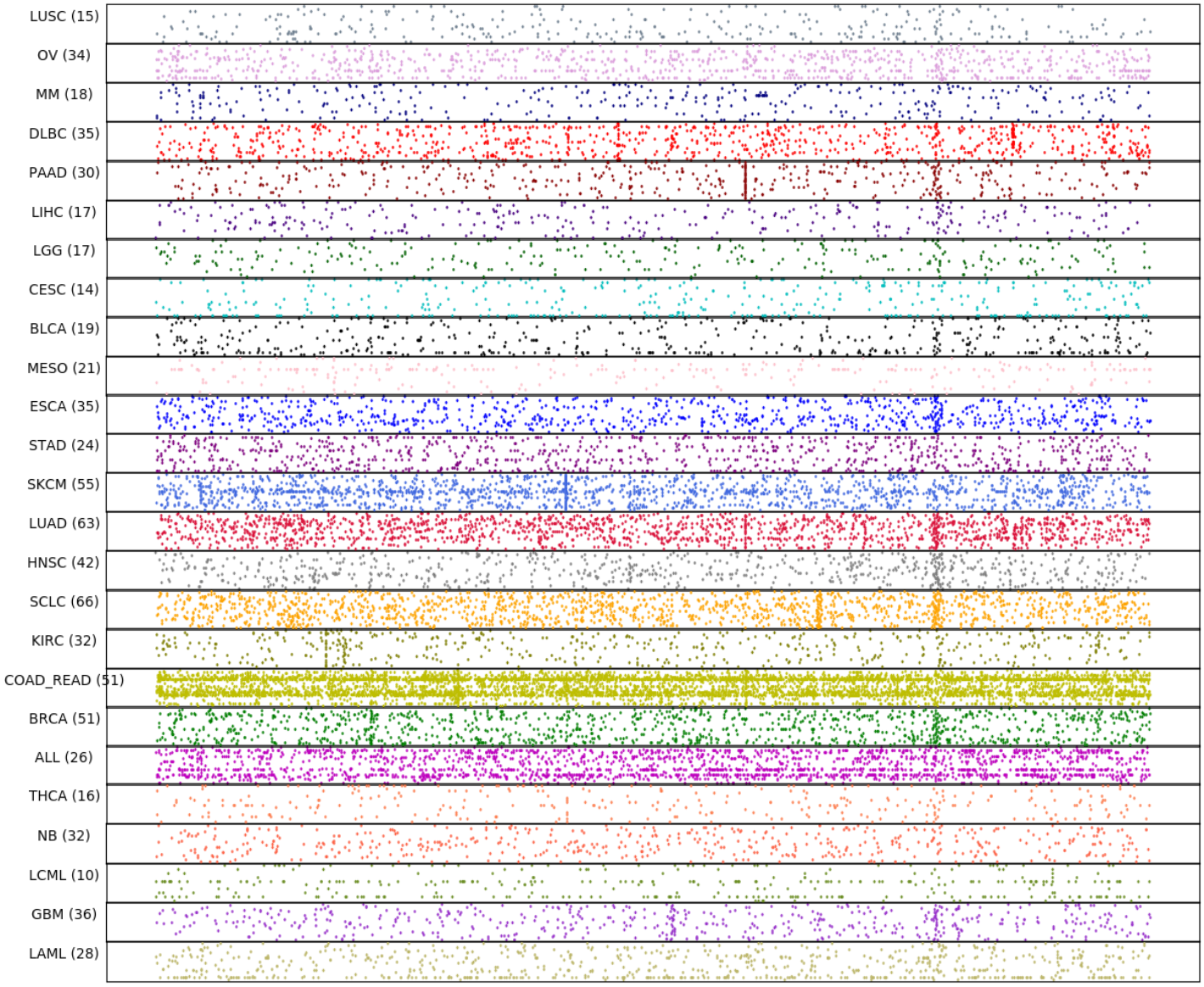

a
master (R^2=0.853)
fully connected (R^2=0.844)
shallow (R^2=0.842)
tanh (R^2=0.839)
unified (R^2=0.838)
Predicted ln(IC_{50})
Observed ln(IC_{50})
10
5
0
-5
-5
0
5
10
b
Error
observed ln(IC_{50}) - predicted ln(IC_{50})
8
6
4
2
0
-2
-4
-6
-8
master
fc
shallow
tanh
unified
Models

암 종류별 예측 결과



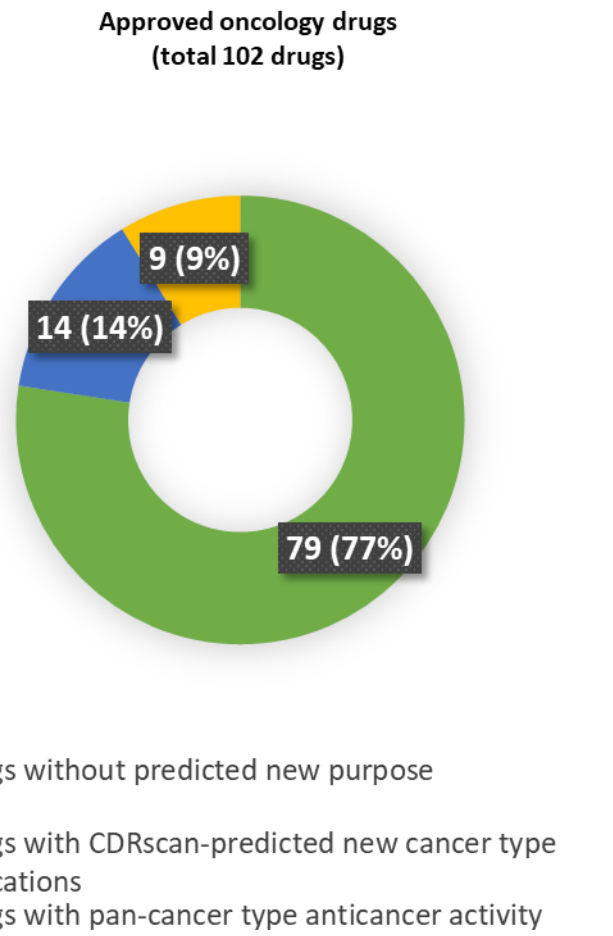
- master
- fully connected
- shallow
- tanh
- unified

암 종류별 변이

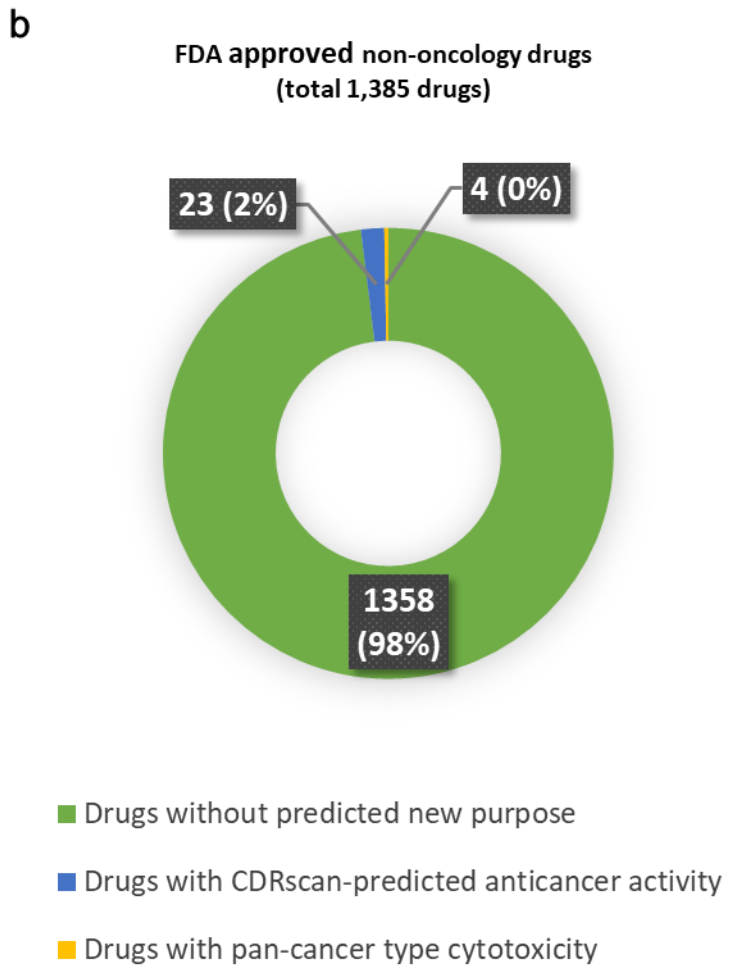

LUSC (15)
OV (34)
MM (18)
DLBC (35)
PAAD (30)
LIHC (17)
LGG (17)
CESC (14)
BLCA (19)
MESO (21)
ESCA (35)
STAD (24)
SKCM (55)
LUAD (63)
HNSC (42)
SCLC (66)
KIRC (32)
COAD_READ (51)
BRCA (51)
ALL (26)
THCA (16)
NB (32)
LCML (10)
GBM (36)
LAML (28)

FDA 승인 약물 기반 예측

a



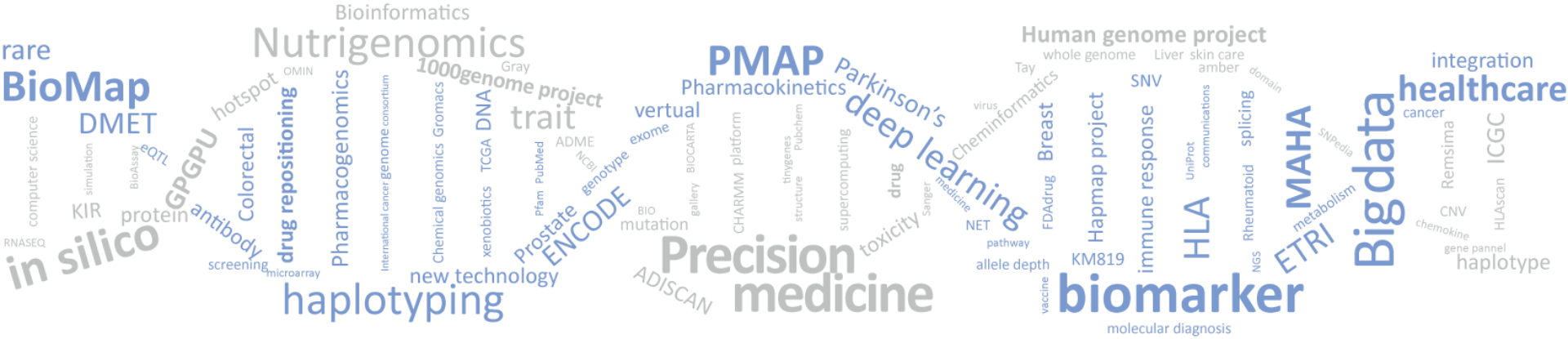
b



Conclusion

- 최근 유전체 분석 등 과학 기술의 발전에 따라 환자 개개인에 최적화된 진단 및 치료를 제공하는 정밀의학의 시대가 도래했다.
- 분자 수준(kinase mutant)과 세포 수준(cancer cell line)의 약물 활성 예측은 약물 저항성 관련 진단과 신약 개발에 도움을 줄 것으로 예상됨.
- Kinase mutant에 대한 약물 활성 예측 모형은 virtual docking 기법을 사용하여 더 높은 정확도를 가짐.
- 암세포주 데이터를 활용한 약물 활성 예측 모형은 약물 저항성 뿐 아니라 신약 개발 단계나 신약 재창출(drug repositioning) 단계에도 사용될 수 있음.

감사합니다.


rare
BioMap
DMET
KIR
protein
in silico
RNASEQ
computer science
simulation
BioAssay
eQTL
GPGPU
hotspot
OMIN
antibody
screening
microarray
Colorectal
drug repositioning
Pharmacogenomics
Bioinformatics
Nutrigenomics
International cancer genome consortium
Chemical genomics
Gromacs
xenobiotics
TCGA
DNA
1000genome project
Gray
trait
ADME
Pfam
PubMed
Prostate
ENCODE
genotype
exome
vertual
BIO
mutation
new technology
haplotyping
gallery
BIOCARTA
CHARMM platform
tinygenes
structure
Pubchem
supercomputing
PMAP
Pharmacokinetics
Parkinson's
deep learning
drug
toxicity
Sanger
medicine
Precision
medicine
ADISCAN
Cheminformatics
virus
Tay
Human genome project
whole genome
Liver
skin care
amber
domain
SNV
Breast
Hapmap project
immune response
splicing
UniProt
communications
FDAdrug
NET
pathway
allele depth
KM819
HLA
Rheumatoid
NGS
vaccine
biomarker
molecular diagnosis
MAHA
metabolism
ETRI
SNPedia
Bigdata
integration
healthcare
cancer
Remsima
ICGC
CNV
chemokine
HLAscan
gene pannel
haplotype