Drug Synergy Study Based on Network and Multi-level Omics Annotation

The 15th Korea-China-Japan Bioinformatics Symposium, June 21-22, Seoul, Korea

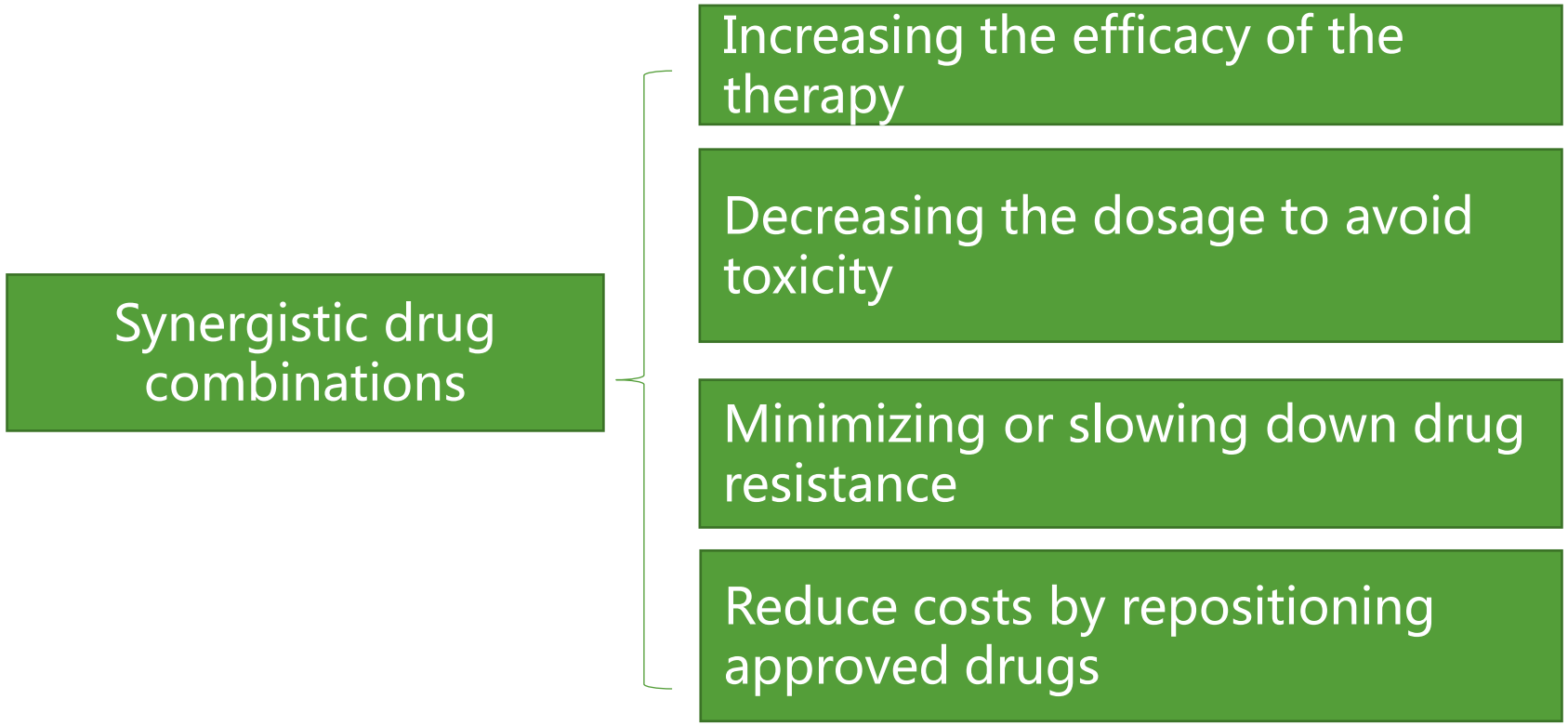
Lu XIE, Shanghai Center for Bioinformation Technology

Necessity of combination therapy

- Genetic, epigenetic and metabolic factors all contribute, in molecular networks that govern growth, development, death and specialization of tumor cells
- Resistance occurs and is recurrent
- Network-based multi-target identification may be crucial

Drug Discovery Today: http://dx.doi.org/10.1016/j.ddtec.2015.06.002

Advantages in combination therapy

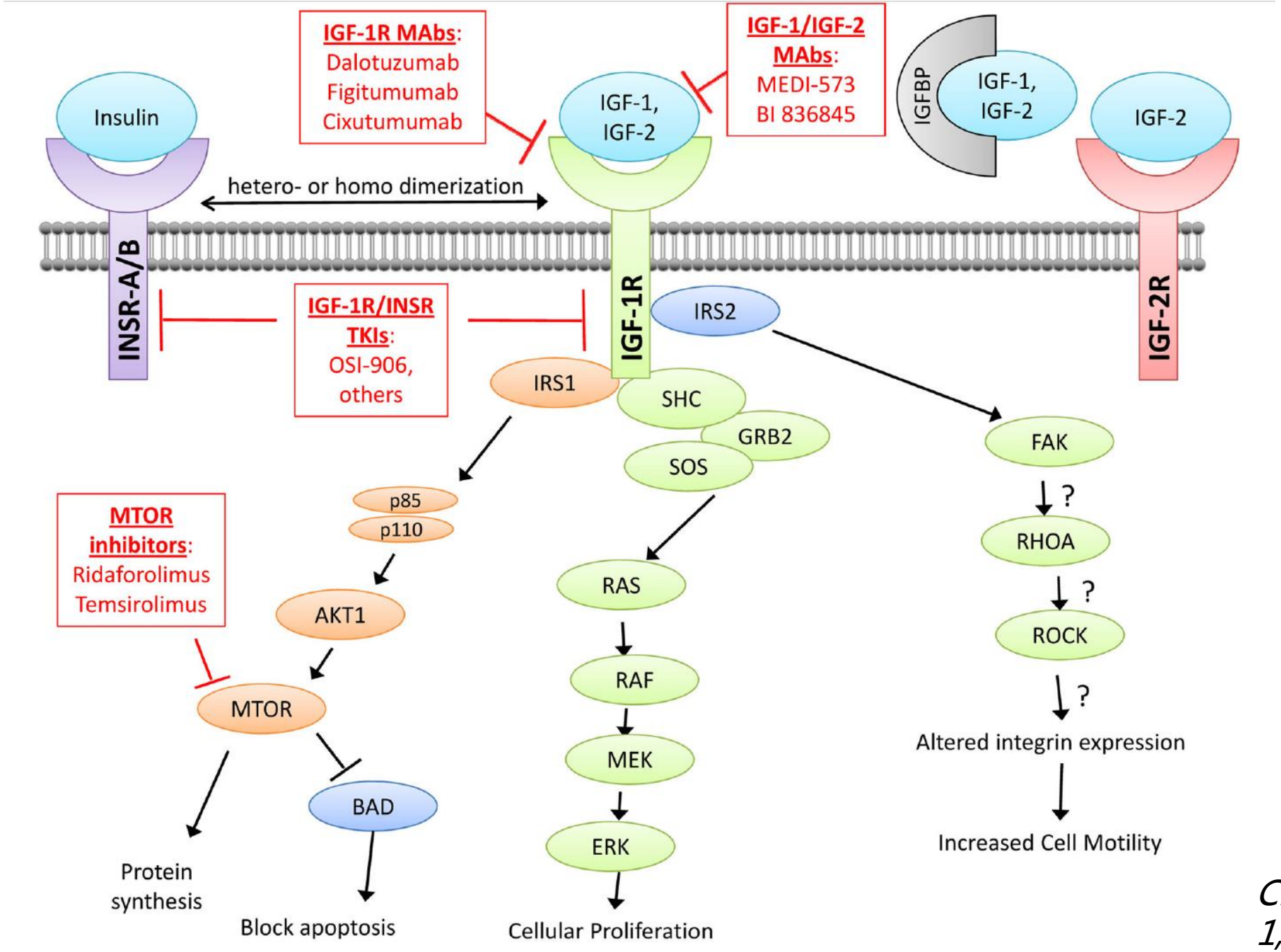


Combination design

Basic rule: Synergistic

1. Space based: extracellular+intracellular; receptor+effector; microenvironment+tumor tissue; …
2. Function based: target therapy+cell cycle chemotherapy (mitosis); anti-proliferation+pro-apoptosis; epigenetic+genetic; singel pathway; cross-talk pathways; …
3. Time based: sequential; repeating;…

Golden standards only based on clinical trials



Example: IGF signaling pathway is a complex and tightly regulated network which is critical for cell proliferation and survival

Clin Cancer Res. 2015 October 1; 21(19): 4270–4277.

Drug synergy modeling based on cell lines

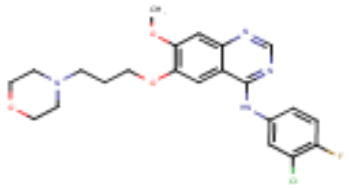
Prediction model of synergistic anti-cancer drug combinations

Supervised learning, DREAM as traingin and CMAP data as test

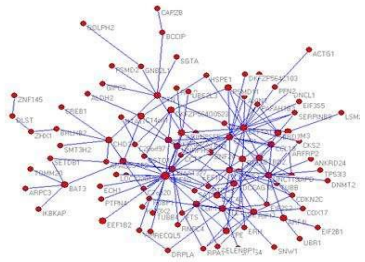
Drug target network features, Pharmacogenomics features, Random forest algorithm

Nat Biotechnol. 2014. A community computational challenge to predict the activity of pairs of compounds

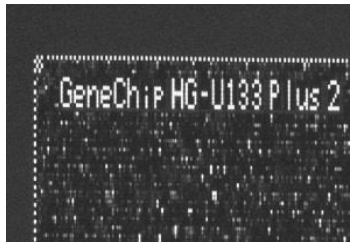
Training dataset



Chemical structure



Drug target network



Gene chip data (NCBI GEO)

①

②

③

DREAM Challenge 7 sub-challenge 2

	Compound Name
1	Aclacinomycin A
2	Blebbistatin
3	Camptothecin
4	Cycloheximide
5	Doxorubicin hydrochloride
6	Etoposide
7	Geldanamycin
8	H-7, Dihydrochloride
9	Methotrexate
10	Mitomycin C
11	Monastrol
12	Rapamycin
13	Trichostatin A
14	Vincristine

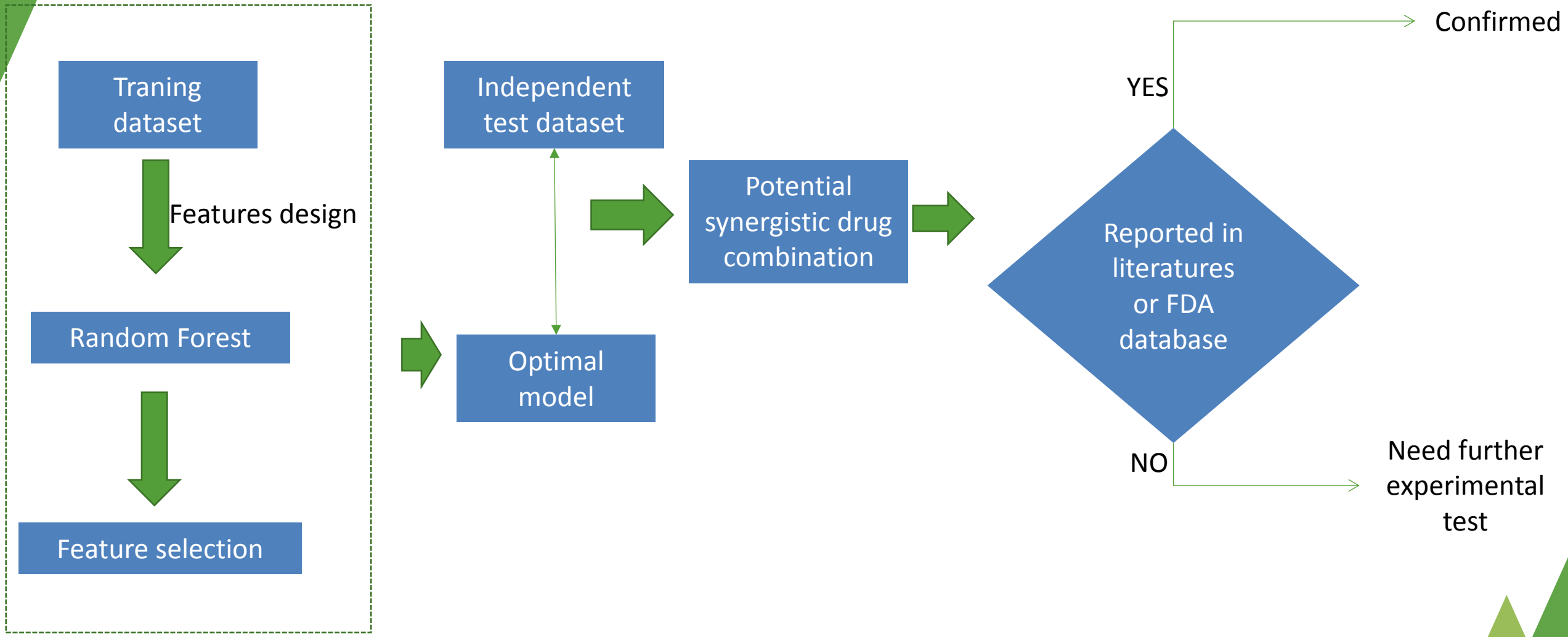
OCI-LY3 diffuse large B-cell lymphoma (DLBCL) cell line

④

	Cmpd A	Cmpd B	Label
1	Doxorubicin	H-7	Synergy
2	H-7	Mitomycin C	Synergy
3	Camptothecin	Mitomycin C	Synergy
4	Doxorubicin	Mitomycin C	Synergy
5	Etoposide	Mitomycin C	Synergy
6	Etoposide	H-7	Synergy
7	Cycloheximide	H-7	Synergy
8	Doxorubicin	Trichostatin A	Synergy
9	Blebbistatin	H-7	Synergy
10	Cycloheximide	Monastrol	Synergy
11	Monastrol	Trichostatin A	Synergy
……	……	……	……
85	Blebbistatin	Camptothecin	Non-synergy
86	Doxorubicin	Monastrol	Non-synergy
87	Etoposide	Monastrol	Non-synergy
88	Aclacinomycin A	Trichostatin A	Non-synergy
89	Blebbistatin	Trichostatin A	Non-synergy
90	Blebbistatin	Cycloheximide	Non-synergy
91	Camptothecin	Rapamycin	Non-synergy

16 positive

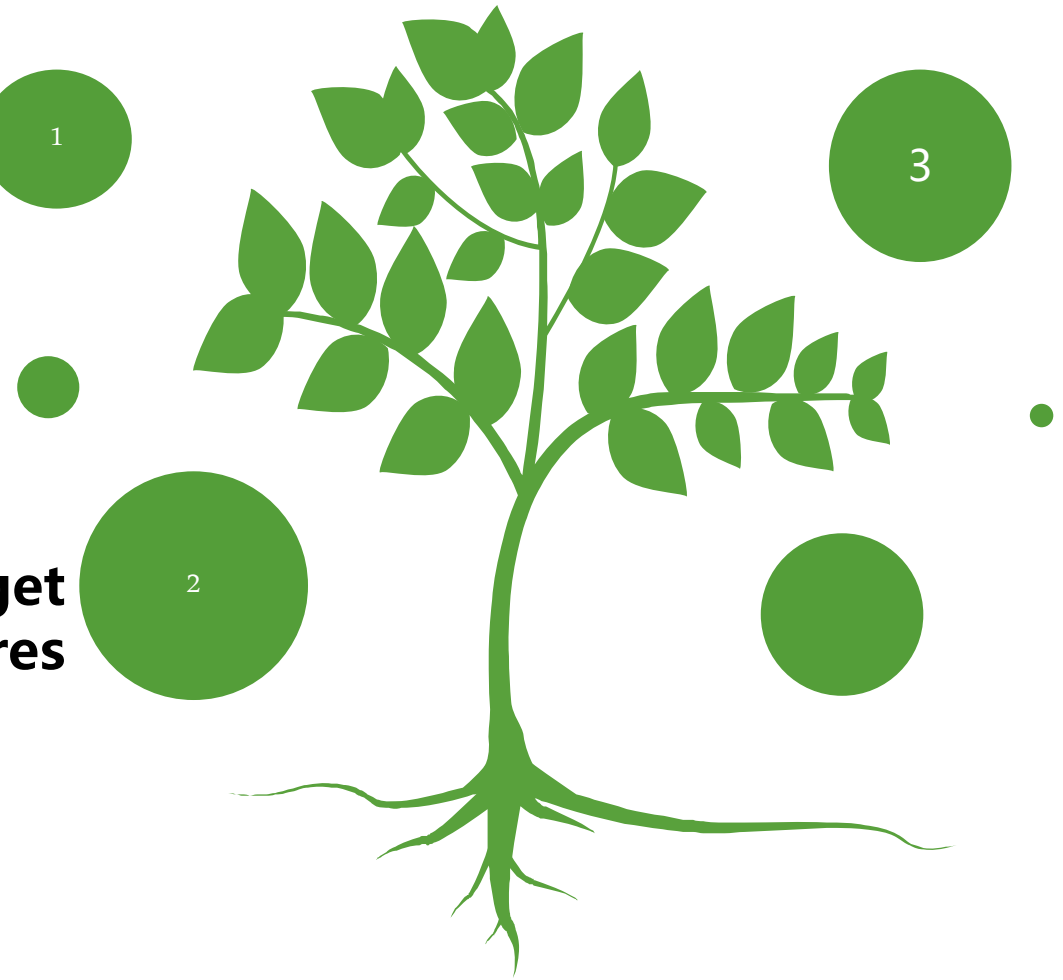
Work flow


Traning dataset
Features design
Random Forest
Feature selection
Independent test dataset
Optimal model
Potential synergistic drug combination
Reported in literatures or FDA database
YES
Confirmed
NO
Need further experimental test

Designing features for drug combinations

Drug chemical structure

Drug target network features



Pharmacogenomics features

Drug chemical structure

(1) Similarity score of drug chemical structure(based on tanimoto coefficient)

$$S_c = \frac{AB}{A + B - AB} \qquad (1)$$

Drug target network features

Shortest distance in PPI network

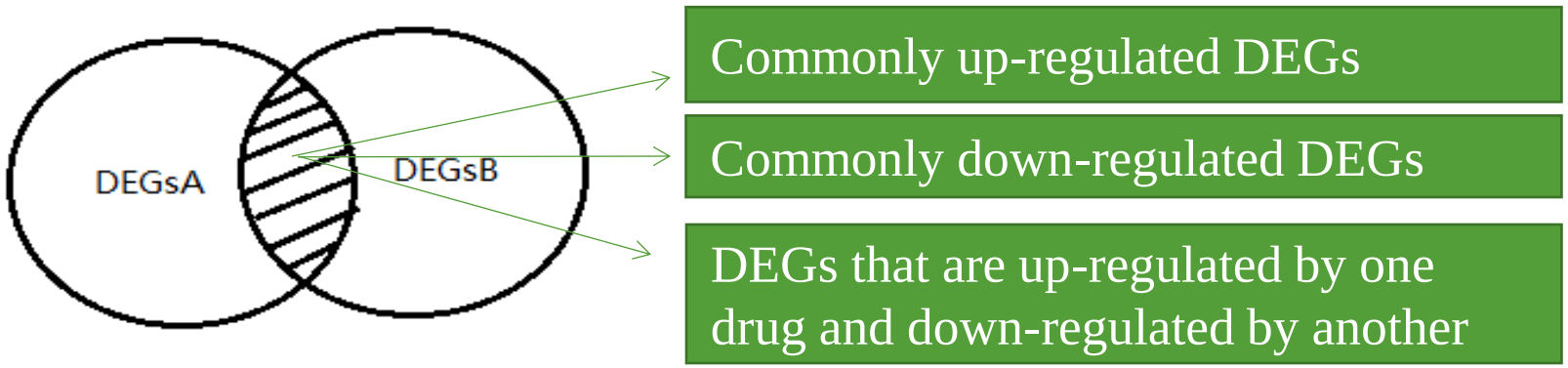
$$S_{dtd}(d_1, d_2) = \frac{\sum_{i=1}^{i=N1} \sum_{j=1}^{j=N2} distance(t_i, t_j)}{N1 * N2} \quad (2)$$

Similarity of drug targeted KEGG pathways

$$s(p_1, p_2) = \frac{|geneset(p_1) \cap geneset(p_2)|}{|geneset(p_1) \cup geneset(p_2)|}$$

$$S_{dtps}(d_1, d_2) = \frac{\sum_{i=1}^{i=M} \sum_{j=1}^{j=N} s(p_i, p_j)}{M * N} \quad (3)$$

Pharmacogenomics features



Also common DEGs in 8 Growth related pathways (GP)

DIGRE. CPT Pharmacomet Syst Pharmacol 4 (2) (2015)

① Features based on commonly up-regulated DEGs

$$Common\_up1 = 0.5*(\frac{N1}{|DEGsA|} + \frac{N1}{|DEGsB|}) \quad (4)$$

$$Common\_up2 = 0.5*(\frac{N1'}{DEGsA} + \frac{N1'}{DEGsB}) \quad (5)$$

$$Common\_up3 = \frac{N1'}{|geneset(GP)|} \quad (6)$$

② Features based on commonly down-regulated DEGs

$$Common\_dn\ 1 = 0.5*(\frac{N2}{|DEGsA|} + \frac{N2}{|DEGsB|}) \quad (7)$$

$$Common\_dn\ 2 = 0.5*(\frac{N2'}{DEGsA} + \frac{N2'}{DEGsB}) \quad (8)$$

$$Common\_dn3 = \frac{N2'}{|geneset(GP)|} \quad (9)$$

③ Features based on DEGs that up-regulated by one drug and down-regulated by another

$$Opposite1 = 0.5*(\frac{N3}{|DEGsA|} + \frac{N3}{|DEGsB|}) \quad (10)$$

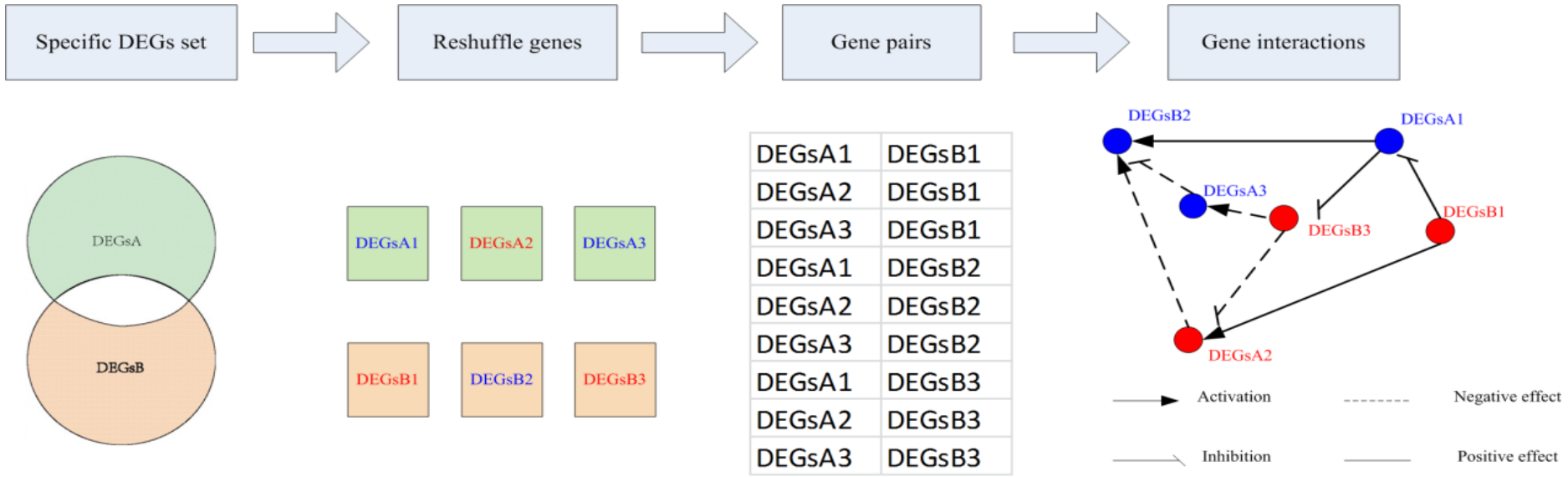
$$Opposite\ 2 = 0.5*(\frac{N3'}{DEGsA} + \frac{N3'}{DEGsB}) \quad (11)$$

$$Opposite3 = \frac{N3'}{|geneset(GP)|} \quad (12)$$

④ Features based on common DEGs in cell growth-related pathways

$$Average\_overlap = 0.5*(\frac{N4+N5}{|geneset\ (GP)|}) \quad (13)$$

Non-common DEGs denote genes that are only differentially expressed after drug A treatment or drug B treatment



The scoring process for non-common DEGs

Regulation relationships of 132 cancer-related pathways(CRPs) were extracted

Nat Commun 6 (2015) 8481

$$Up\_dn\_po1 = 0.5*(\frac{N6}{DEGsA} + \frac{N6}{DEGsB}) \quad (14)$$

$$Up\_dn\_ne1 = 0.5*(\frac{N7}{DEGsA} + \frac{N7}{DEGsB}) \quad (15)$$

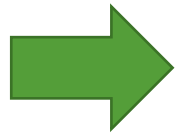
$$Up\_dn\_po2 = \frac{N6}{|geneset(CRPs)|} \quad (16)$$

$$Up\_dn\_ne2 = \frac{N7}{|geneset(CRPs)|} \quad (17)$$

Features based on differential expressed drug transpoter genes: drug efflux genes and drug influx genes

Table 3-1 drug efflux genes and drug influx genes

Gene Category	Gene Name
Drug efflux genes	SLC47A1 SLC47A2 ABCB1 ABCC2 ABCC3 ABCC4 ABCG2
Drug influx genes	SLC15A1 SLC15A2 SLC22A1 SLC22A2 SLC22A4 SLC22A5 SLC22A6 SLC22A7 SLC22A8 SLC22A9 SLCO1A2 SLCO1B1 SLCO1B3 SLCO2B1 SLCO4C1

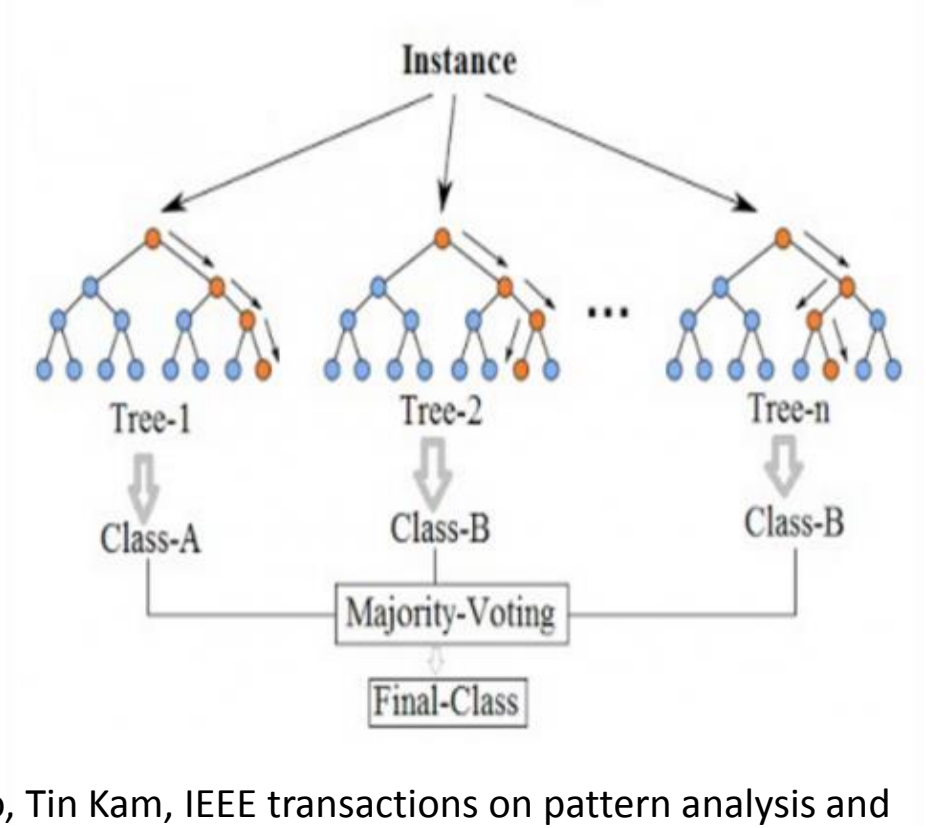


$$efflux\_po = N_8 \quad (18)$$
$$efflux\_ne = N_9 \quad (19)$$
$$influx\_po = N_{10} \quad (20)$$
$$influx\_ne = N_{11} \quad (21)$$

Drug efflux genes belong to ATP-binding cassette (ABC) family
Drug influx genes belong to Solute Carrier (SLC) family

Random forest algorithm

Random forest is operated by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.



Ho, Tin Kam, IEEE transactions on pattern analysis and machine intelligence 20.8 (1998): 832-844.

Advantages of Random Forest：

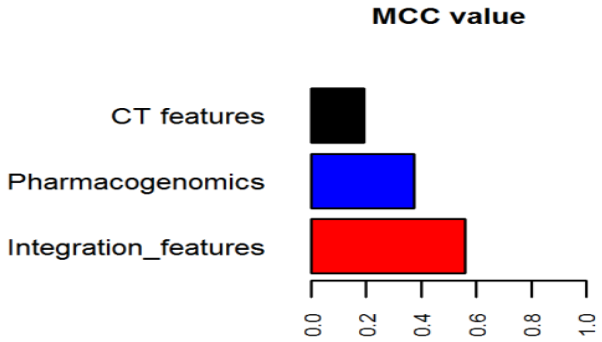
①**It can handle thousands of input variables and identify most significant variables**

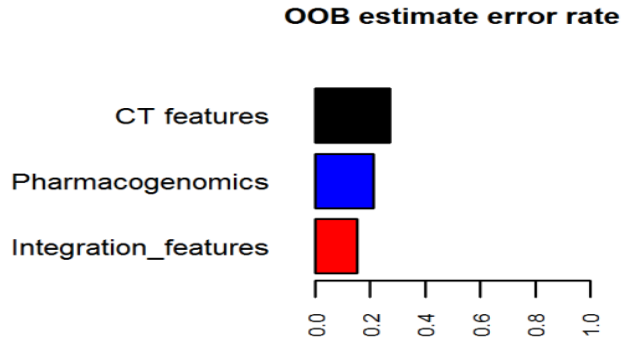
②It has an effective method for **estimating missing data** and maintains accuracy when a large proportion of the data are missing

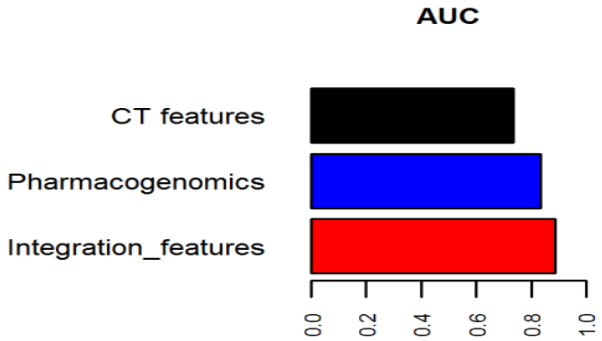
③It has methods for **balancing errors** in data sets where classes are imbalanced

④ It has been implemented in the R package (RandomForest) or python scikit-learn module

Feature optimization







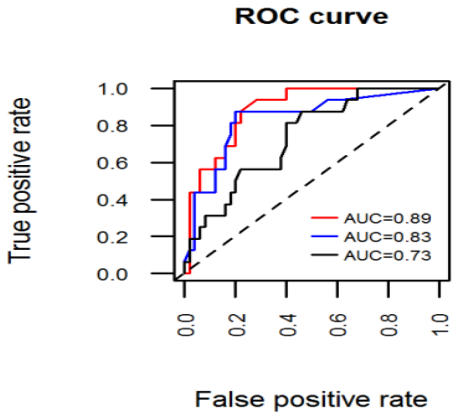


Fig. 4-1 Comparison of different optimal model based on different features combinations

Optimal model based on drug chemical structure and drug target features

Optimal model based on pharmacogenomics features

Optimal model based on drug chemical structure, drug target features and pharmacogenomics features

Optimal model

The feature combinations contributing to the optimal model

Drug target network features

Drug chemical structure

Pharmacogenomics features

Order	Features	Descriptions
1	Sdtps	The average similarity score of KEGG pathways targeted by each drug in a drug combination.
2	Sc	The similarity score of drug chemical structure.
3	commom_up1	The proportion of overlapped DEGs in DEGsA and DEGsB which are both up-regulated by two drugs.
4	opposite1	The proportion of overlapped DEGs in DEGsA and DEGsB which are up-regulated by one drug but down-regulated by the other.
5	common_up2	The proportion of commonly up-regulated DEGs in DEGsA and DEGsB which belong to geneset(GP) as well as DEGsA and DEGsB.
6	opposite2	The proportion of opposite genes in DEGsA and DEGsB which belong to geneset(GP) as well as DEGsA and DEGsB.
7	up_dn_po2	An assessment indictor used to evaluate the positive permutation effect on cancer related pathways by a drug combination.
8	common_dn3	The proportion of commonly down-regulated DEGs in geneset(GP) which belong to geneset(GP) as well as DEGsA and DEGsB.
9	influx_po	The number of up-regulated DEGs which belongs to the drug influx genes set in Table 1.

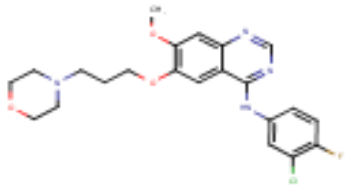
The description of best feature combinations. The combination of pathway similarity, drug chemical structure, differentially expressed genes following drug treatment and the up-regulated influx genes all contributed to SyDRa (synergy of drug based on random forest).

Xiangyi Li, et al. Artificial Intelligence in Medicine, 2017.

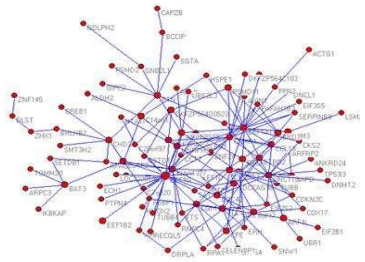
Independent test

- A total of 170 approved anti-cancer drugs were collected from FDA.
- By mapping these drugs to CMap dataset, 17 out of 170 were used to treat MCF7 cell line. the gene expression profiles from CMap database as independent test dataset.
- In total, 187 drug combinations for these 17 drugs. Twenty-eight were predicted as synergistic drug combinations

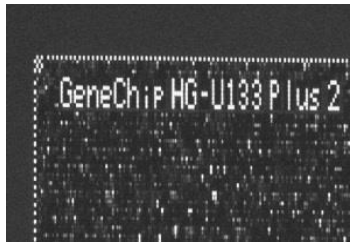
Test dataset



Chemical structure



Target network



Gene chip data (Connectivity Map)

①

②

③

Test dataset

Order	Drug Name
1	azacitidine
2	carmustine
3	dacarbazine
4	decitabine
5	exemestane
6	flutamide
7	imatinib
8	letrozole
9	mercaptopurine
10	methotrexate
11	nilutamide
12	paclitaxel
13	streptozocin
14	tamoxifen
15	thalidomide
16	trifluridine
17	vorinostat

④

	Drug 1	Drug 2	Label
1	decitabine	paclitaxel	effective
2	exemestane	tamoxifen	effective
3	imatinib	vorinostat	effective
4	azacitidine	thalidomide	effective
5	carmustine	streptozocin	effective
6	imatinib	paclitaxel	effective
7	carmustine	tamoxifen	?
…	…	…	
122	exemestane	mercaptopurine	?
123	exemestane	streptozocin	?
124	imatinib	paclitaxel	?
125	letrozole	streptozocin	?
126	letrozole	thalidomide	?
127	letrozole	thalidomide	?
128	mercaptopurine	paclitaxel	?
129	mercaptopurine	tamoxifen	?
130	mercaptopurine	trifluridine	?
131	methotrexate	nilutamide	?
132	methotrexate	streptozocin	?
133	methotrexate	trifluridine	?
134	nilutamide	trifluridine	?
135	tamoxifen	trifluridine	?
136	thalidomide	trifluridine	?

Result

Table 4-2 Potential synergistic drug combinations predicted by optimal model

Drug 1	Concentration(M)	Drug 2	Concentration(M)
azacitidine	1.64E-05	letrozole	1.40E-05
azacitidine	1.64E-05	nilutamide	1.26E-05
azacitidine	1.64E-05	thalidomide	1.54E-05
carmustine	0.0001	imatinib	1.00E-05
carmustine	0.0001	letrozole	1.40E-05
carmustine	0.0001	streptozocin	1.50E-05
carmustine	0.0001	tamoxifen	1.00E-06
decitabine	1.00E-07	exemestane	1.00E-08
decitabine	1.00E-07	imatinib	1.00E-05
decitabine	1.00E-07	letrozole	1.40E-05
decitabine	1.00E-07	mercaptopurine	0.0001
decitabine	1.00E-07	nilutamide	1.26E-05
exemestane	1.00E-08	imatinib	1.00E-05
exemestane	1.00E-08	mercaptopurine	0.0001
exemestane	1.00E-08	streptozocin	1.50E-05
imatinib	1.00E-05	paclitaxel	1.00E-07
letrozole	1.40E-05	streptozocin	1.50E-05
letrozole	1.40E-05	thalidomide	0.0001
letrozole	1.40E-05	thalidomide	1.54E-05
mercaptopurine	0.0001	paclitaxel	1.00E-07
mercaptopurine	0.0001	tamoxifen	1.00E-06
mercaptopurine	0.0001	trifluridine	1.36E-05
methotrexate	8.80E-06	nilutamide	1.26E-05
methotrexate	8.80E-06	streptozocin	1.50E-05
methotrexate	8.80E-06	trifluridine	1.36E-05
nilutamide	1.26E-05	trifluridine	1.36E-05
tamoxifen	1.00E-06	trifluridine	1.36E-05
thalidomide	1.54E-05	trifluridine	1.36E-05

28 pairs
(/187)

Myelodysplastic syndromes and acute myeloid leukemia

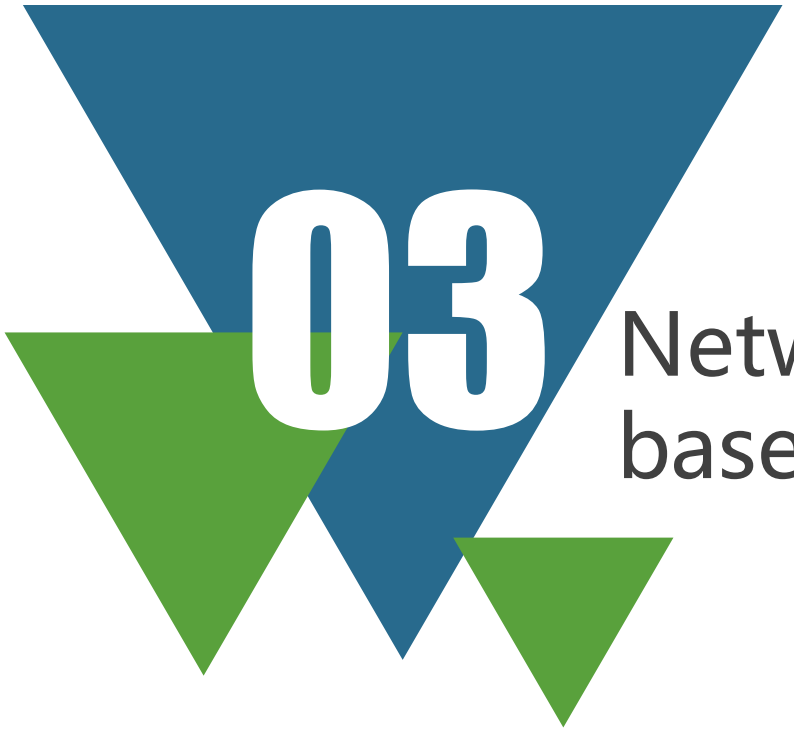
Advanced cancer

Advanced or metastatic solid tumor

Raza A et al. Cancer 2008, 113(7):1596-1604.
Micetich K et al. Journal of the National Cancer Institute 1992, 84(4):256-261.
Pishvaian MJ et al. Cancer chemotherapy and pharmacology 2012, 70(6):843-853.

Summary (I)

- **Based on single drug treated cell line expression profiling, a prediction model for drug synergy was constructed**
- **Both drug phenotype data (i.e. drug chemical structure and drug target-related information) and pharmacogenomics contribute to drug synergism; drug metabolic process may play an important role in drug synergism**
- **The model could be used to predict other drug synergy based on single drug cell line treatment data (preclinical modeling)**



Network medicine annotation based on multi-level omics data

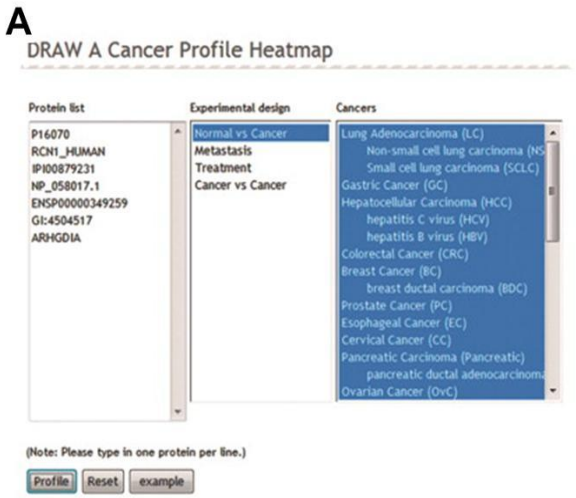
Network medicine annotation workflow

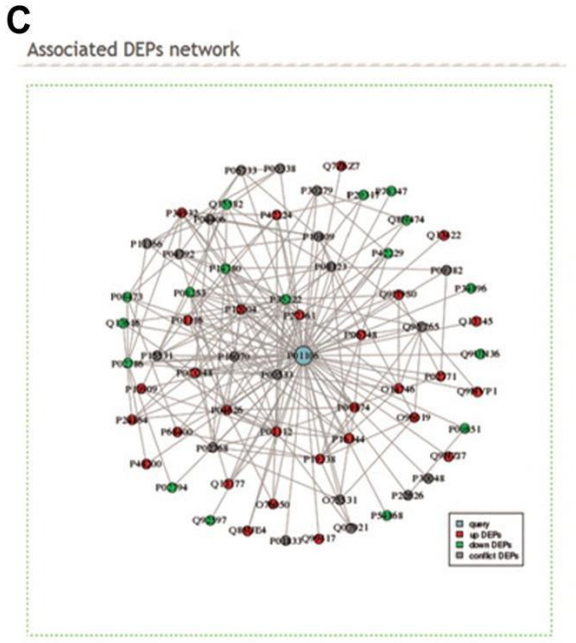
Background network construction

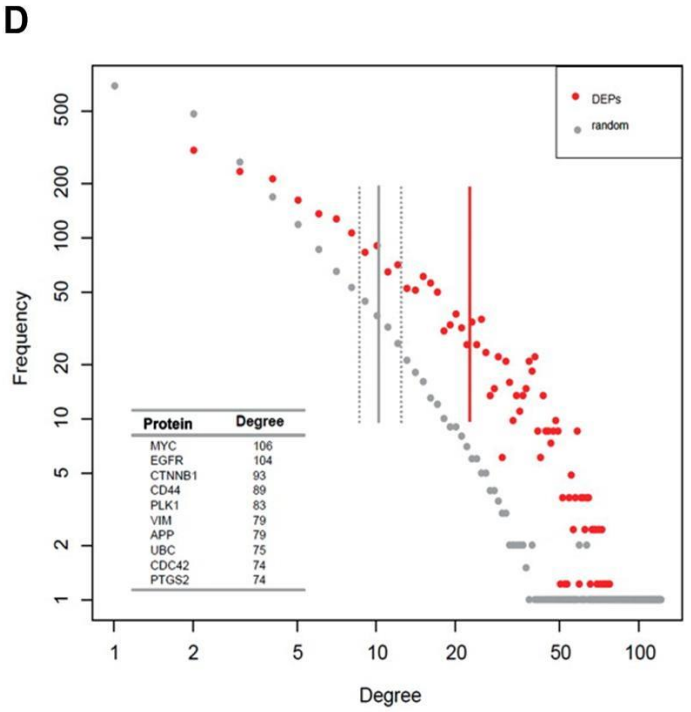
Network omics data annotation

Network medicine annotation

Single omics data annotation for cancer: Proteins differentially expressed in cancer-based on MS proteomics data

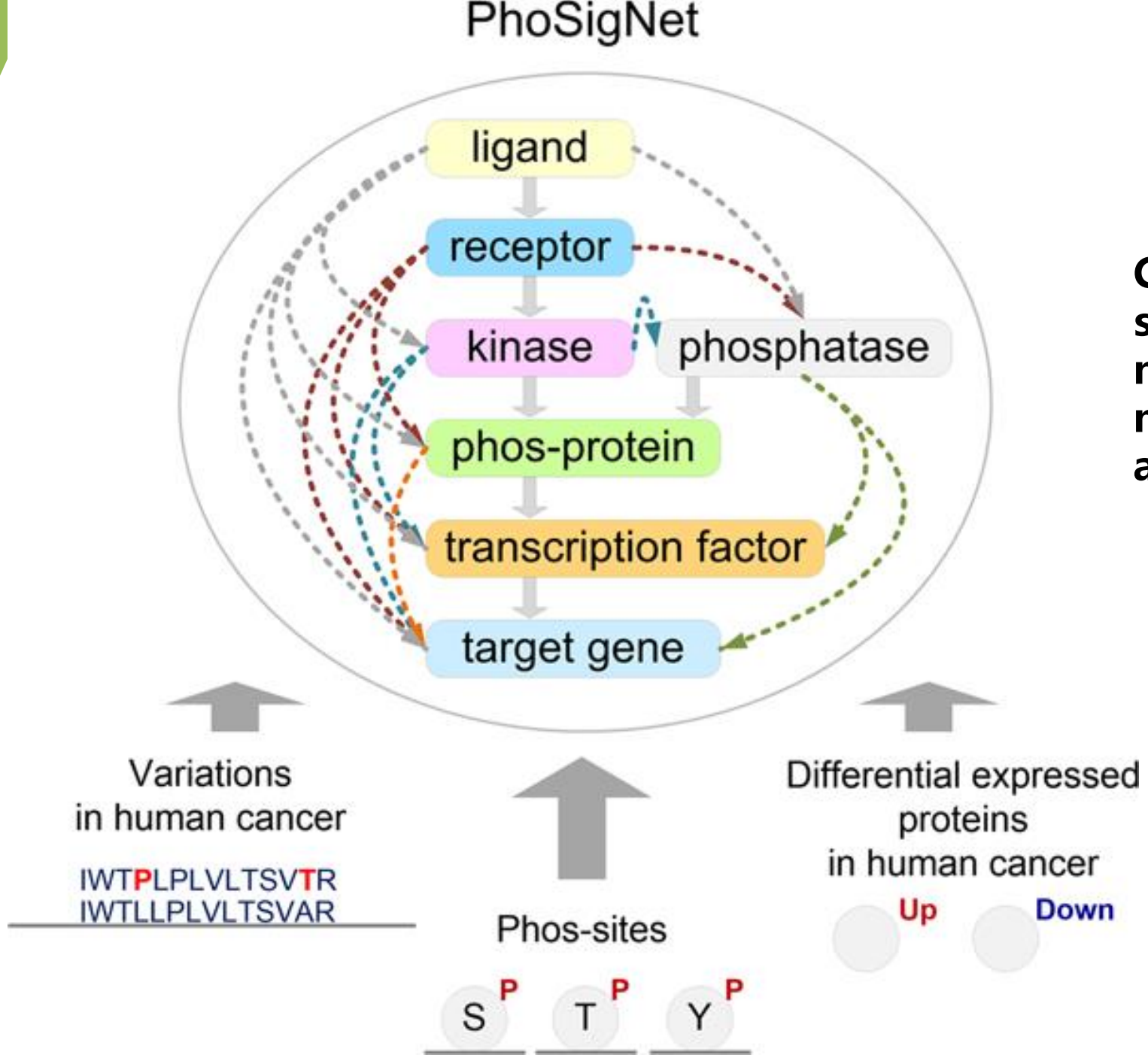






Hong Li, et al. dbDEPC. *NAR,* 2009

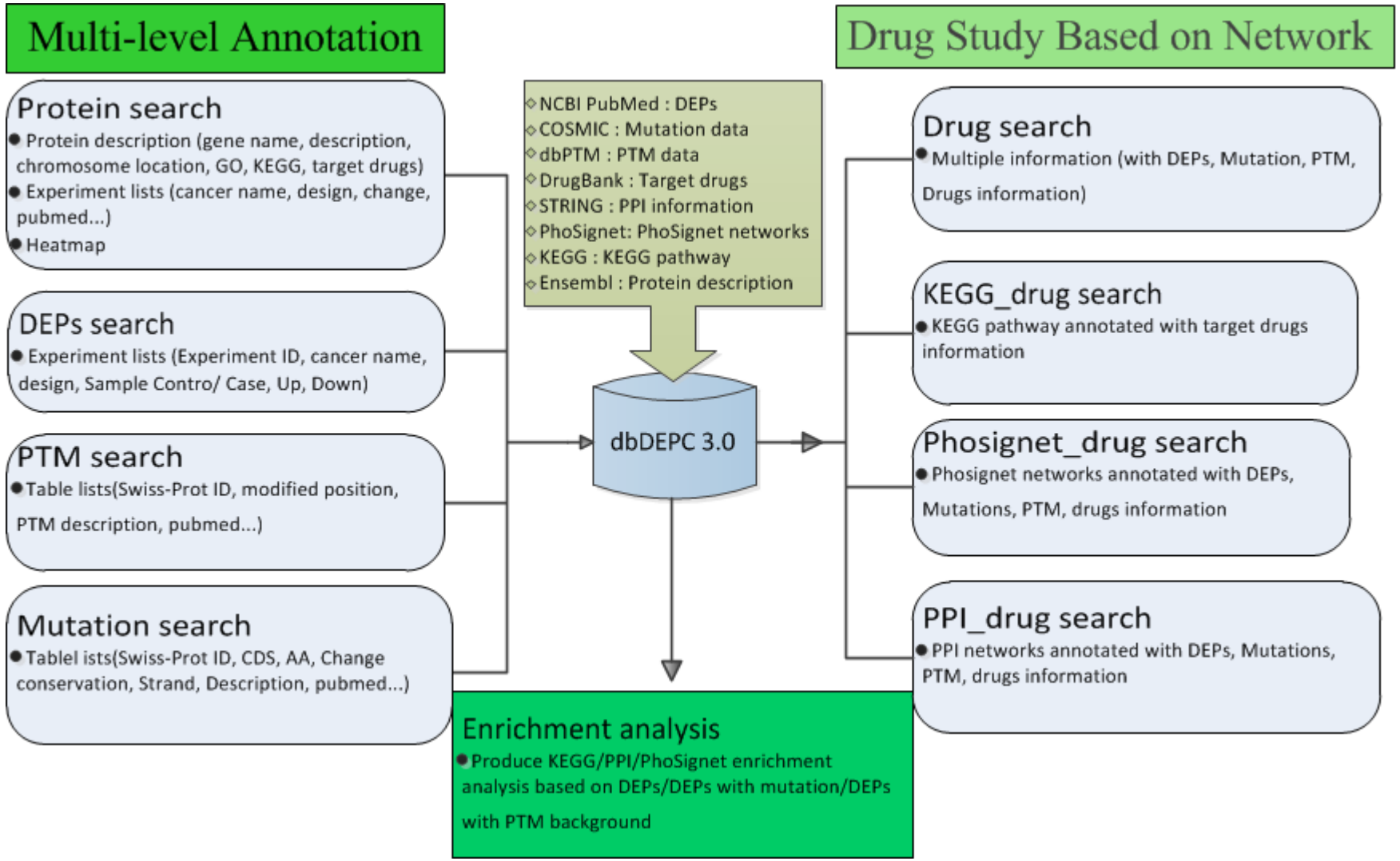
Ying He, et al. dbDEPC 2.0: updated database of differentially expressed proteins in human cancers. *NAR ,* 2012



Construction of background signaling transduction network;
multi-level omics data annotation

Menghuan Zhang, et al. Construction and deciphering of human phosphorylation mediated signaling transduction network. *J of Proteome Res.* 2015

Network medicine annotation based on multi-level omics data



Qingmin Yang, et al. dbDEPC 3.0, 2017, unpublished

dbDEPC_2.0:

- Protein:4029
- Cancer(Subtybe):20(18)
- MS Experiment:331
- Literature:241

dbDEPC_3.0:

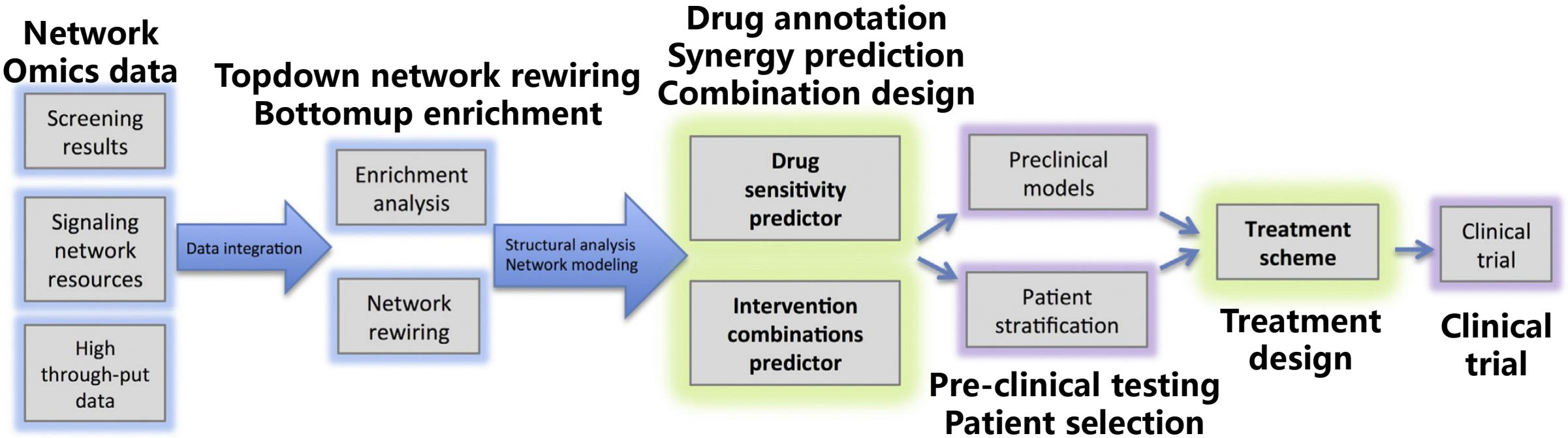
- Protein:11669
- Cancer(Subtybe):26(28)
- MS Experiment:779
- Literature:623

In this updated version, dbDEPC 3.0 has expanded to over 11000 protein entries, curated from 779 experiments across 26 types of human cancers

Unpublished

Summary (III)

- To make network medicine annotation, several resources should be collected: network structure, drug annotation, disease data
- Top-down annotation: known pathways, network, annotation; bottom-up annotation: enrichment pathway, network, annotation
- The pathway or network medicine annotation could be theoretically mapped to individual patient
- Clinical trials to follow or to design



Biochemical and Biophysical Research Communications 464 (2015) 386e391 Network-based approaches for drug response prediction and targeted therapy development in cancer

Acknowledgement

Current members

- LI Xiang-yi, M.S.
- CUI Hui, Ph.D. candidate
- YANG Qing-min, MS candidate
- ZHANG Meng-huan, Ph.D

Previous members

- LI Hong, Ph.D
- HE Ying, Ph.D

THANKS