MICHIGAN STATE UNIVERSITY

ChenLab
Michigan State University
http://binchenlab.org

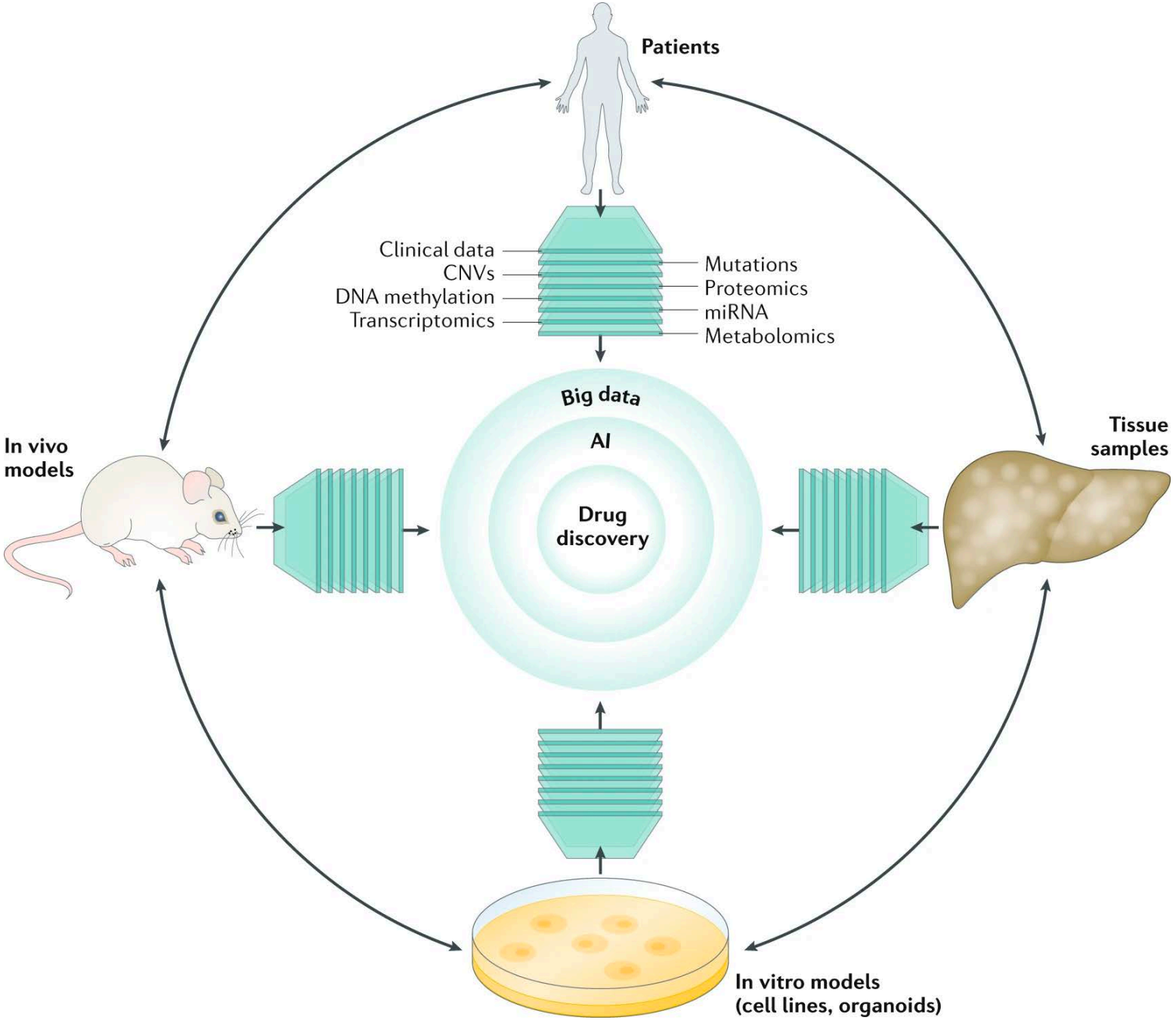
from big data to drug discovery

Bin Chen

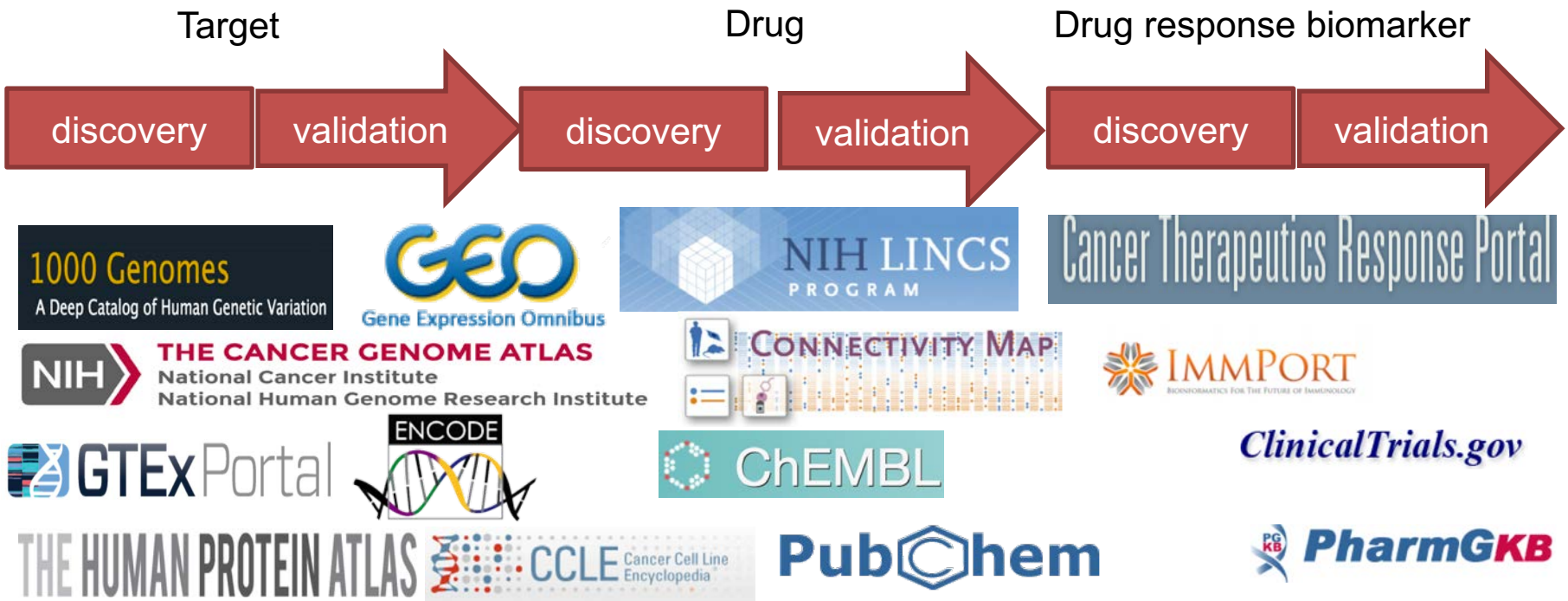
Assistant Professor
Dept. of Pediatrics and Human Development
Dept. of Pharmacology and Toxicology
College of Human Medicine
Michigan State University

Bin.Chen@hc.msu.edu @DrBinChen

http://binchenlab.org



Chen B, Nat Rev Gastroenterol Hepatol. 2020

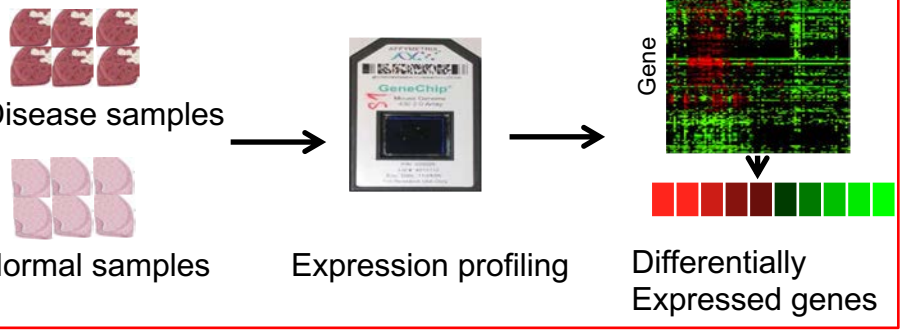


Bin Chen, Clin Pharmacol Ther. 2016

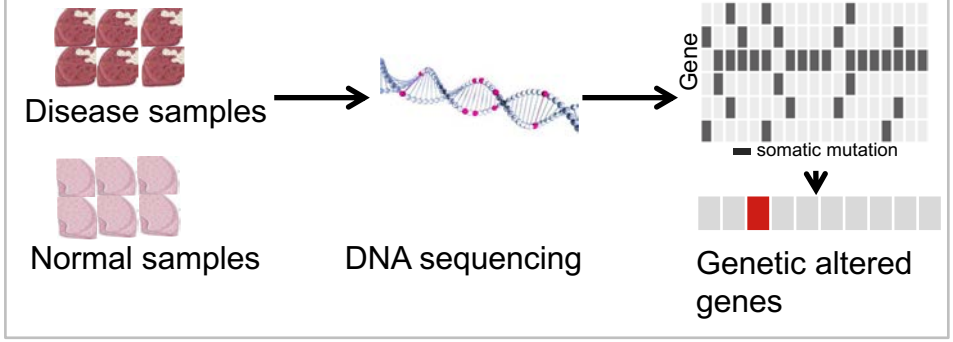
Target Discovery Using Big Data

→ **Experimental Validation**

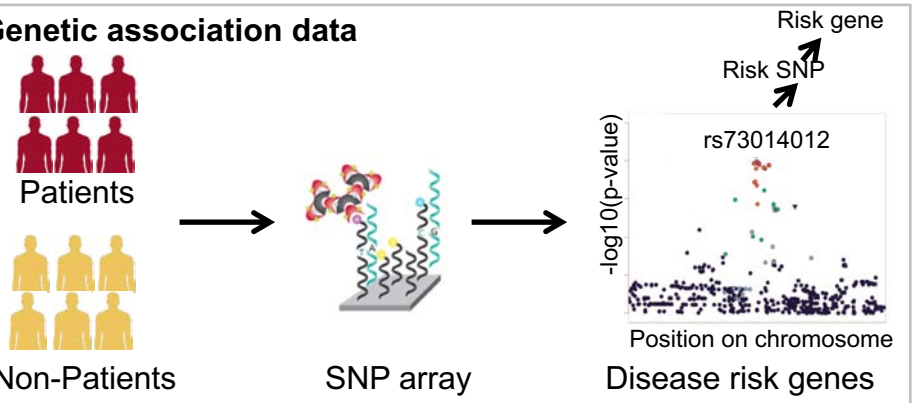
Gene expression data



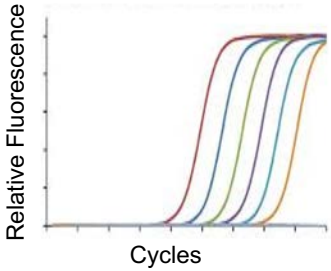
Somatic mutation data



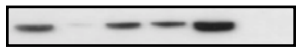
Genetic association data



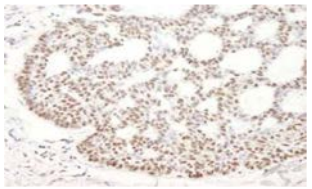
Expression validation
(*in vitro*, *in vivo*)



mRNA expression
using RT-PCR

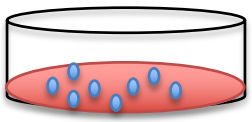


Protein expression
using western blot



Protein expression and location
using Immunohistochemistry

Functional validation
(*in vitro*, *in vivo*)

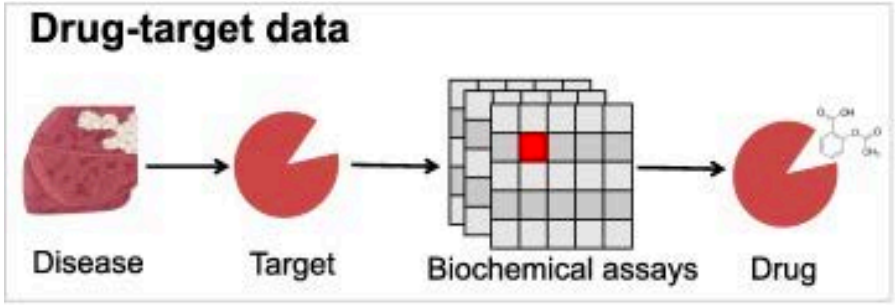


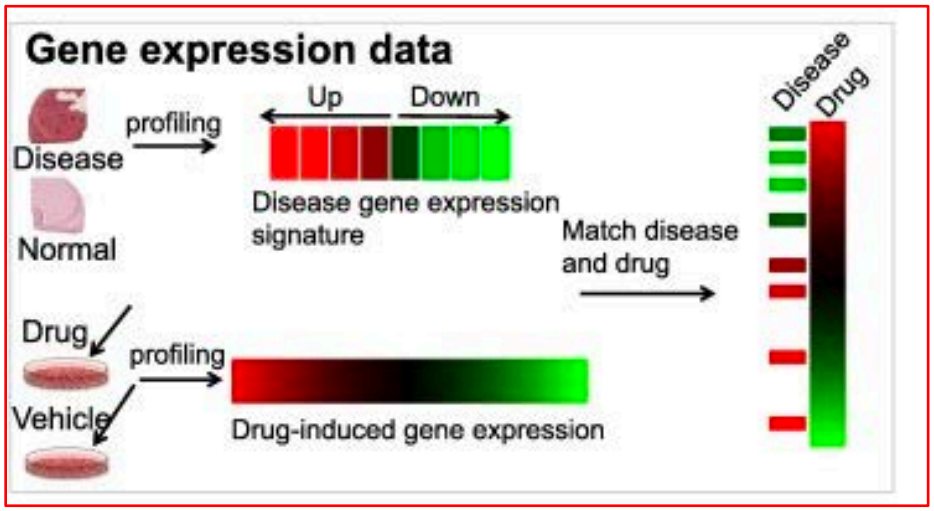
Cell viability after loss of gene
function *in vitro*

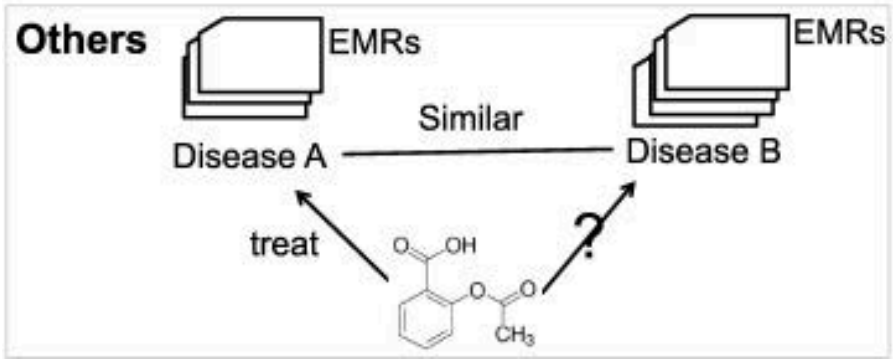
Tumor growth after loss of
gene function *in vivo*

Bin Chen, Clin Pharmacol Ther. 2016

Drug Discovery Using Big Data → Experimental Validation

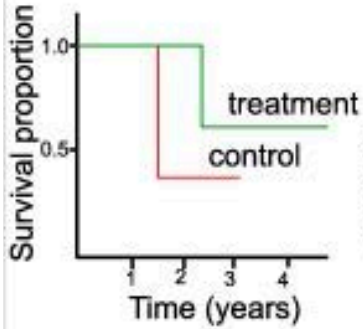






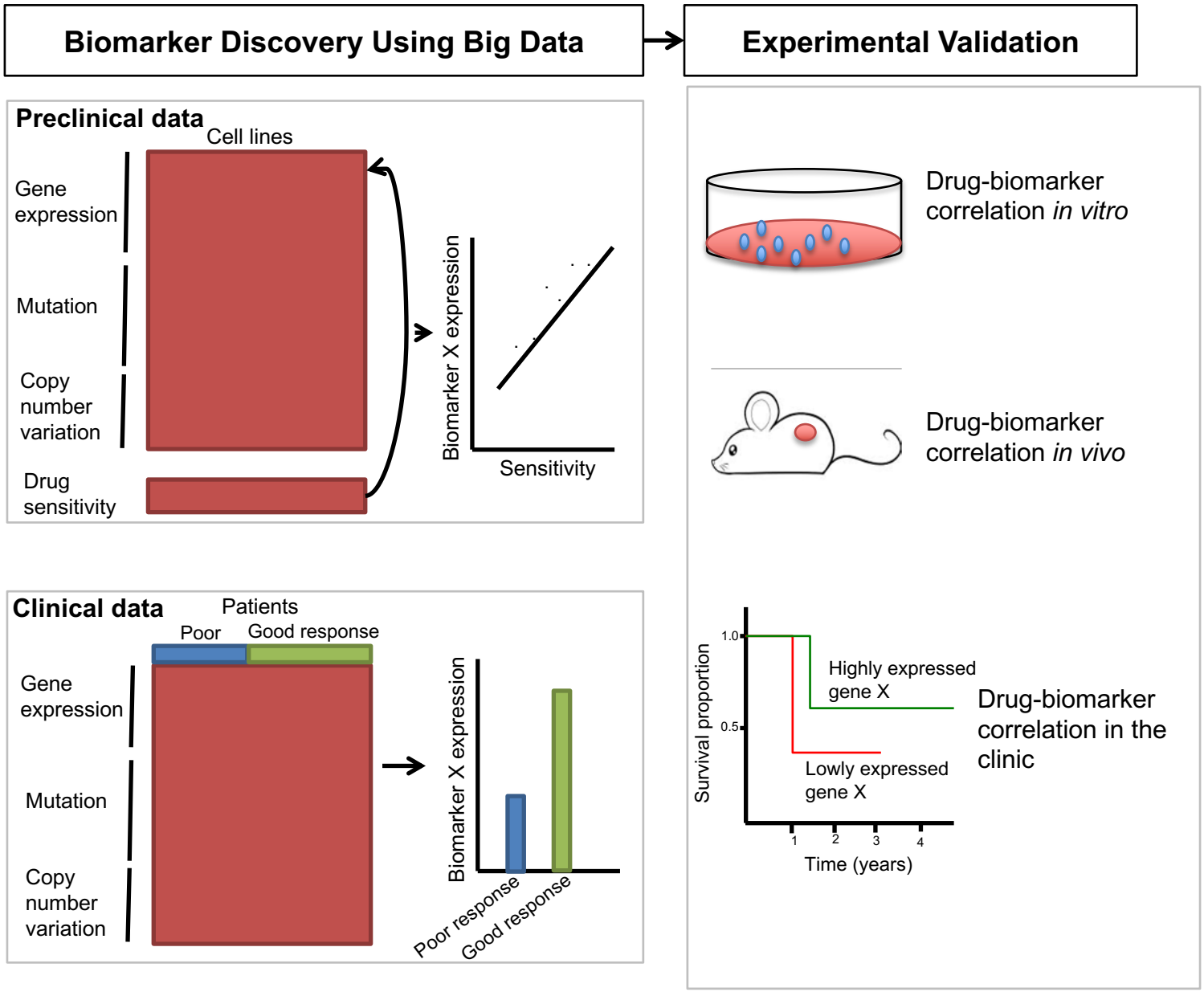
Cell viability after drug treatment *in vitro*

Tumor growth after drug treatment *in vivo*



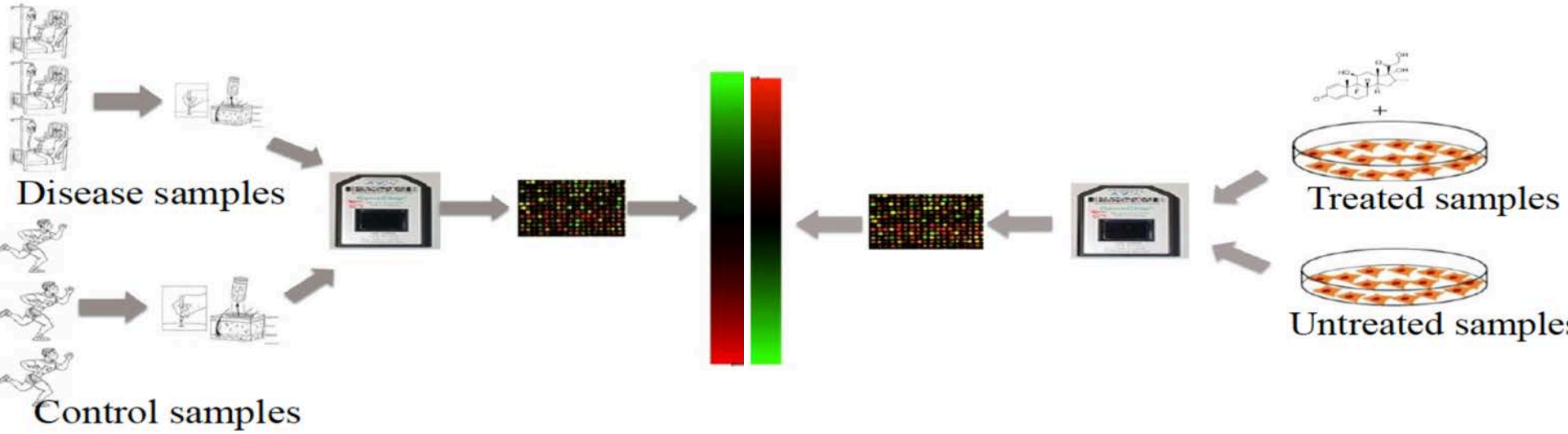
Survival analysis of drug treatment in the clinic

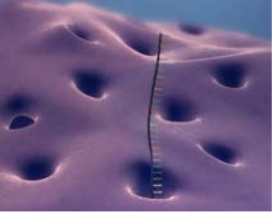
Bin Chen, Clin Pharmacol Ther. 2016

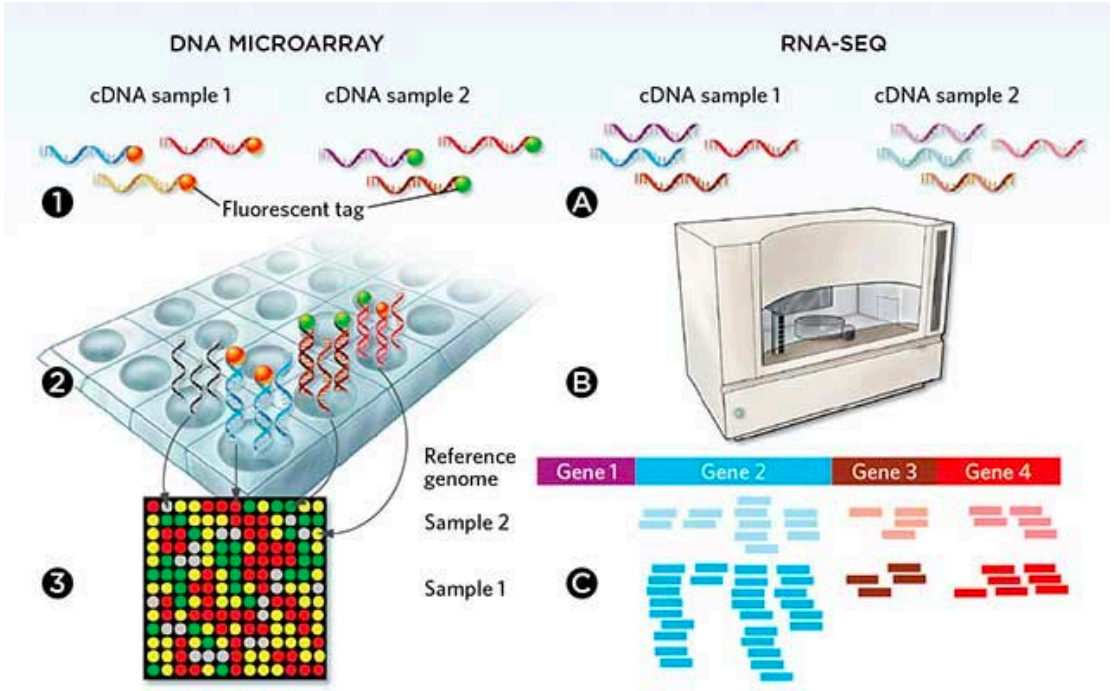


Bin Chen, Clin Pharmacol Ther. 2016

System-based drug discovery







~20,000 patient samples

60K genes

Gene Expression (numeric)

1210 cell lines

60K genes

Gene Expression (numeric)

https://www.the-scientist.com/lab-tools/an-array-of-options-35381

GEO (Gene Expression Omnibus)

GEO DataSets

GEO DataSet | breast cancer

Save search Advanced

Show additional filters

Display Settings: Summary, 20 per page, Sorted by Default order

Send to:

Entry type

DataSets (141)

Series (1626)

Samples (40631)

Platforms (35)

Organism

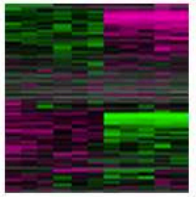
Select ...

Study type

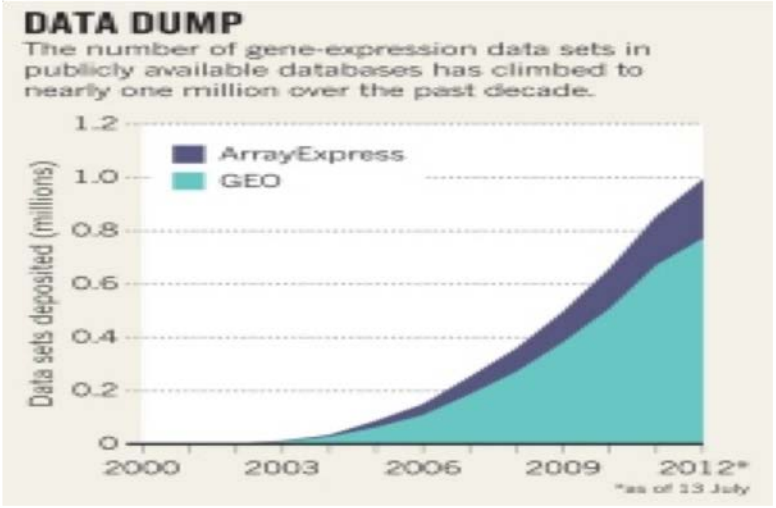
Results: 1 to 20 of 42433

<< First < Prev Page 1 of 2122 Next > Last >>

1. Leukemia inhibitory factor effect on Sin3a-silenced MCF7 **breast cancer** cell line



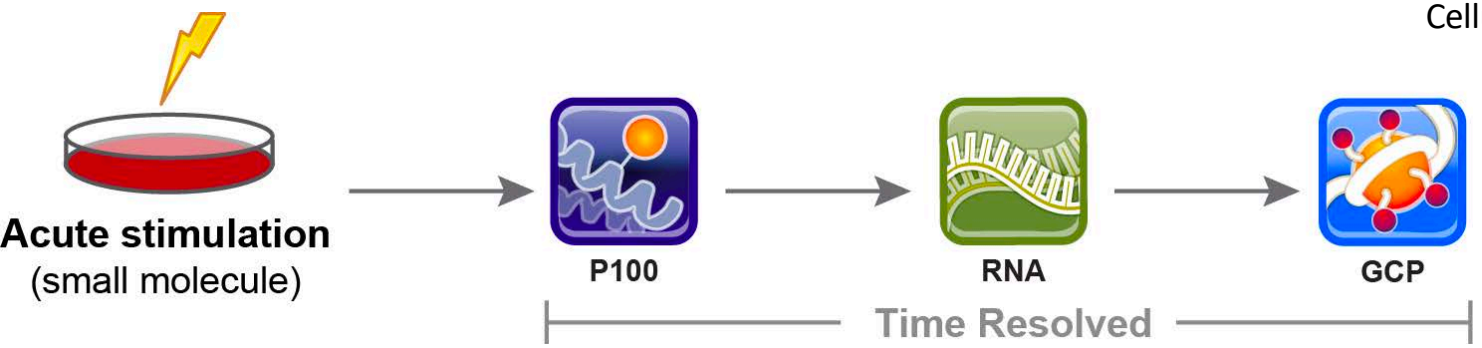
Analysis of SIN3 transcription regulator homolog A (Sin3a)-depleted MCF7 cells stimulated with LIF cytokine to activate signal transducer and activator of transcription 3 (STAT3). STAT3 transcription factor is a potent oncogene. Results provide insight into role of Sin3a in mediating STAT3 activity

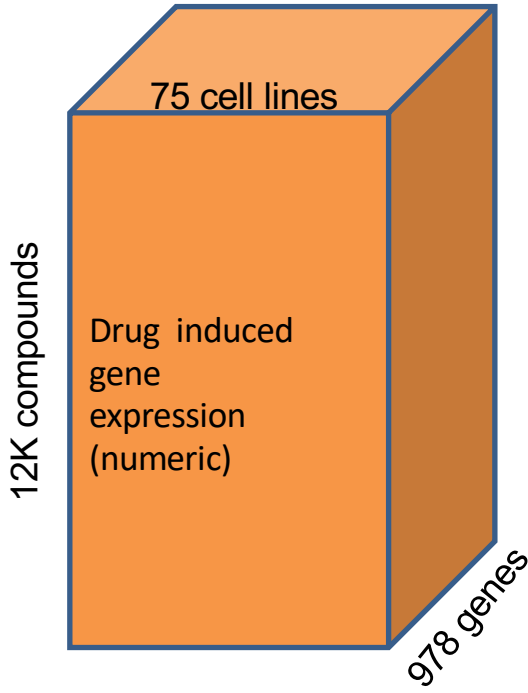


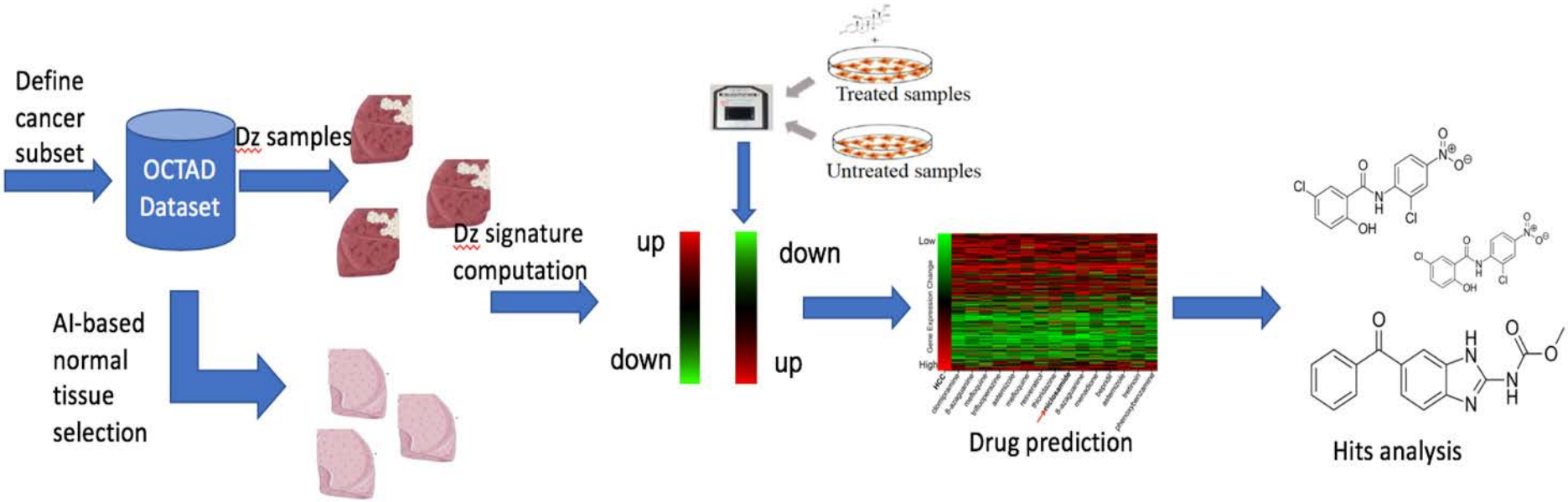
Repository Browser

DataSets:	4348
Series:	95960
Platforms:	18242
Samples:	2425625

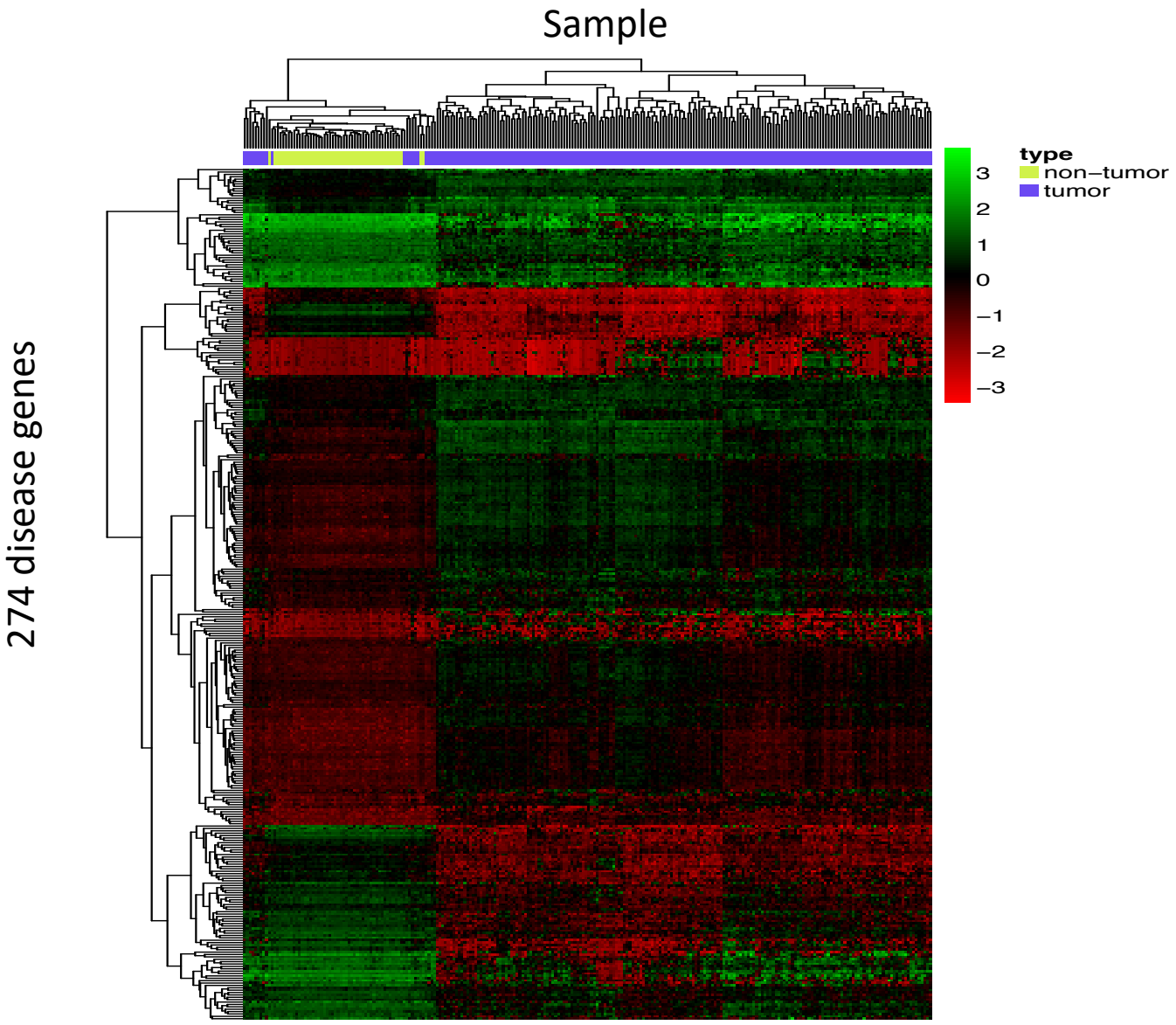
Cellular response/gene expression



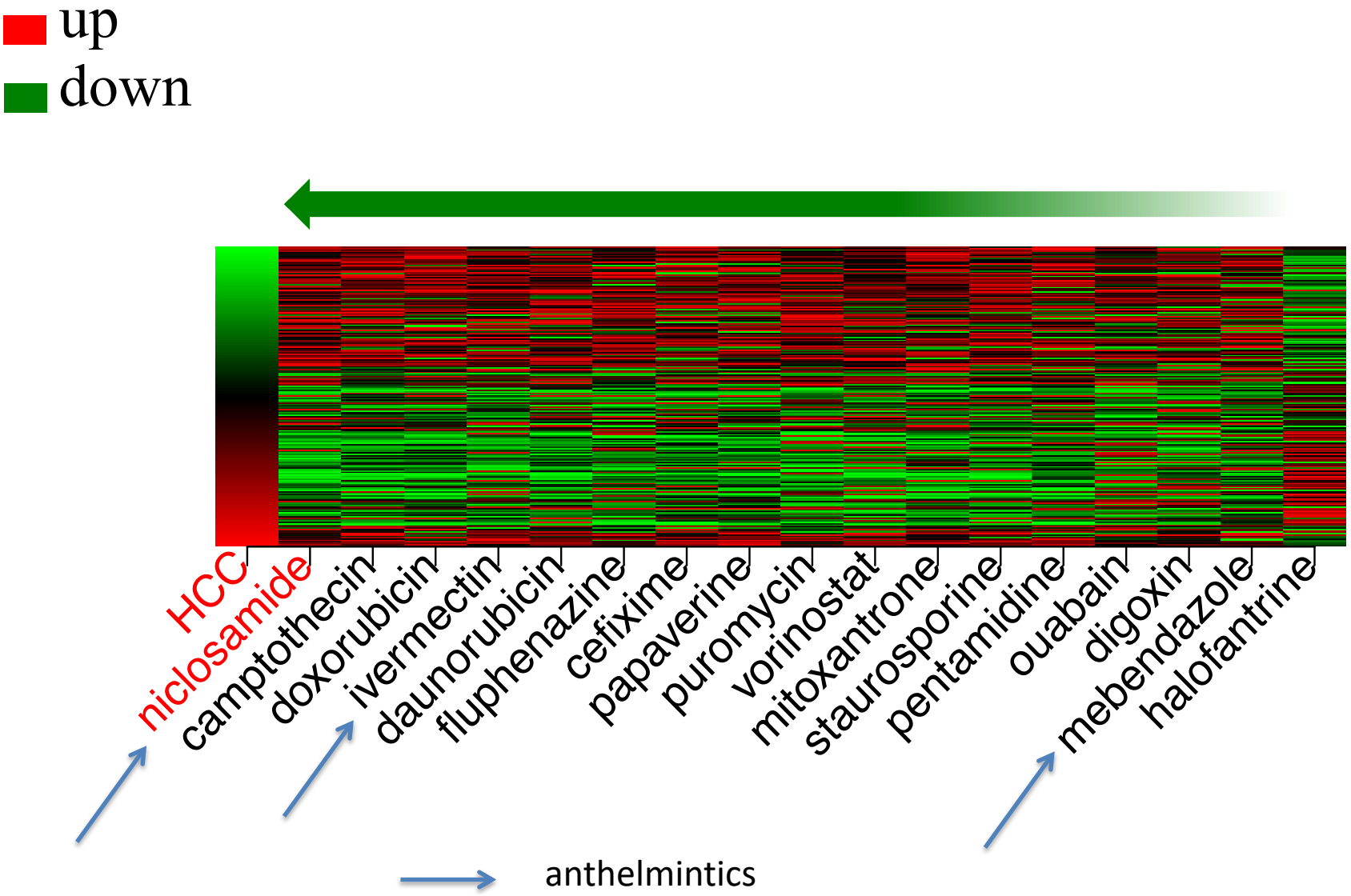



Define cancer subset
OCTAD Dataset
Dz samples
AI-based normal tissue selection
Dz signature computation
up
down
down
up
Treated samples
Untreated samples
Low
Gene Expression Change
High
HCC
Drug prediction
Hits analysis

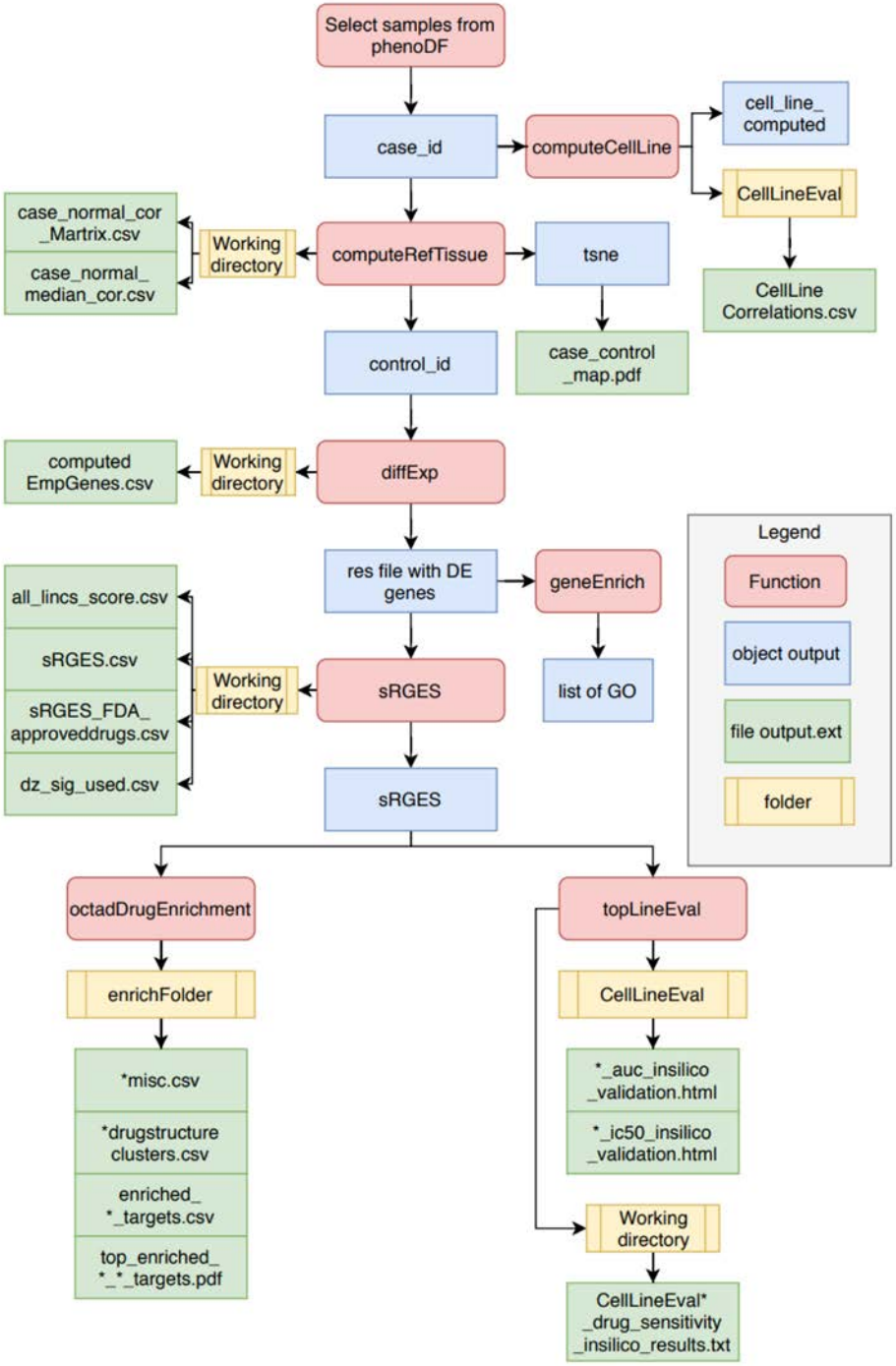
HCC disease gene expression signature



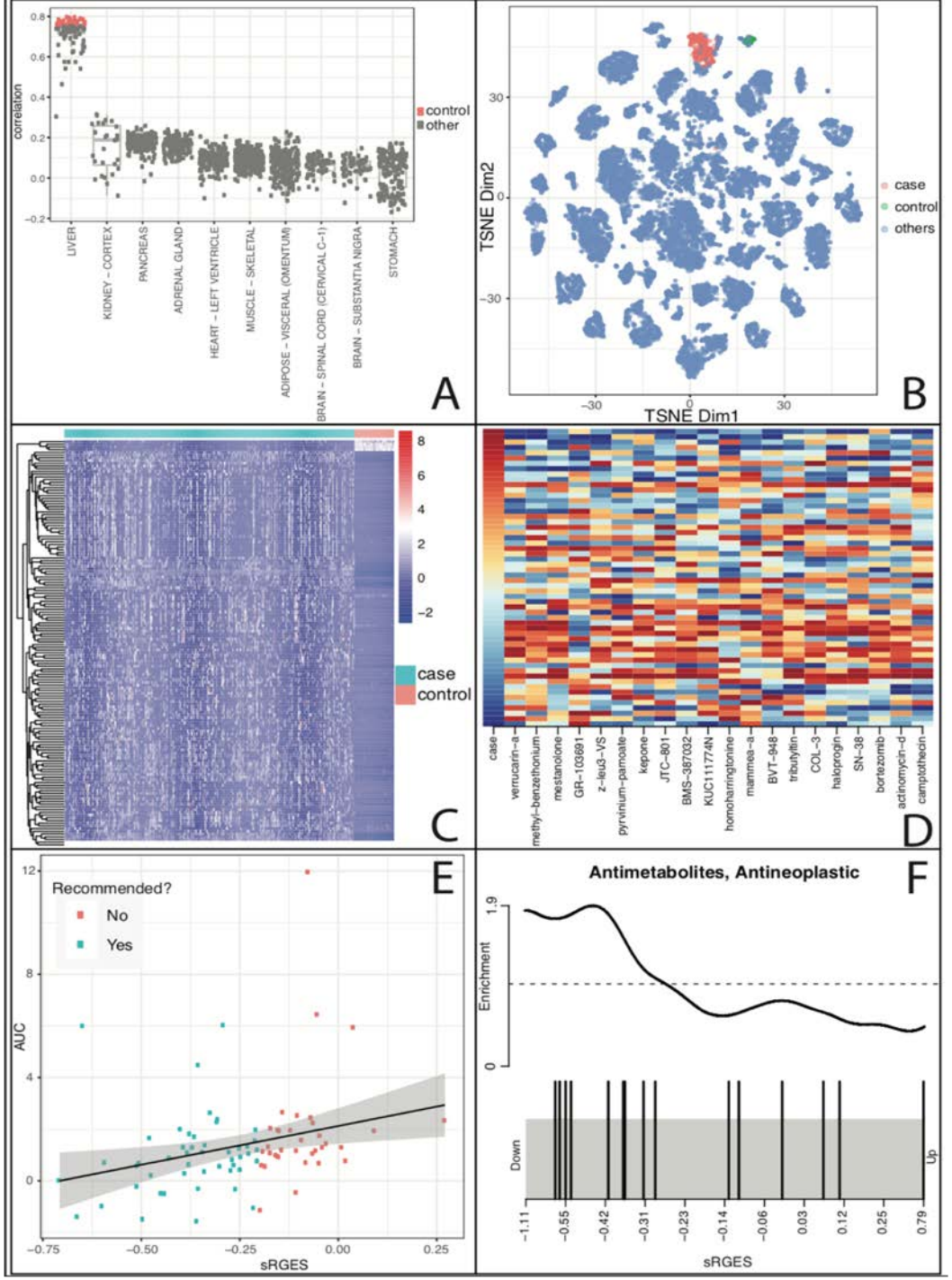
Drugs that reverse HCC gene expression signature



Chen, Gastroenterology, 2017



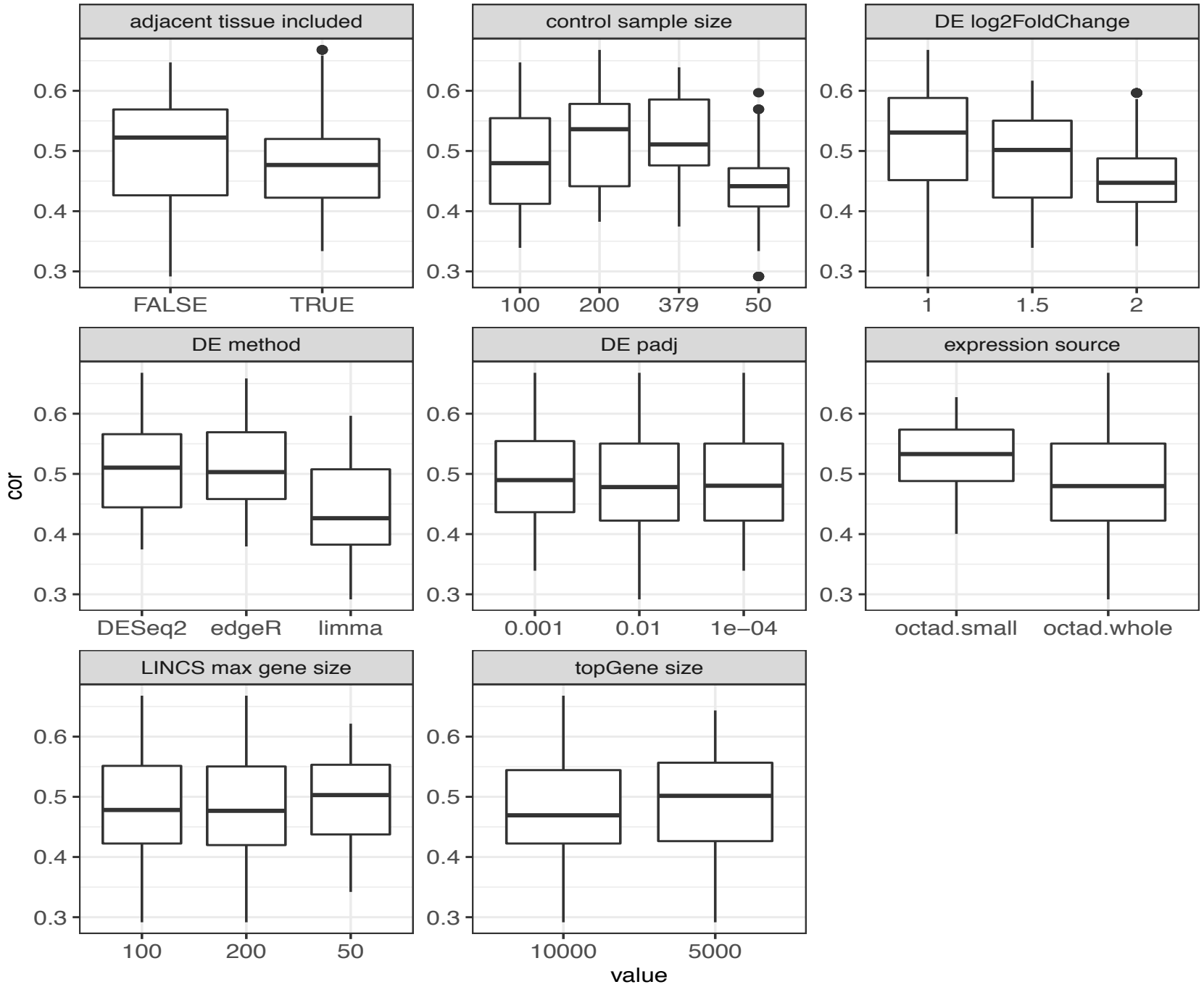
http://octad.org/



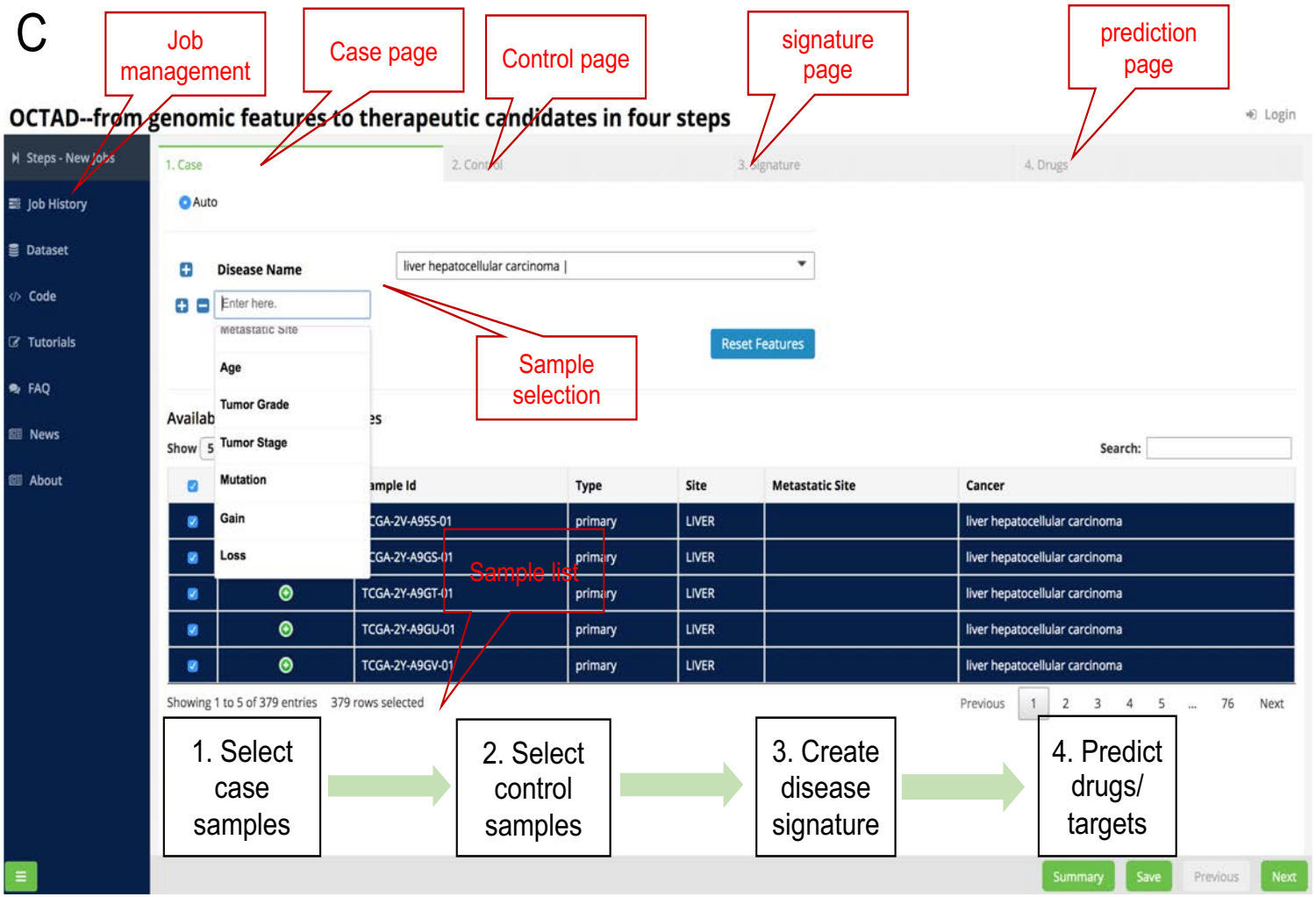
http://octad.org/

OCTAD R package

```
case_id=subset(phenoDF,cancer=='liver hepatocellular carcinoma'&sample.type == 'primary', select = c("sample.id")


HCC_adjacent=subset(phenoDF,cancer=='liver hepatocellular carcinoma'&sample.type == 'adjacent'&data.source == 'TCGA', select = c("sample.id")


res=diffExp(case_id,control_id,source='octad.whole',output=T,n_topGenes=10000,file ='octad.counts.and.tpm.h5')


sRGES=runsRGES(res,max_gene_size=500,permutations=10000)


head(sRGES)
```

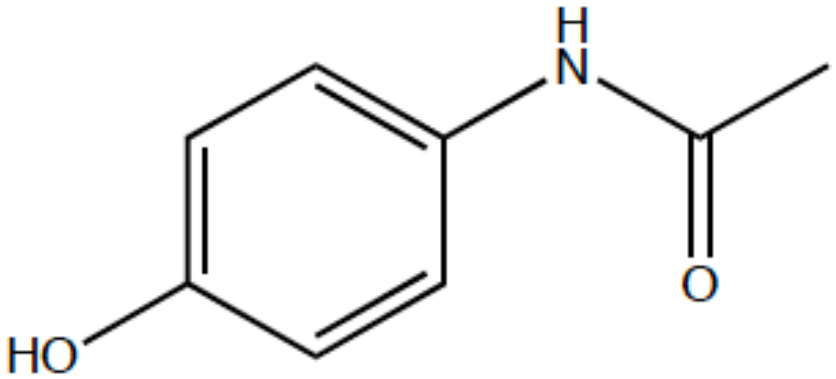

adjacent tissue included
FALSE
TRUE
control sample size
100
200
379
50
DE log2FoldChange
1
1.5
2
DE method
DESeq2
edgeR
limma
DE padj
0.001
0.01
1e−04
expression source
octad.small
octad.whole
LINCS max gene size
100
200
50
topGene size
10000
5000
cor
value
0.3
0.4
0.5
0.6

OCTAD Web portal



#Billy Zeng, #Benjamin S. Glicksberg, #Patrick Newbury, #Evgenii Chekalin, Jing Xing, Ke Liu, Anita Wen, Caven Chow, Bin Chen, OCTAD: an open workplace for virtually screening therapeutics targeting precise cancer patient groups using gene expression features accepted, Nature Protocols

Drug/Chemical compounds



SMILES: **c1c(O)ccc(NC(=O)C)c1**

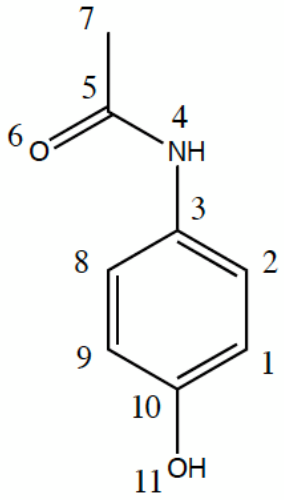
Acetaminophen

(Tylenol / Paracetamol)

https://djwild.pages.iu.edu/ic/

2D- compound representation in computer

Atom Number	Atom Type
1	C
2	C
3	C
4	N
5	C
6	O
7	C
8	C
9	C
10	C
11	O

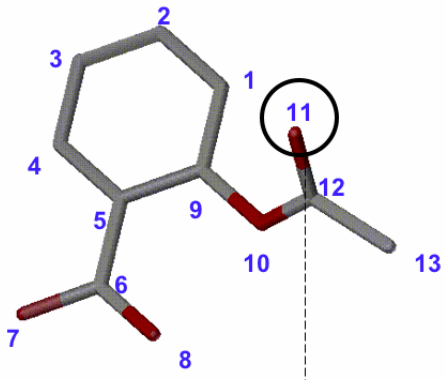


	1	2	3	4	5	6	7	8	9	10	11
1	0	1	0	0	0	0	0	0	0	2	0
2	**1**	0	2	0	0	0	0	0	0	0	0
3	**0**	**2**	0	1	0	0	0	1	0	0	0
4	**0**	**0**	**1**	0	1	0	0	0	0	0	0
5	**0**	**0**	**0**	**1**	0	2	1	0	0	0	0
6	**0**	**0**	**0**	**0**	**2**	0	0	0	0	0	0
7	**0**	**0**	**0**	**0**	**1**	**0**	0	0	0	0	0
8	**0**	**0**	**1**	**0**	**0**	**0**	**0**	0	2	0	0
9	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**2**	0	1	0
10	**2**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	0	1
11	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	0

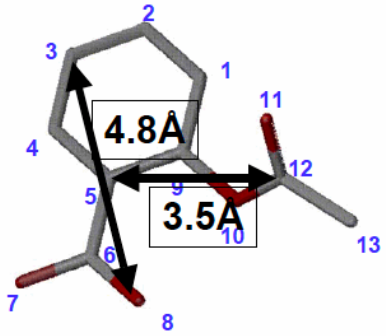
https://djwild.pages.iu.edu/ic/

connection table/ MOL/SDF

3D- compound representation in computer



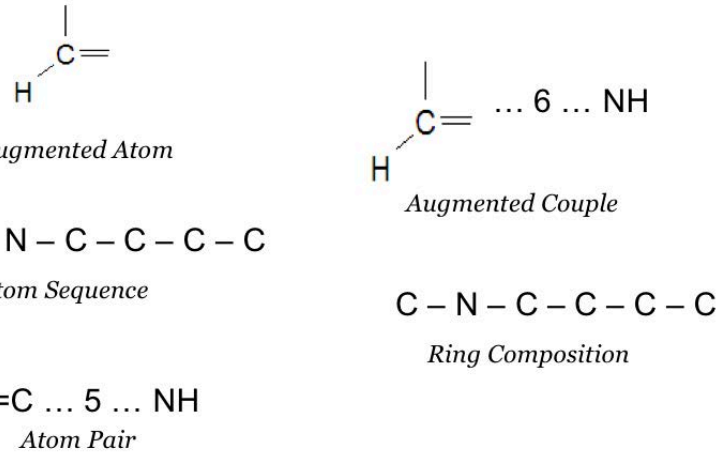
Atom	Label	X	Y	Z
1	C	-1.8920	-0.9920	-1.5760
2	C	-1.3680	-2.1480	-0.9880
3	C	-0.0760	-2.1440	-0.4640
4	C	0.7080	-0.9840	-0.5200
5	C	0.2000	-0.1560	-1.1960
6	C	-0.1080	0.1600	-1.6520
7	O	2.0840	-1.0280	0.1040
8	O	2.5320	-2.0320	0.6360
9	C	2.8760	0.0240	0.1120
10	O	0.7520	1.3320	-1.0840
11	O	0.6680	2.0240	0.0320
12	C	1.3000	3.0600	0.1520
13	C	-0.2400	1.5760	1.4440

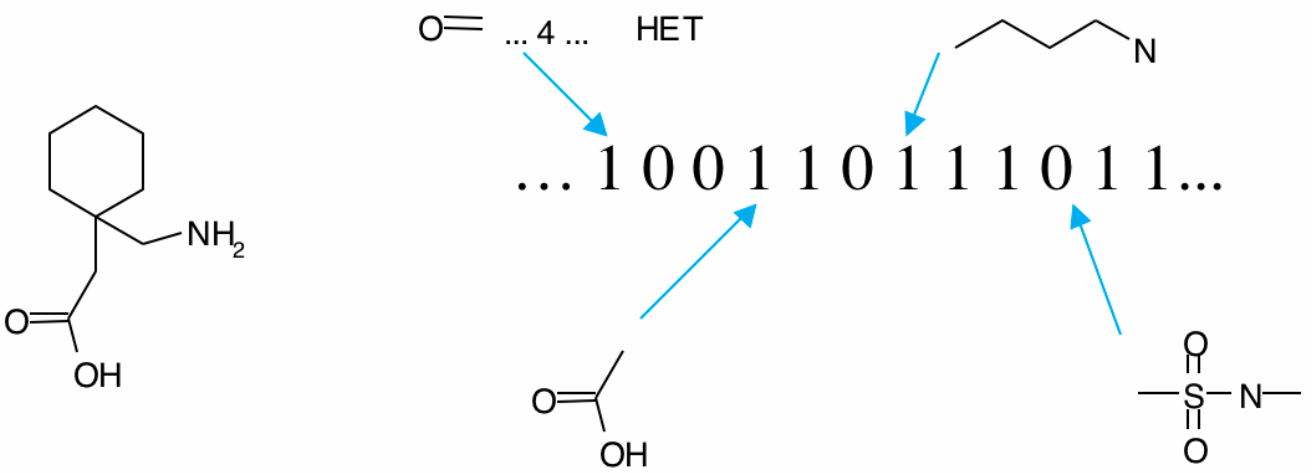


	1	2	3	4	5	6	7	8	9	10	11	12	13
1		1.4	2.4	2.8	2.4	3.8	4.8	4.2	1.4	2.4	2.7	2.9	4.3
2			1.4	2.4	2.8	4.3	5.1	5.0	2.4	3.7	3.9	4.2	5.6
3				1.4	2.4	3.8	4.2	4.8	2.8	4.2	4.7	4.9	6.4
4					1.4	2.5	2.8	3.6	2.4	3.7	4.7	4.6	6.1
5						1.5	2.4	2.3	1.4	2.3	3.7	3.5	4.8
6							1.3	1.2	2.5	2.8	4.4	3.9	5.0
7								2.2	3.7	4.1	5.7	5.2	6.3
8									2.8	2.5	4.2	3.5	4.3
9										1.4	2.6	2.3	3.7
10											2.2	1.3	2.5
11												1.2	2.4
12													1.5
13													

https://djwild.pages.iu.edu/ic/

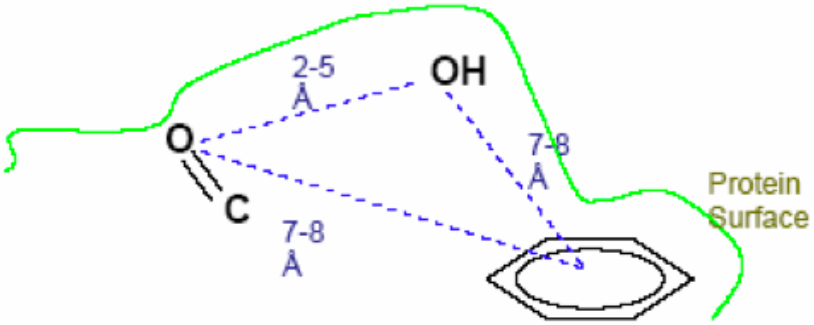
molecular fingerprint

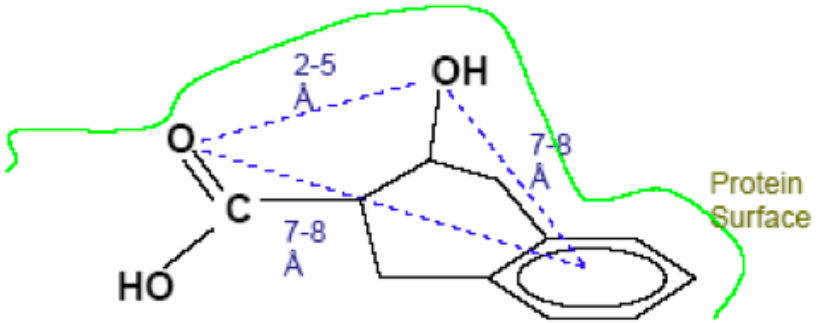




https://djwild.pages.iu.edu/ic/

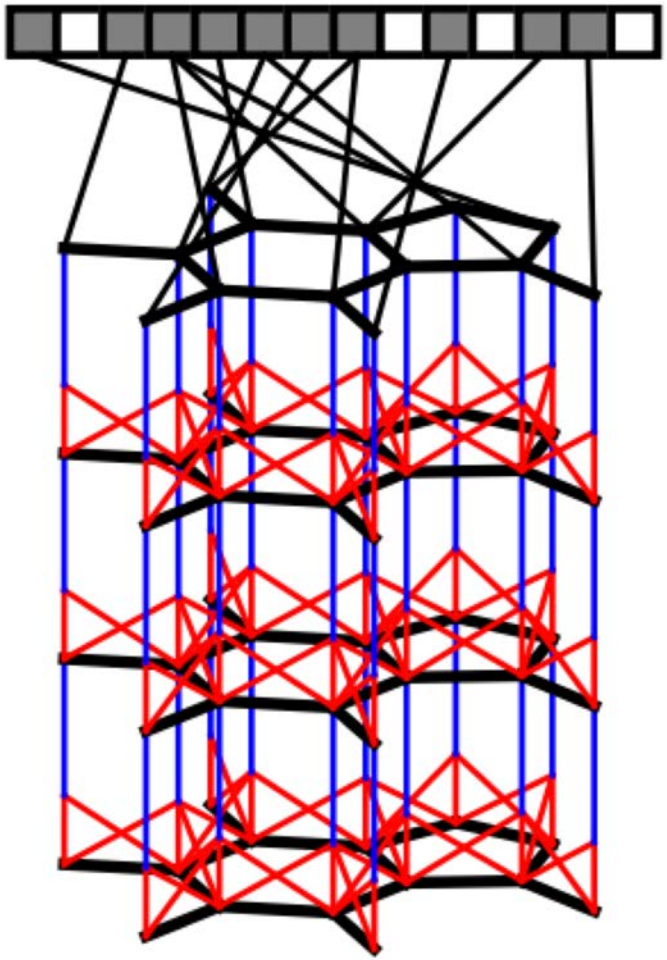
Pharmacophore modeling





https://djwild.pages.iu.edu/ic/

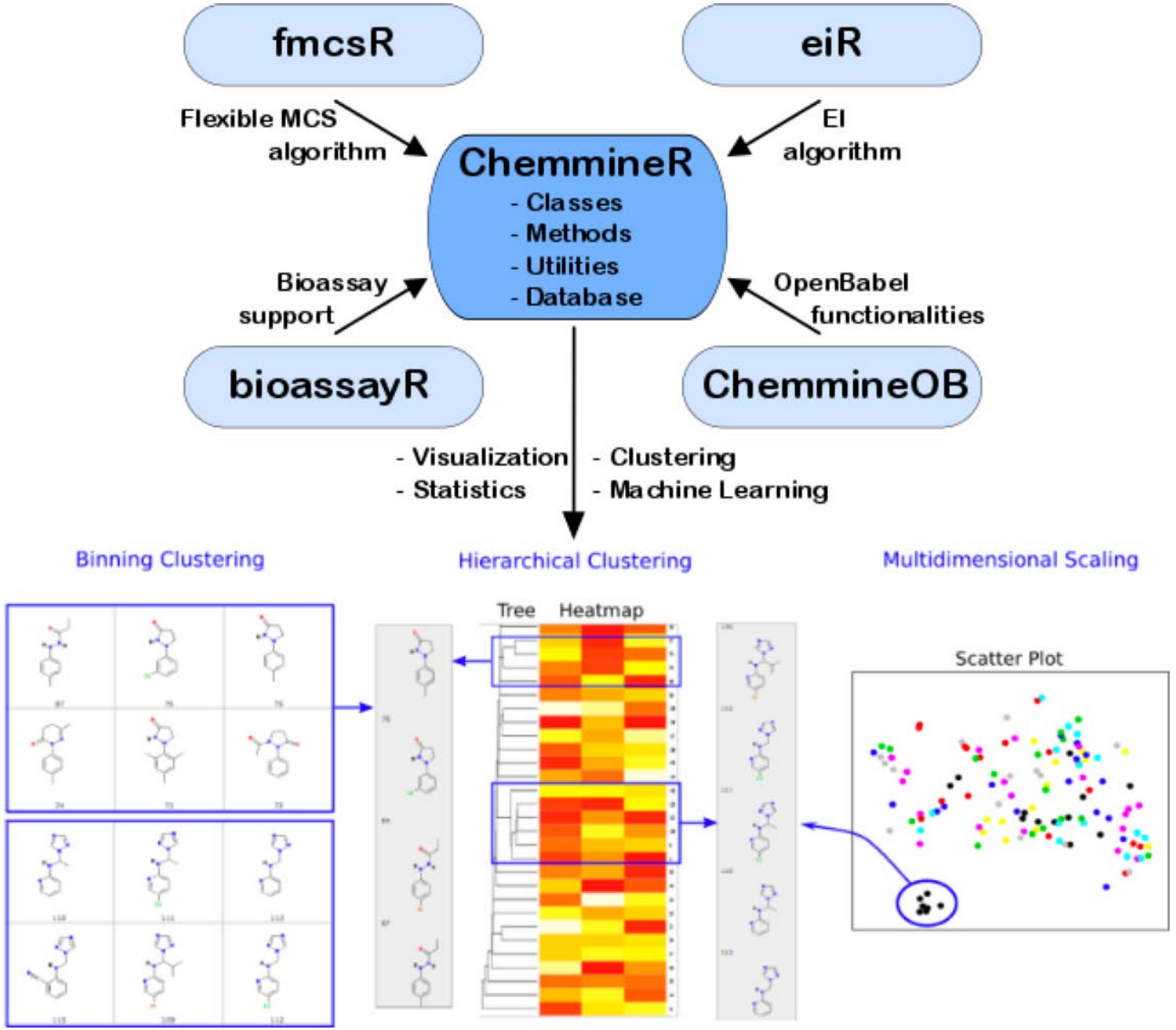
Graph Convolutional Networks



David Duvenaud, NIPS, 2015

applications

- Computer similarity between two compounds
- Search structurally similar compounds
- Cluster compounds
- Prediction of compound biological/physical properties


fmcsR
eiR
Flexible MCS algorithm
EI algorithm
ChemmineR
- Classes
- Methods
- Utilities
- Database
Bioassay support
OpenBabel functionalities
bioassayR
ChemmineOB
- Visualization
- Statistics
- Clustering
- Machine Learning
Binning Clustering
Hierarchical Clustering
Multidimensional Scaling
Tree
Heatmap
Scatter Plot

ChemmineR