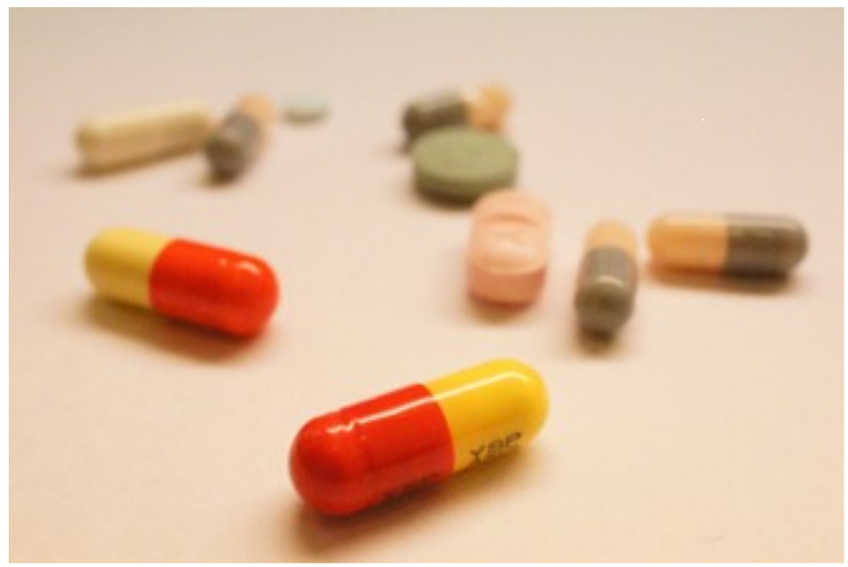
Connectivity Map: the use for drug-repurposing

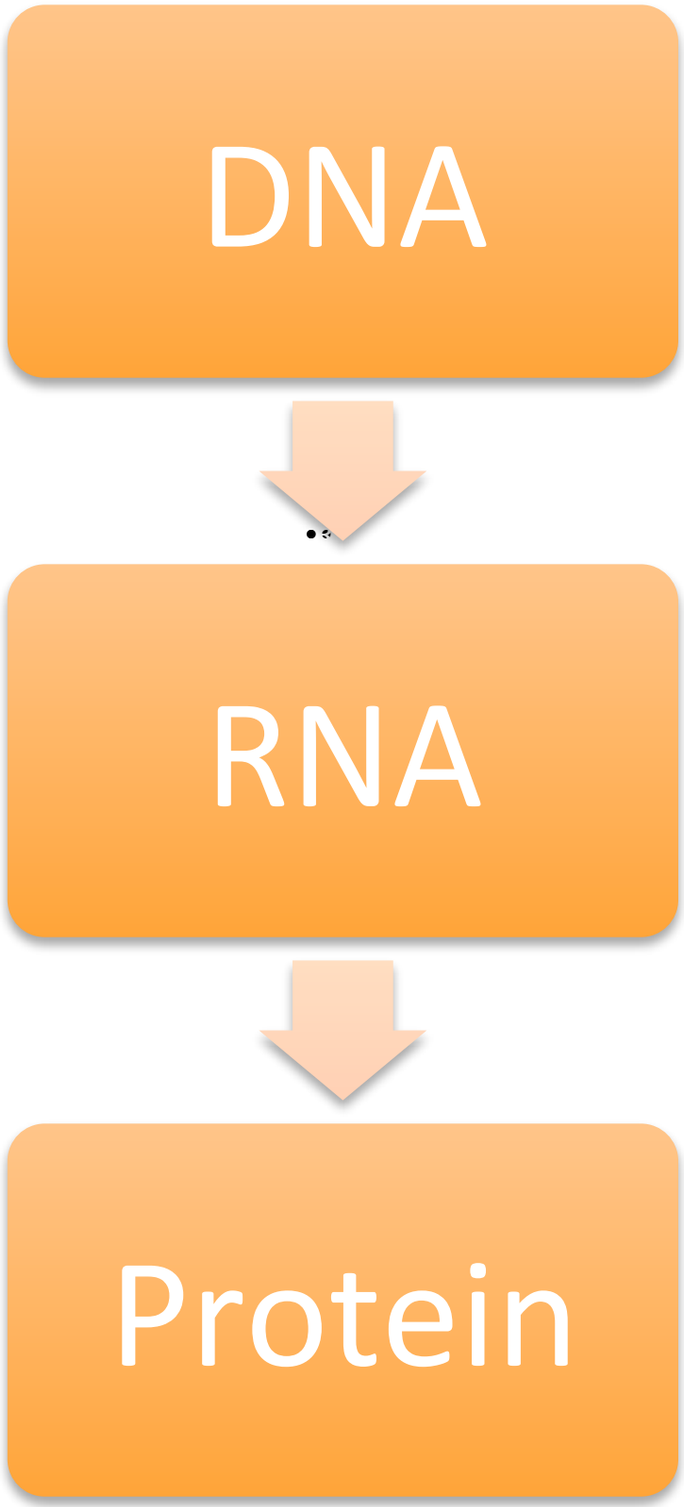
An introduction

Speaker	Wu-Lung Roger Yang	楊伍隆
Advisors	Prof. Chi-Ying Huang	黃奇英
	Prof. Kun-Mao Chao	趙坤茂



May 14, 2012

Central Dogma

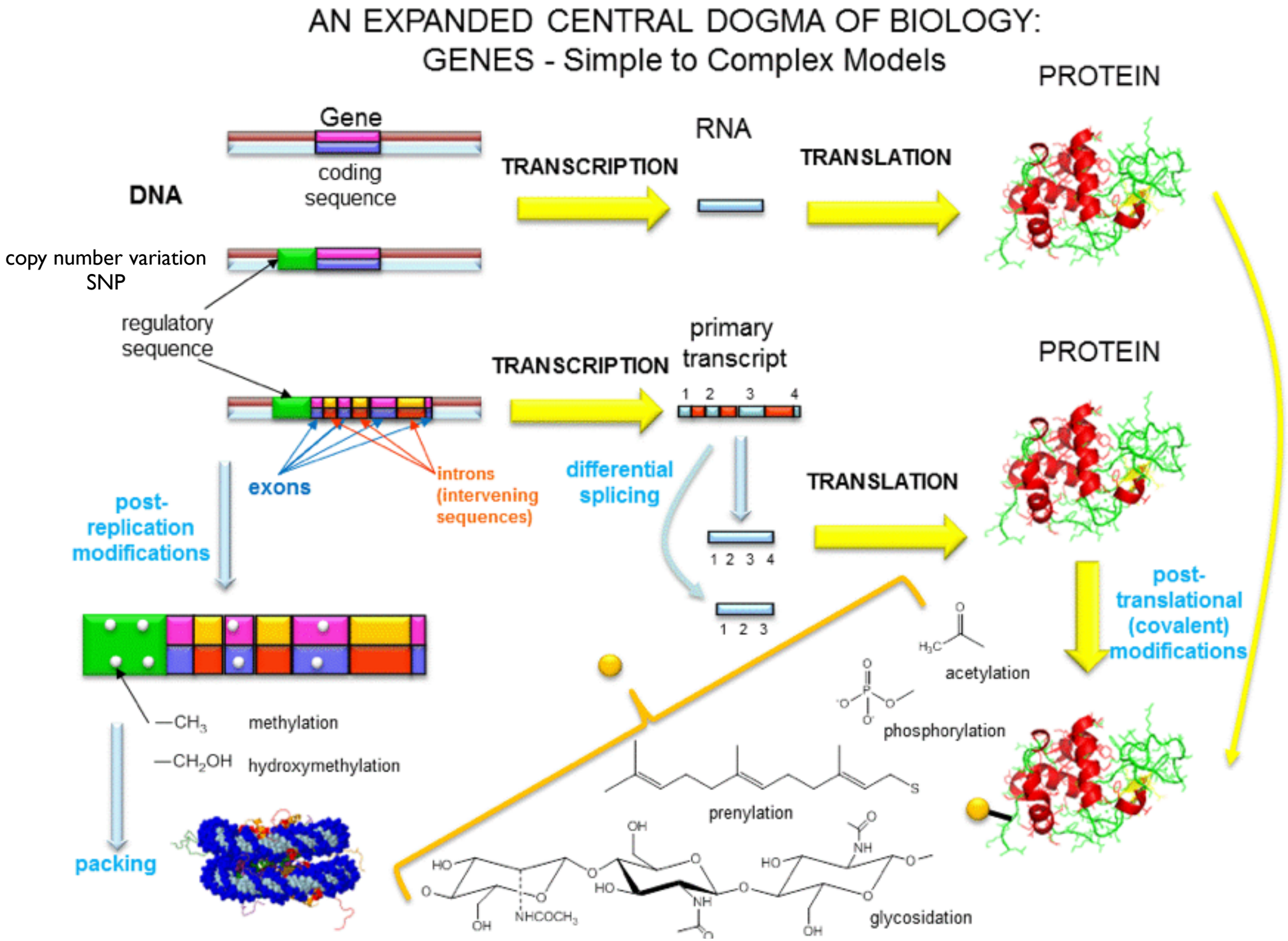


duplication

transcription

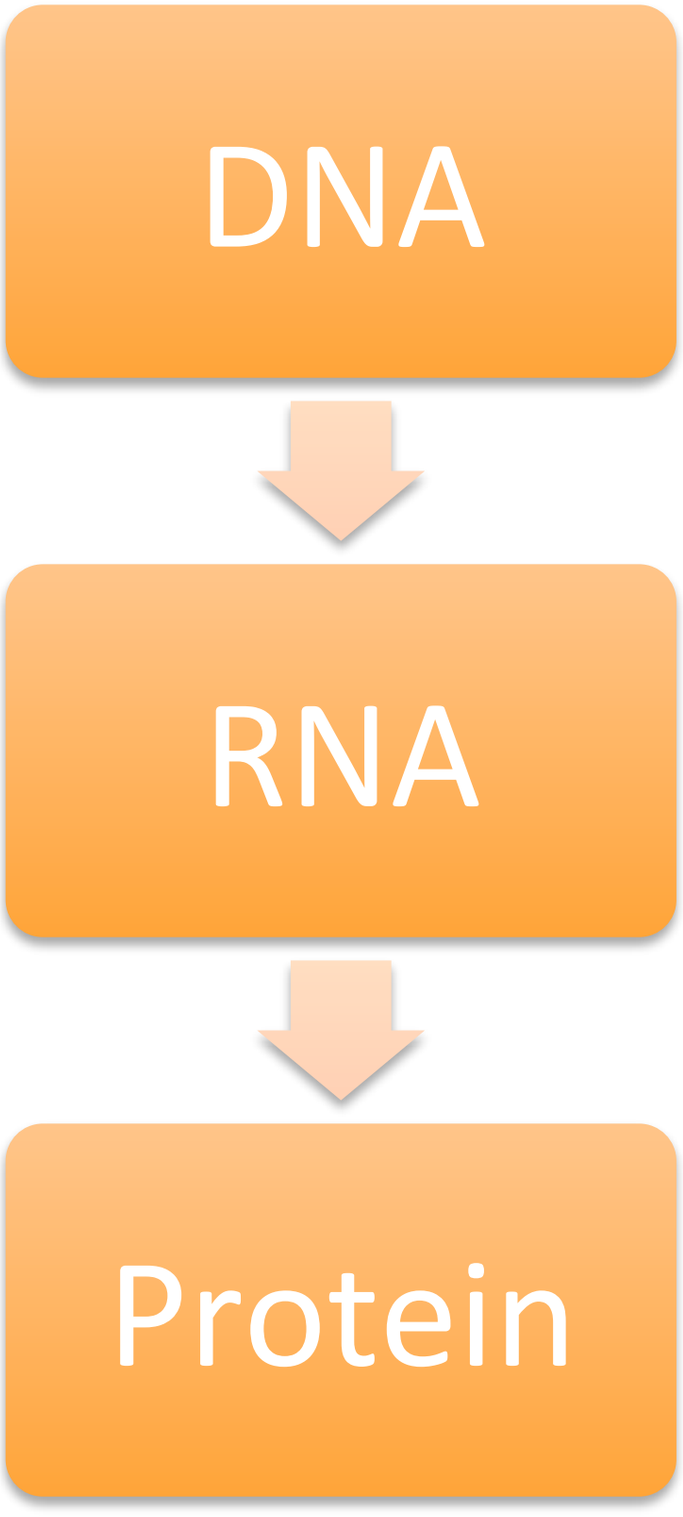
translation

Extended central dogma



http://employees.csbsju.edu/hjakubowski/classes/ch331/dna/genessimplecomplex.gif

Technology available



- DNA
  - Sequencing, Next generation sequencing (NGS)
  - SNP array
- RNA
  - microarray
  - NGS
- Protein
  - protein array
  - mass spectrometry

References

The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease

Justin Lamb,[1*] Emily D. Crawford,[1†] David Peck,[1] Joshua W. Modell,[1] Irene C. Blat,[1] Matthew J. Wrobel,[1] Jim Lerner,[1] Jean-Philippe Brunet,[1] Aravind Subramanian,[1] Kenneth N. Ross,[1] Michael Reich,[1] Haley Hieronymus,[1,2] Guo Wei,[1,2] Scott A. Armstrong,[2,3] Stephen J. Haggarty,[1,4] Paul A. Clemons,[1] Ru Wei,[1] Steven A. Carr,[1] Eric S. Lander,[1,5,6] Todd R. Golub[1,2,3,5,7*]

***Science* 313(5795): 1929-1935.**

Cited in Scopus: 750 (as May 4, 2013)

Cited in Scopus: 137 (as May 4, 2013)

INNOVATION

The Connectivity Map: a new tool for biomedical research

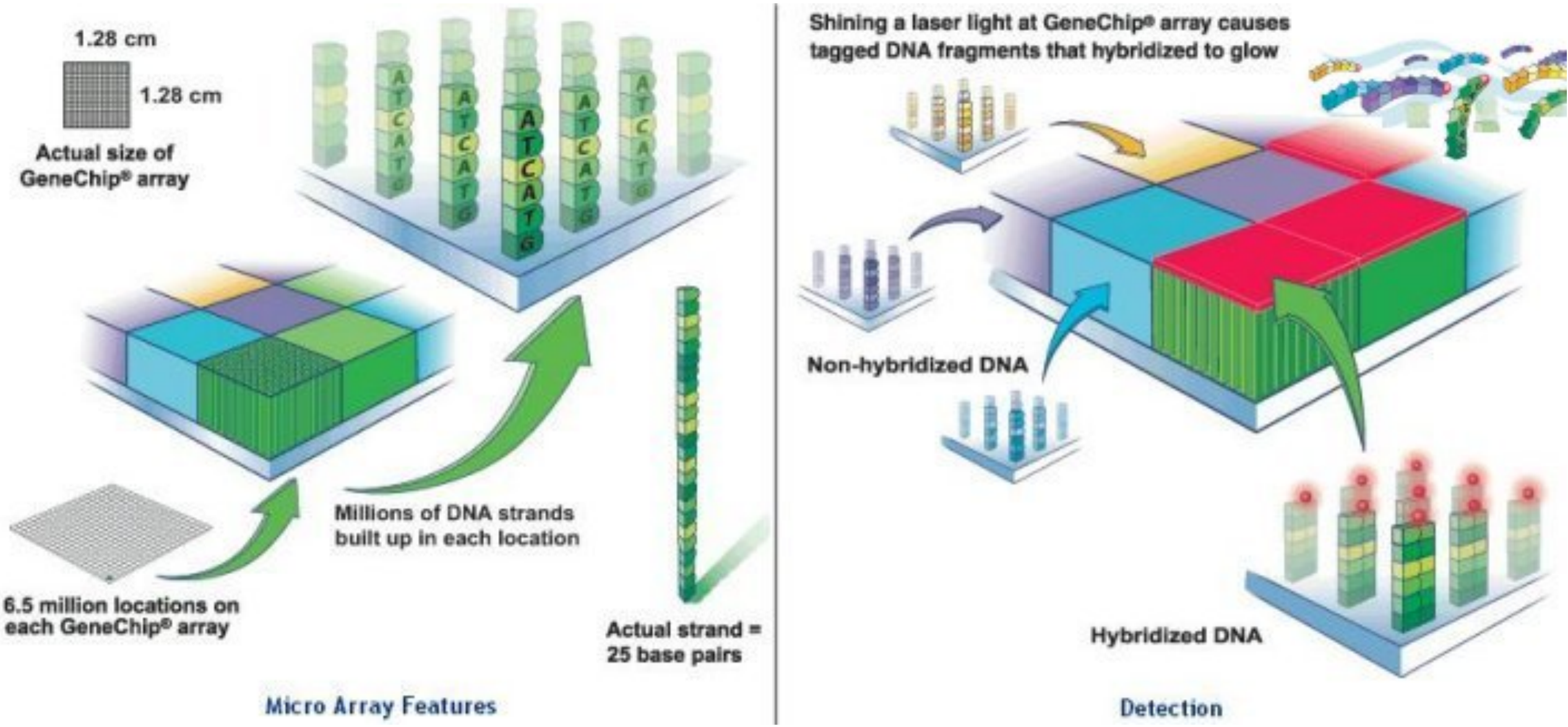
Justin Lamb

***Nature Reviews Cancer* 7(1): 54-60.**

Goals

- Create a large drug-response profile database.
- Use microarray profiles to query best matching or mismatching drugs
- Drug repurposing

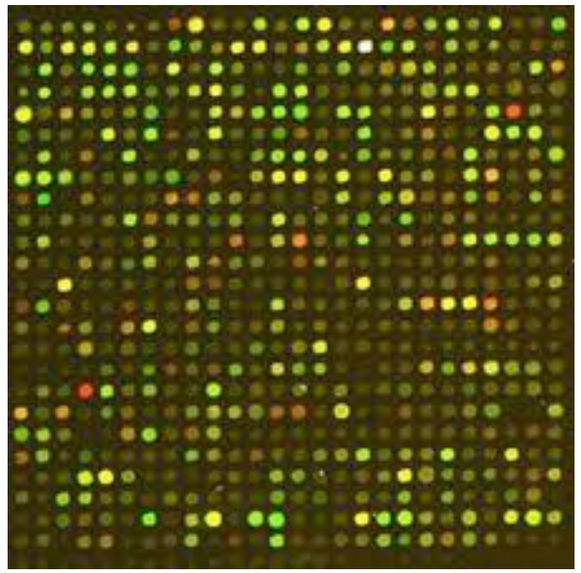
Microarray



http://2.bp.blogspot.com/_cwleTqQx7ks/TIkHQJSm-ul/AAAAAAAAABU/feudFmSg4MY/s1600/DNA+microarray.jpg

Goals

Control	Treatment
Cell	Cell + Drug

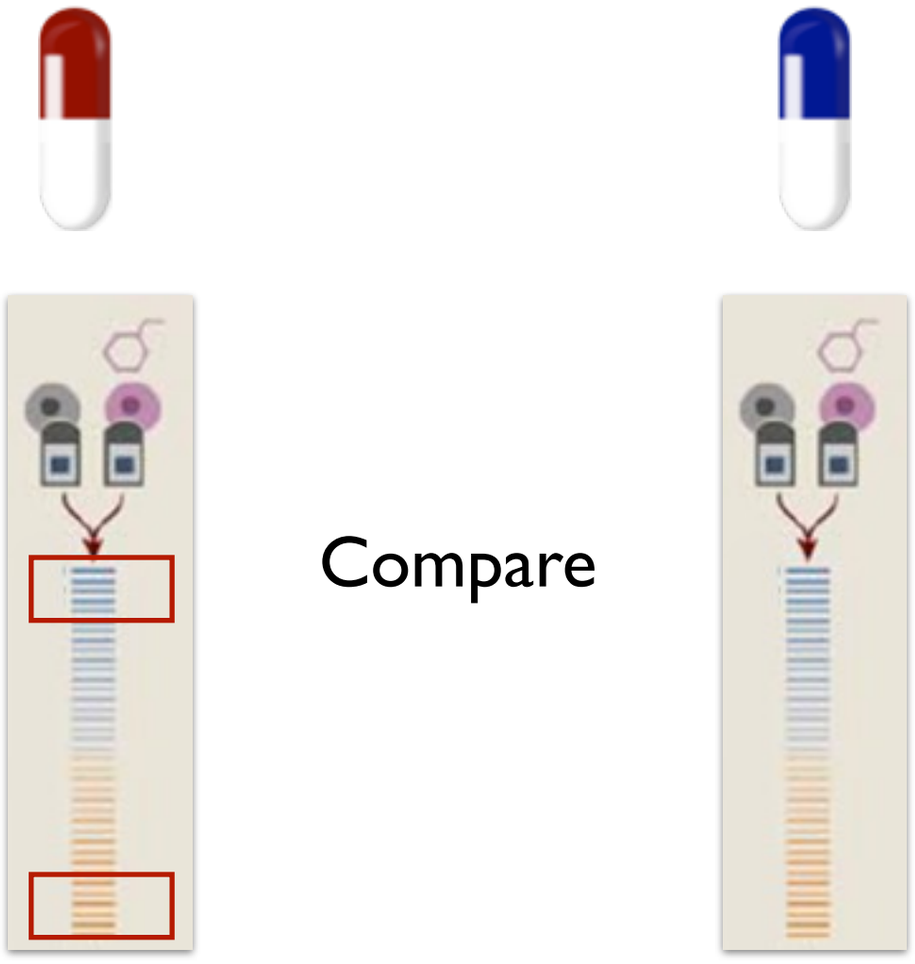


Gene	Control	Treatment	Fold Change
A	1	5	5/1
B	1	1	1/1
C	5	1	1/5
...			

measurement of mRNA

	Drug 1	Drug 2	Drug 3	Drug ...	Drug N
Gene 1					
Gene 2					
Gene 3					
Gene 4					
Gene 5					
Gene 6					
Gene ...					
Gene M					

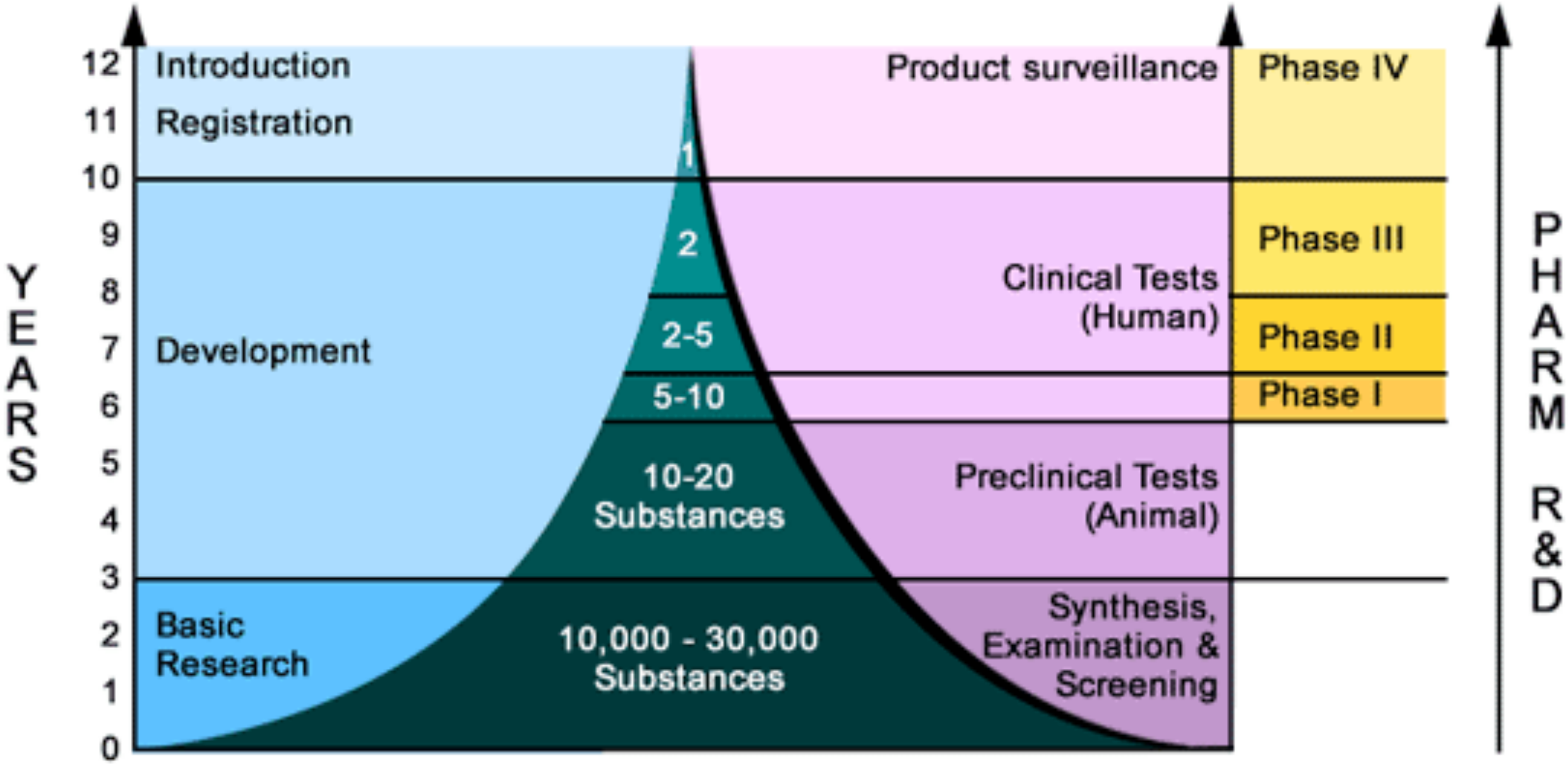
Drug Similarity



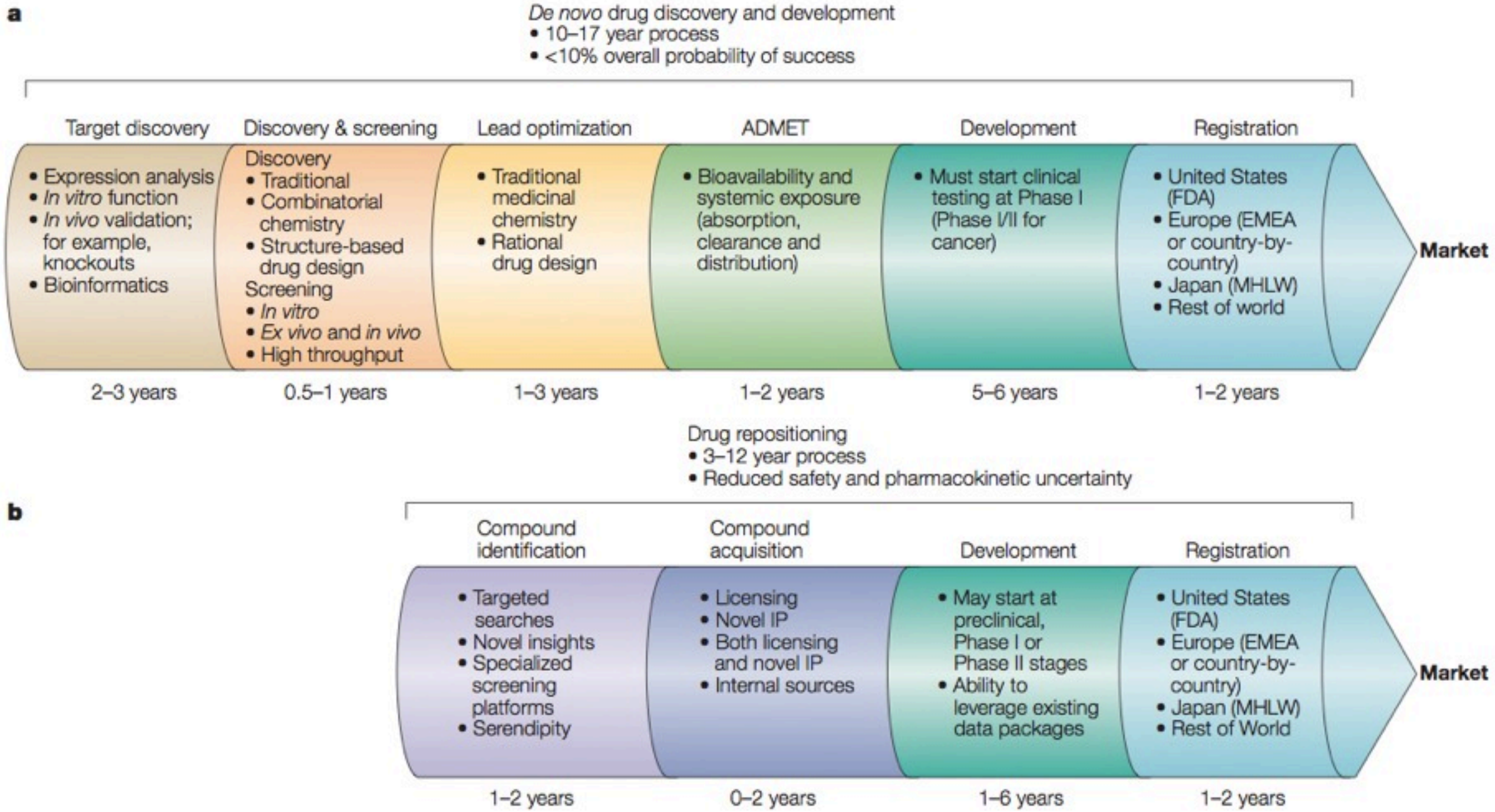
Compare

Drug Repurposing

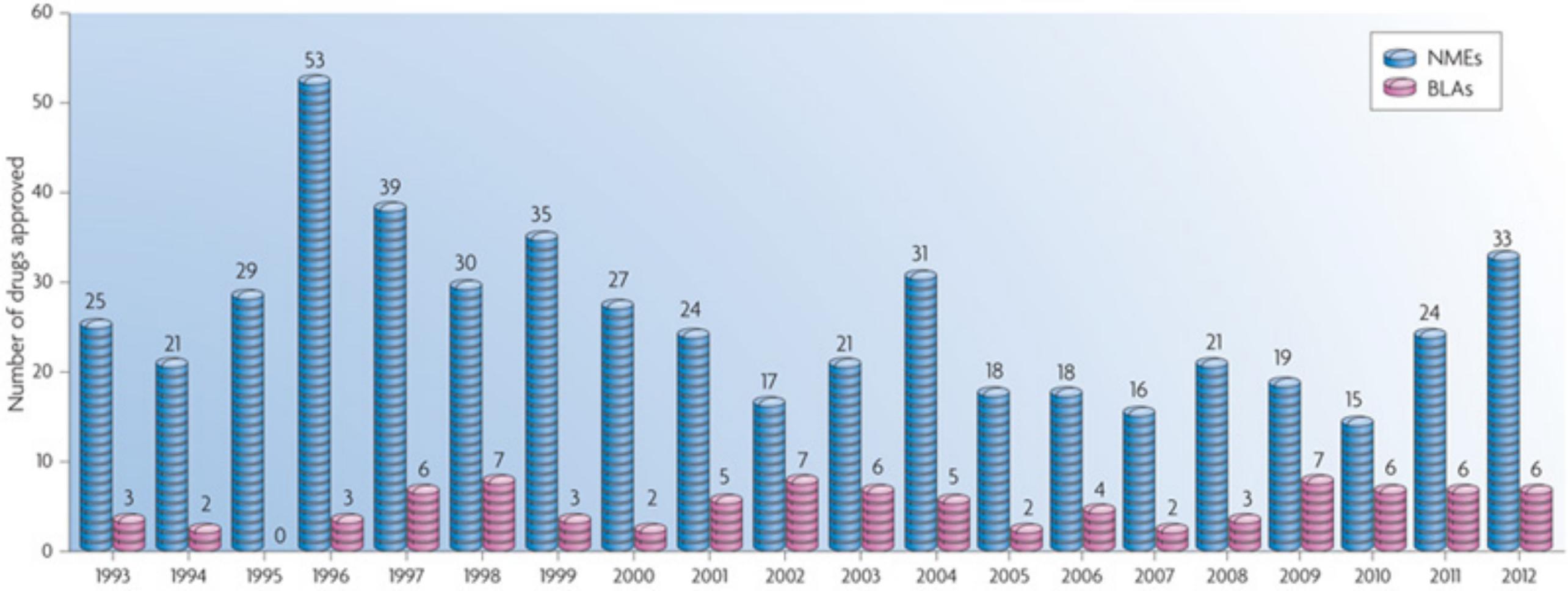
Drug Development Process



http://www.msd.com.hk/health_info/drug_education/e_ddp_introduction.html



Ashburn, Ted T, and Karl B Thor. 2004. "Drug repositioning: identifying and developing new uses for existing drugs." Nature Reviews Drug Discovery



Nature Reviews | Drug Discovery

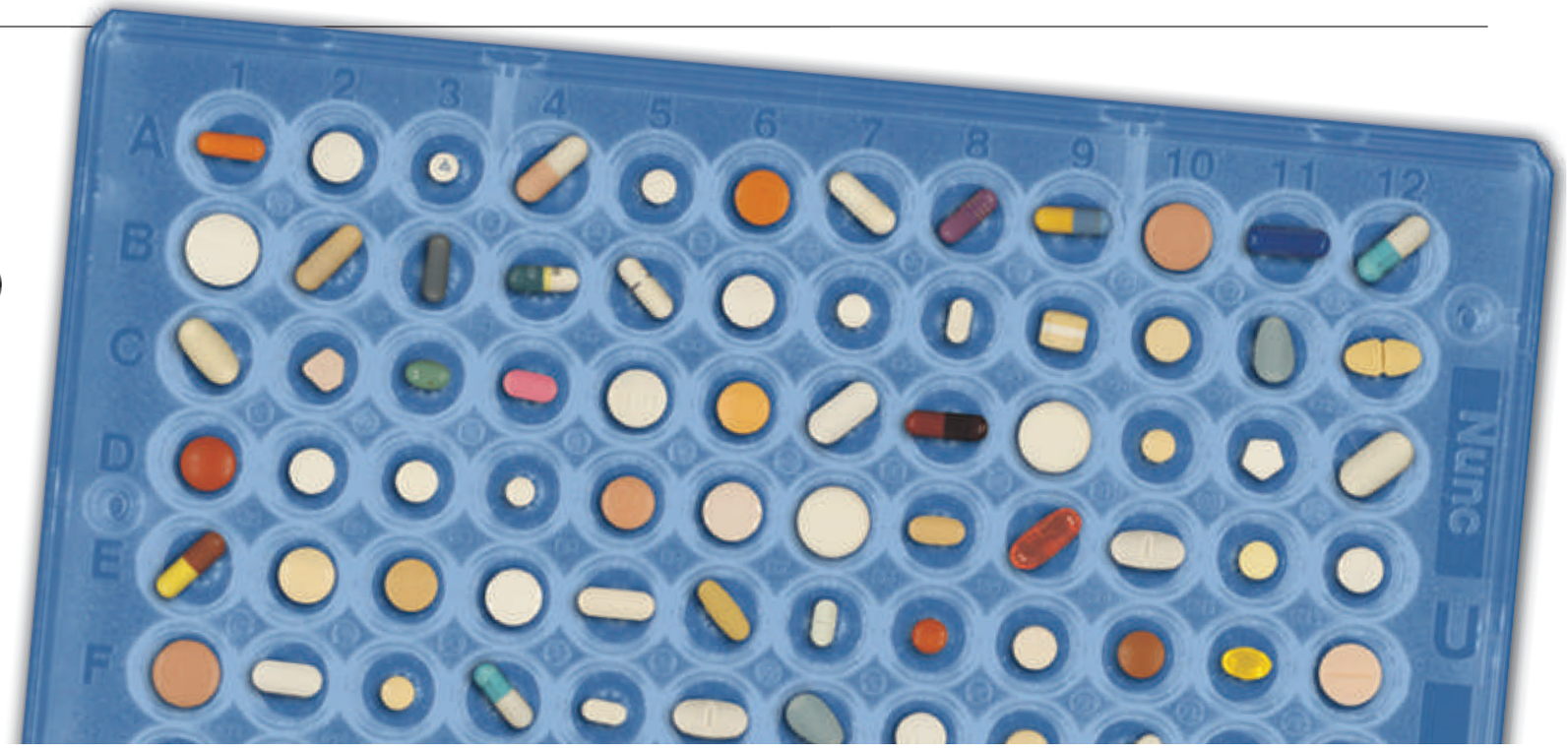
Nature Reviews Drug Discovery 12, 87-90 (February 2013) | doi:10.1038/nrd3946

- 2007: 69 drugs, but 16 are new drugs
- average 15 years and US$800 million for a new drug to market
- New drugs approved by FDA each year remain at 20~30 compounds.

New uses for old drugs

New uses for old drugs

It takes too long and costs too much to bring new drugs to market. So let's beef up efforts to screen existing drugs for new uses, argue **Curtis R. Chong** and **David J. Sullivan Jr.**



Chong, Curtis R, and David J Sullivan. 2007. "New uses for old drugs." Nature 448(7154): 645–646.

"The most fruitful basis for the discovery of a new drug is to start with an old drug"
-- Nobel laureate James Black

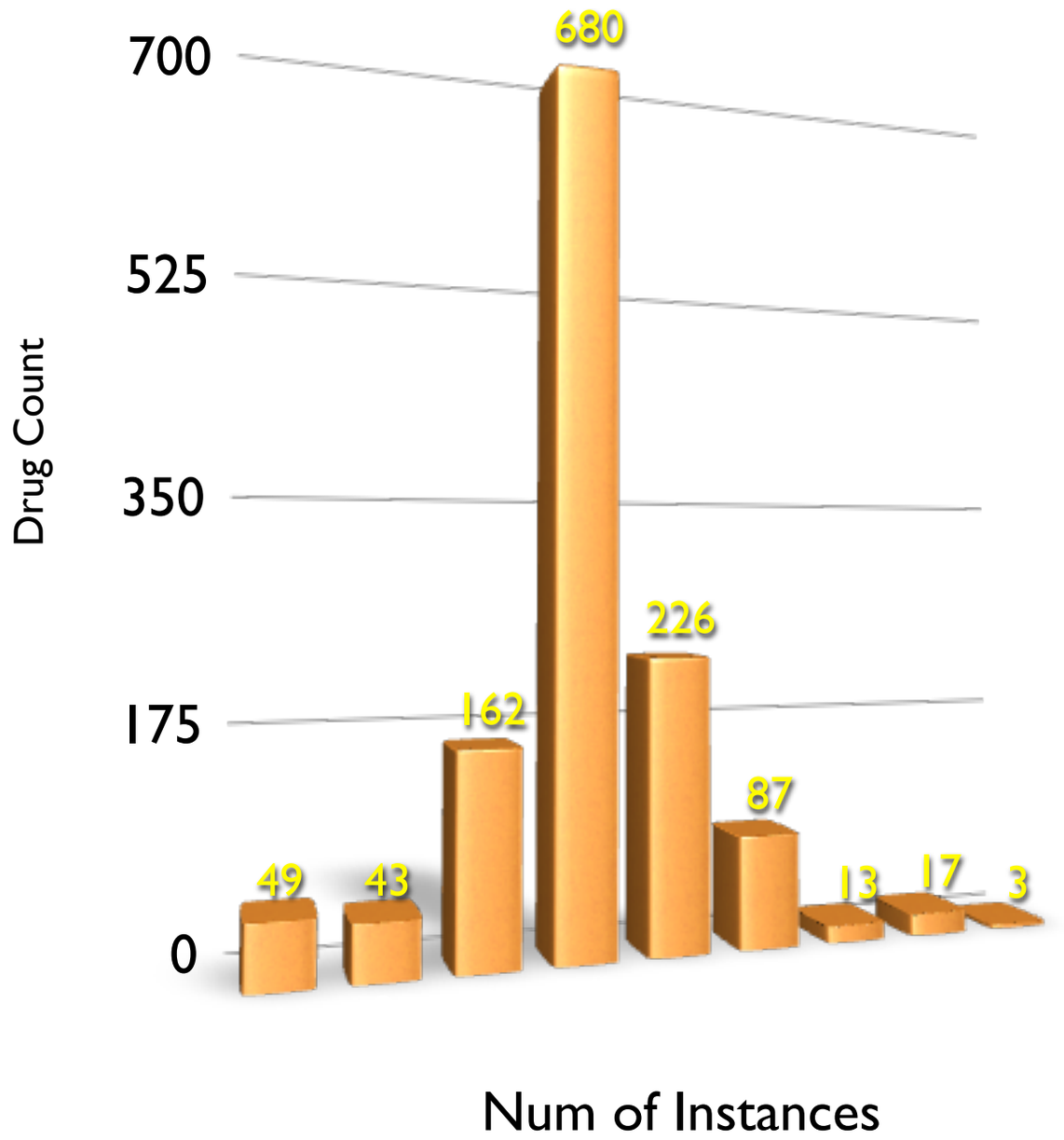
Drug repurposing examples

- Viagra for erectile dysfunction
- Nelfinavir for cancer
- Tamoxifen for bipolar disorder
- Gleevec for rheumatoid arthiritis
- Pentylenetetrazole for Down Syndrome
- Astemizole for malaria
- Lipitor for alzhemimers
- Lipitor for influenza mortality
- Metformin for cancer

CDD Community Group Meeting, SFO, Oct 1, 2009

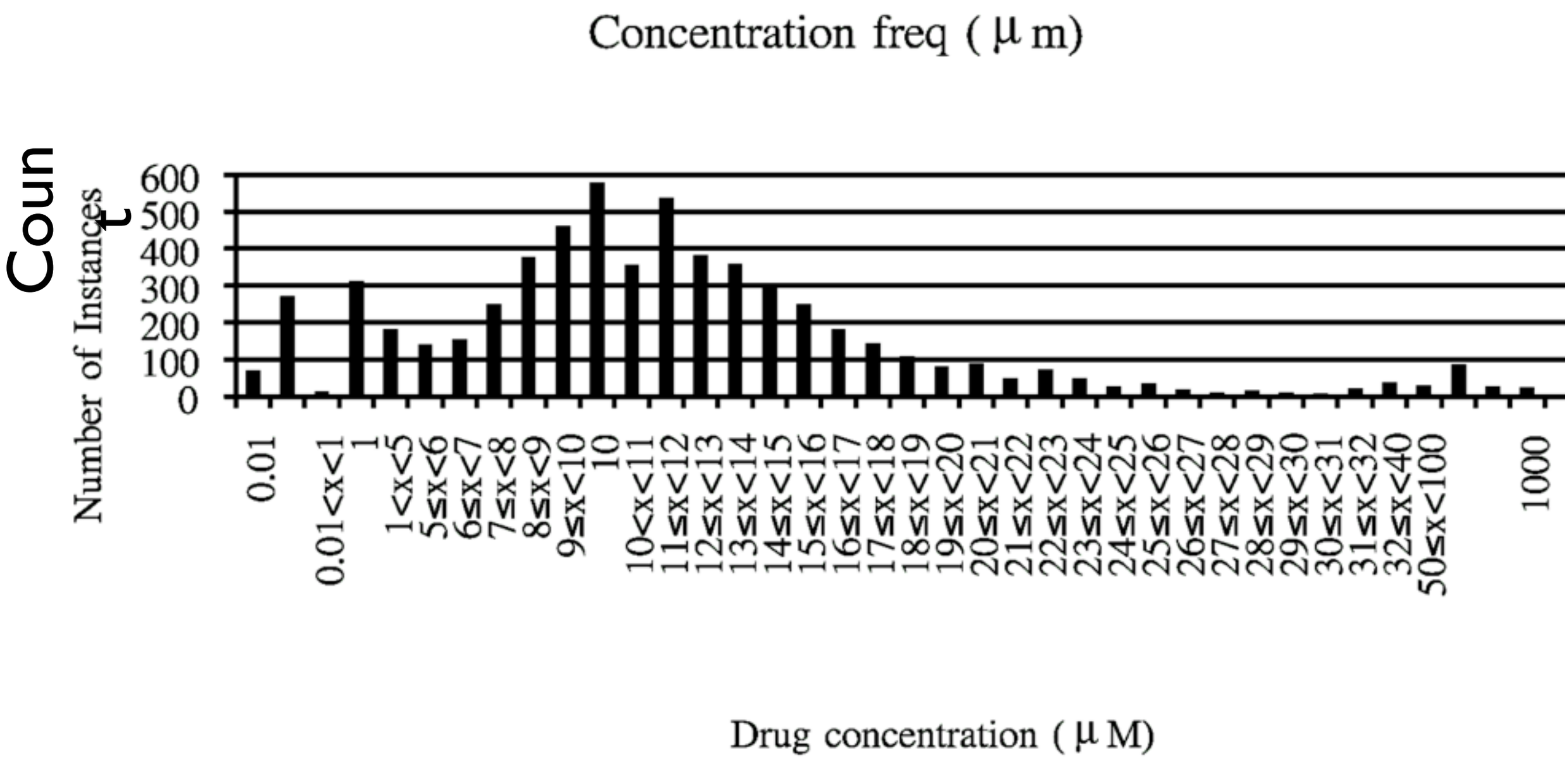
Database

Drugs



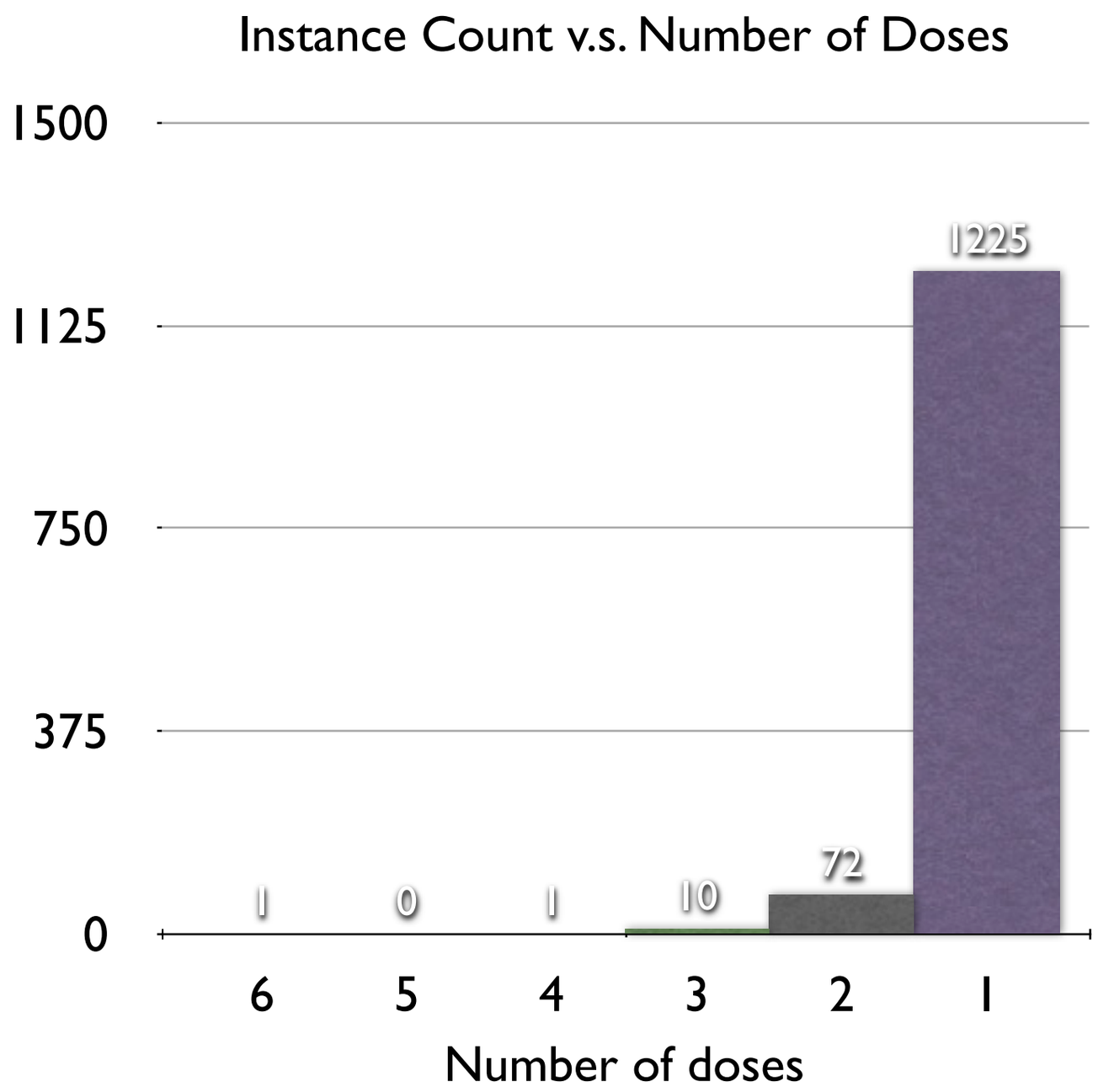
- 1309 small molecules
- 680 drugs have been conducted 4 times each.
- 6100 instances
- 7056 microarrays

Drug dosage



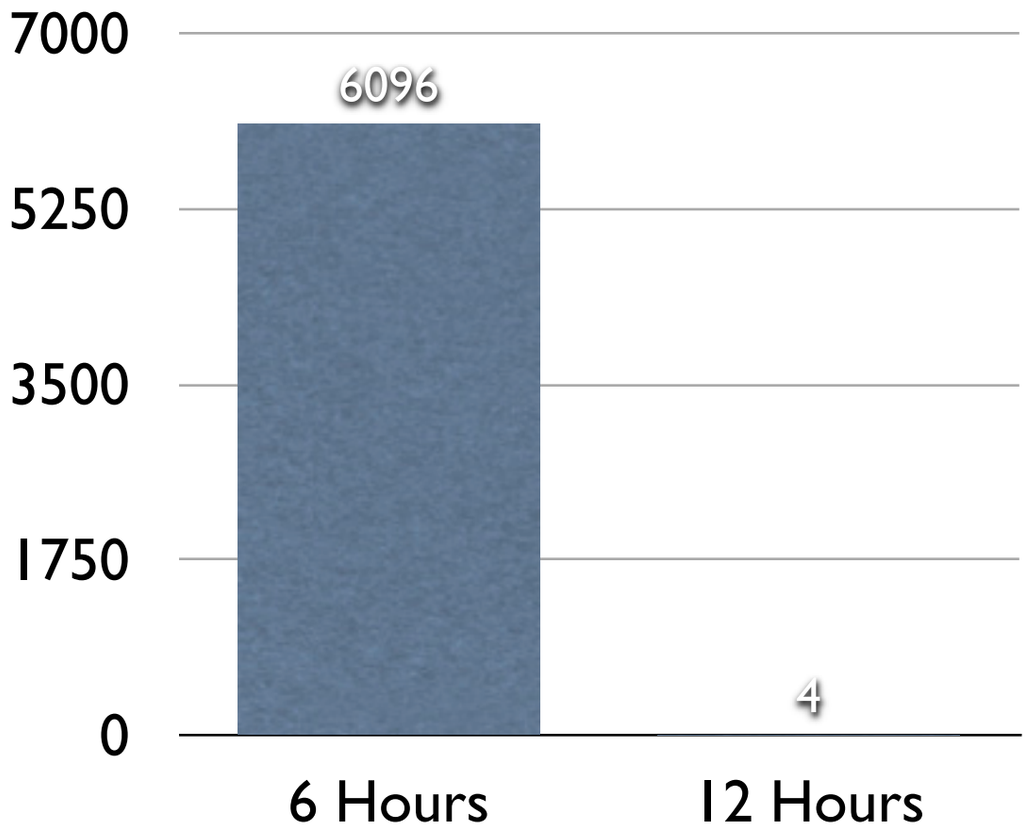
- Most drugs are tested between 1μM and 20μM

Dose count



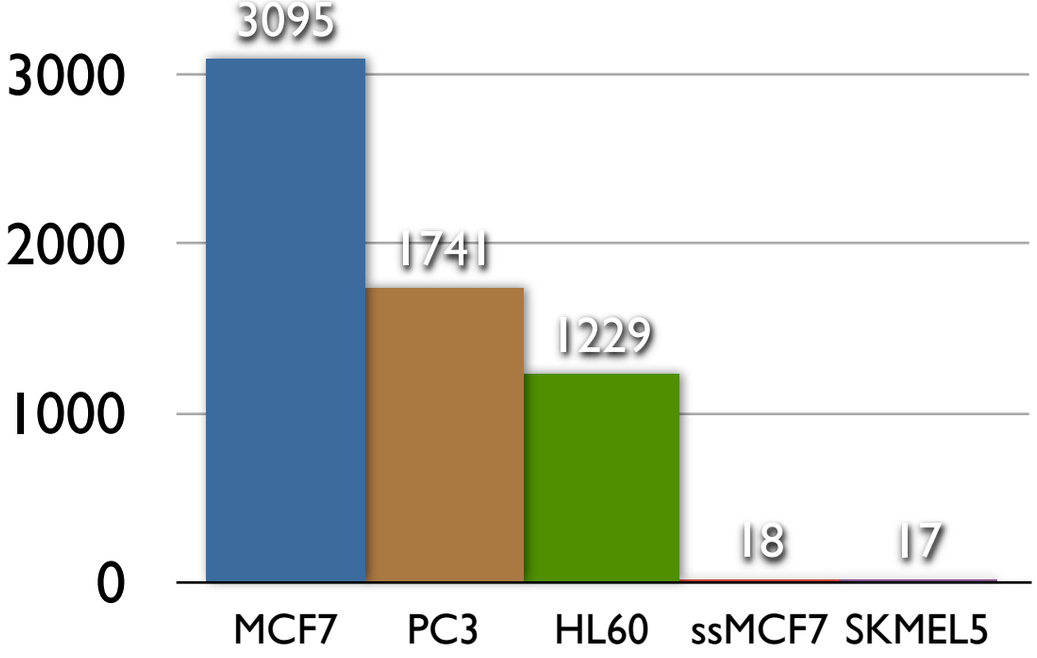
20

Duration



- Most drugs are conducted for 6 hours.
- Primary effects, not second effects

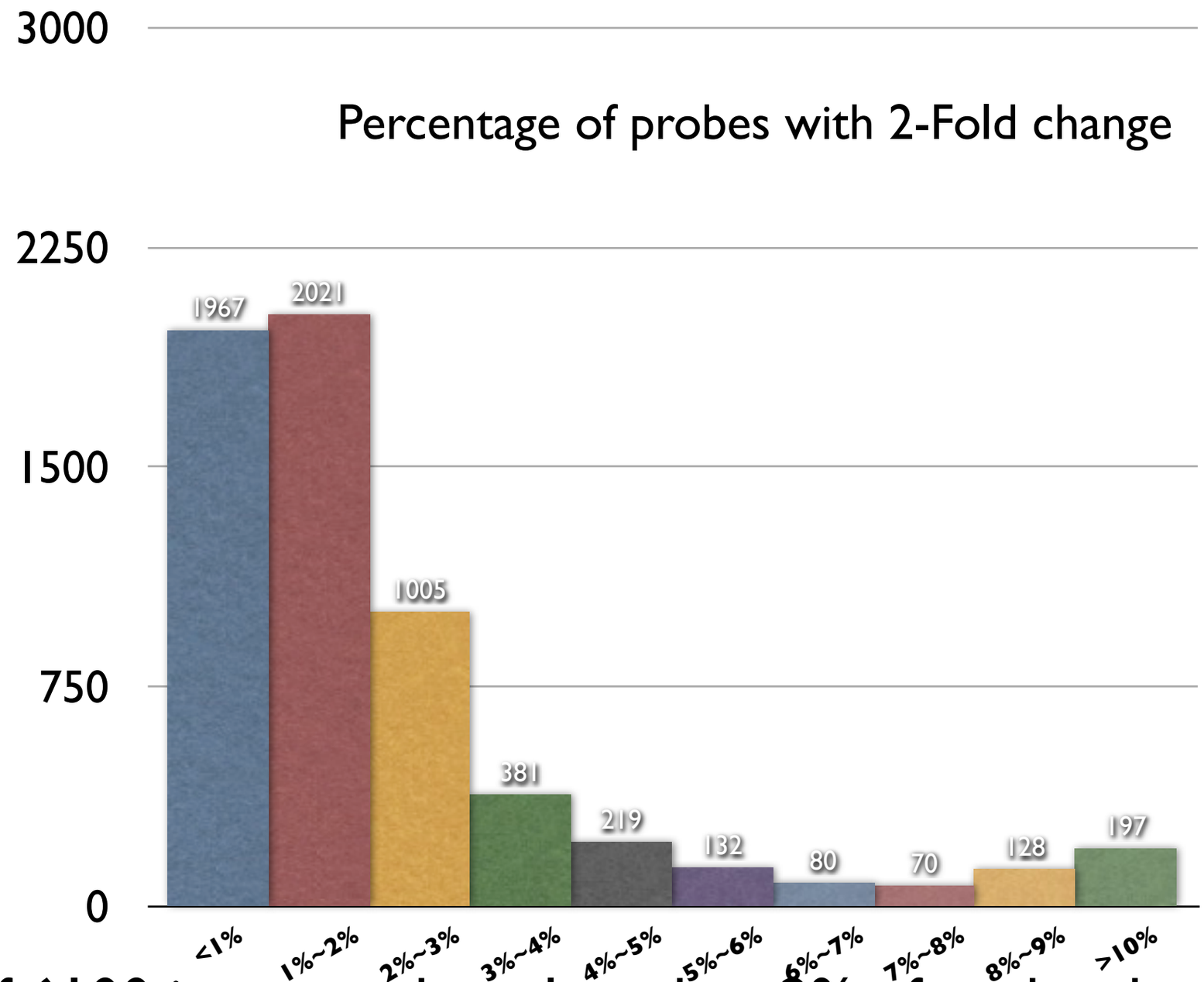
Cell Type



MCF7	Breast cancer
PC3	Prostate Cancer
HL60	Leukemia
ssMCF7	Breast cancer charcoal-stripped serum
SKMEL5	Skin Cancer

22

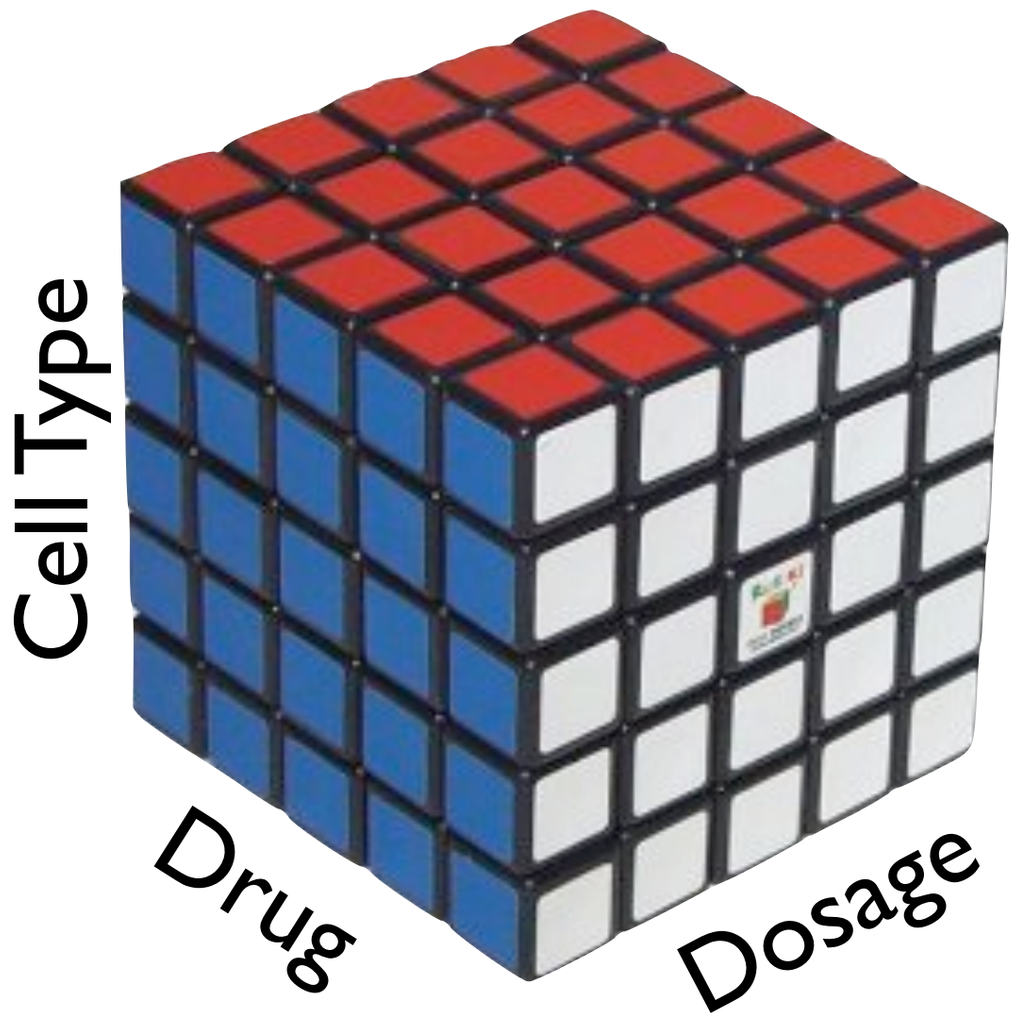
Fold Change



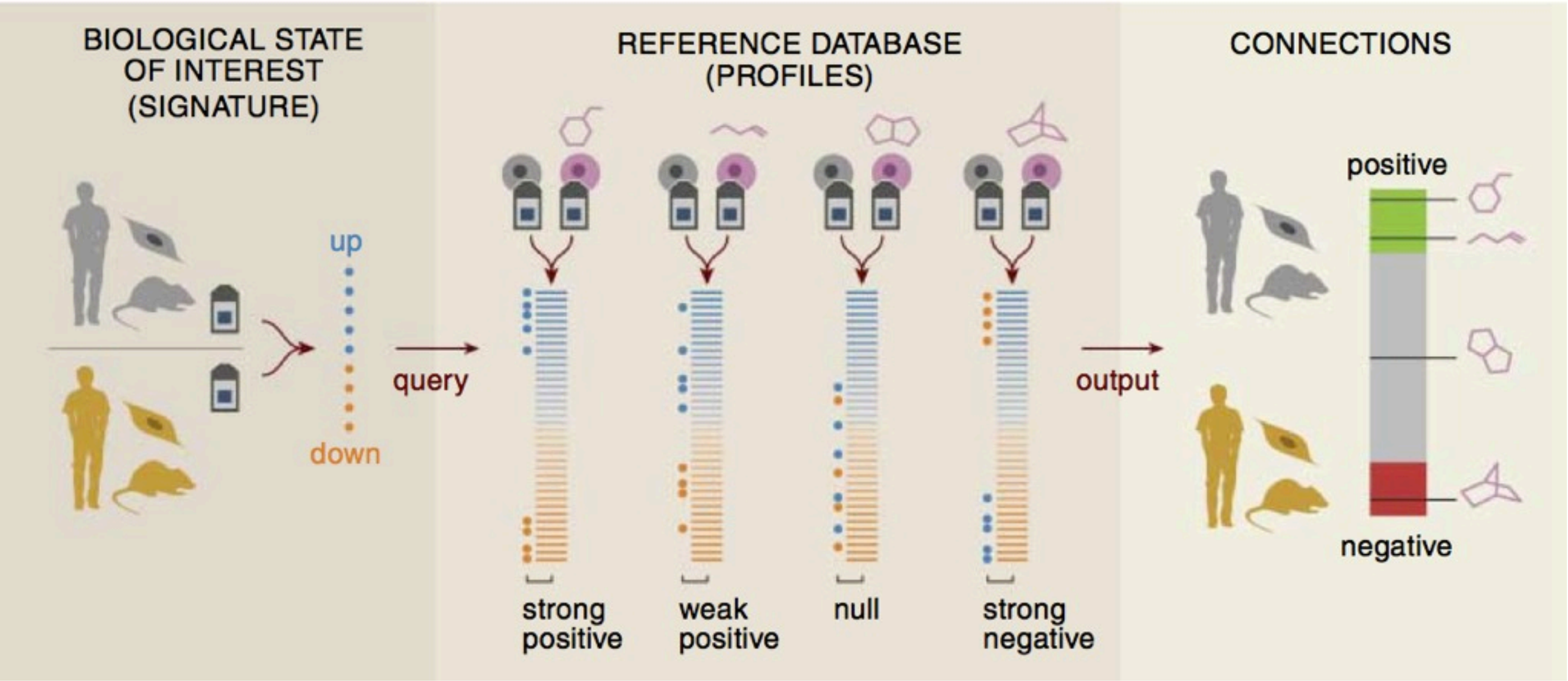
- 64% of 6100 instances have less than 2% of probes having 2-Fold change.

Instances

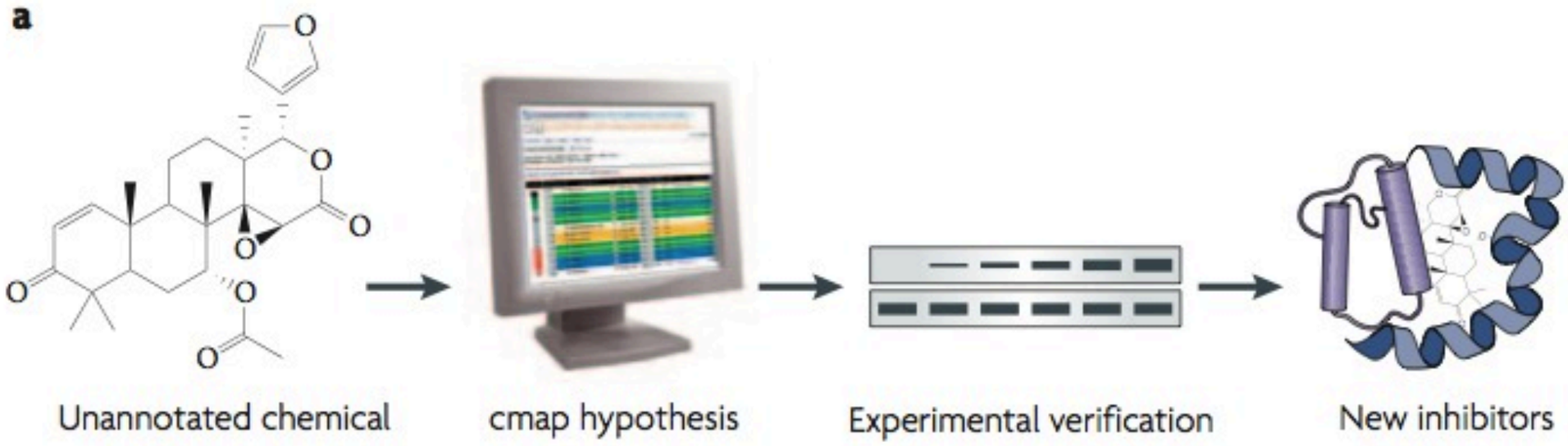
- Each instance is defined as one drug treatment over a sample of controls.
  - e.g. cells treated with vorinostat versus untreated ones.
- Instance = Drug type x Cell type x Duration x Dose
  - e.g. 10μM vorinostat on MCF7 for 6 hours
- 6100 instances over 1309 drugs



2008 price: $250 US



Lamb, J et al. 2006. "The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease." Science 313(5795):



Lamb, Justin. 2007. "The Connectivity Map: a new tool for biomedical research." Nature Reviews Cancer 7(1): 54–60.

ScreenFlow File Edit Insert Font Actions View Window Purchase Help 7:32

Connectivity Map sign-in

CONNECTIVITY MAP 02

username: spardy

password: ••••••••••

sign in

email me my password | register as a new user

The Connectivity Map (also known as cmap) is a collection of genome-wide transcriptional expression data from cultured human cells treated with bi algorithms that together enable the discovery of functional connections between drugs, genes and diseases through the transitory feature of commc cmap from our papers in *Science* and *Nature Reviews Cancer*.

This web interface provides access to the current version (**build 02**) of Connectivity Map which contains more than 7,000 expression profiles repres pharmacologists, chemists and clinical scientists to use cmap without the need for any specialist ability in the analysis of gene-expression data. The accessed here.

A brief tutorial can be found by clicking 'getting started' under the 'help' tab after log in. Detailed help and a definition of cmap terms can be found t else, please contact us.

The Connectivity Map is based at The Broad Institute of MIT and Harvard in Cambridge, Massachusetts. The cmap team is Justin Lamb, Xiaodong Lu Blat, Josh Modell, Jim Lerner, Elizabeth Liu and Emily Crawford. Jean-Philippe Brunet, Ken Ross, Michael Reich, Paul Clemons, Kathy Seiler, Steve Ha Christopher Johnson, Stephen Johnson, the MSigDB curation team, and the Genetic Analysis Platform contribute invaluable expertise and assistance leadership for the project.

privacy statement | terms and conditions

BROAD INSTITUTE

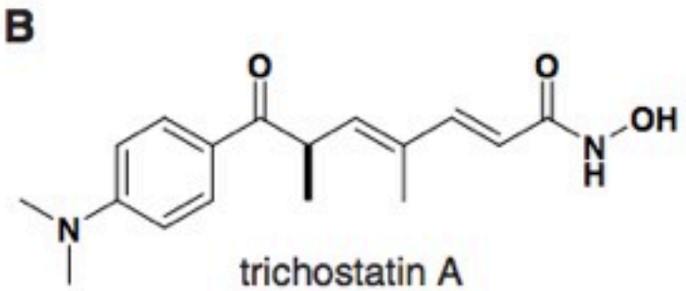
The Broad Institute is a research collaboration of MIT, Harvard and its affiliated Hospitals, and the Whitehead Institute, created to bring the power of genomics to medicine.

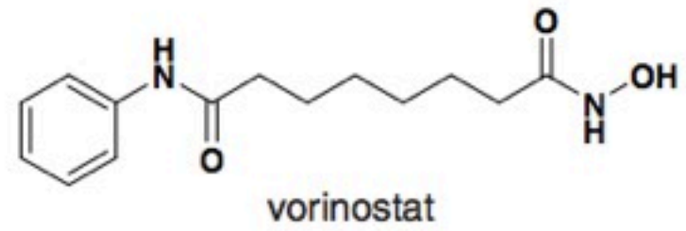
Drug-Drug similarity

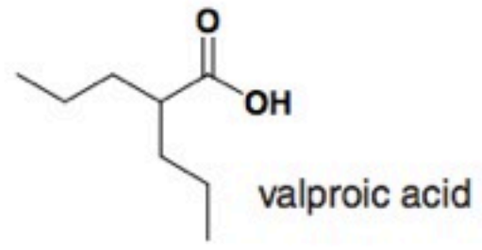
A

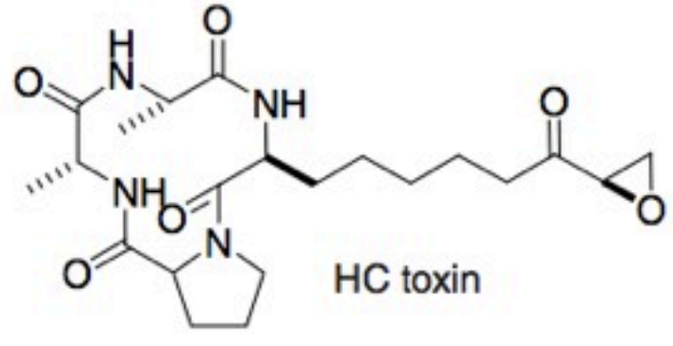
rank	perturbagen	dose	cell	score
1	vorinostat [1000]	10 µM	MCF7	1
2	trichostatin A [873]	1 µM	MCF7	0.969
3	trichostatin A [992]	100 nM	MCF7	0.931
4	trichostatin A [1050]	100 nM	MCF7	0.929
5	vorinostat [1058]	10 µM	MCF7	0.917
6	trichostatin A [981]	1 µM	MCF7	0.915
7	HC toxin [909]	100 nM	MCF7	0.914
8	trichostatin A [1112]	100 nM	MCF7	0.908
9	trichostatin A [1072]	1 µM	MCF7	0.906
10	trichostatin A [1014]	1 µM	MCF7	0.893
11	trichostatin A [332]	100 nM	MCF7	0.882
12	trichostatin A [331]	100 nM	MCF7	0.846
13	trichostatin A [448]	100 nM	PC3	0.788
14	valproic acid [345]	10 mM	MCF7	0.743
15	valproic acid [23]	1 mM	MCF7	0.735
16	valproic acid [1047]	1 mM	MCF7	0.733
17	trichostatin A [413]	100 nM	ssMCF7	0.725
18	valproic acid [410]	10 mM	HL60	0.725
19	valproic acid [458]	1 mM	PC3	0.680
33	valproic acid [409]	1 mM	HL60	0.634
39	valproic acid [1020]	500 µM	MCF7	0.619
52	valproic acid [346]	2 mM	MCF7	0.582
61	valproic acid [1078]	500 µM	MCF7	0.563
71	valproic acid [629]	1 mM	SKMEL5	0.539
72	valproic acid [347]	500 µM	MCF7	0.539
73	valproic acid [989]	1 mM	MCF7	0.538
76	valproic acid [433]	1 mM	PC3	0.528
89	trichostatin A [364]	100 nM	HL60	0.507
92	valproic acid [497]	1 mM	ssMCF7	0.501
297	valproic acid [348]	50 µM	MCF7	0
388	valproic acid [994]	200 µM	MCF7	0
403	valproic acid [1002]	50 µM	MCF7	0
419	valproic acid [1060]	50 µM	MCF7	-0.537

B









Input: 13 HDAC gene signature

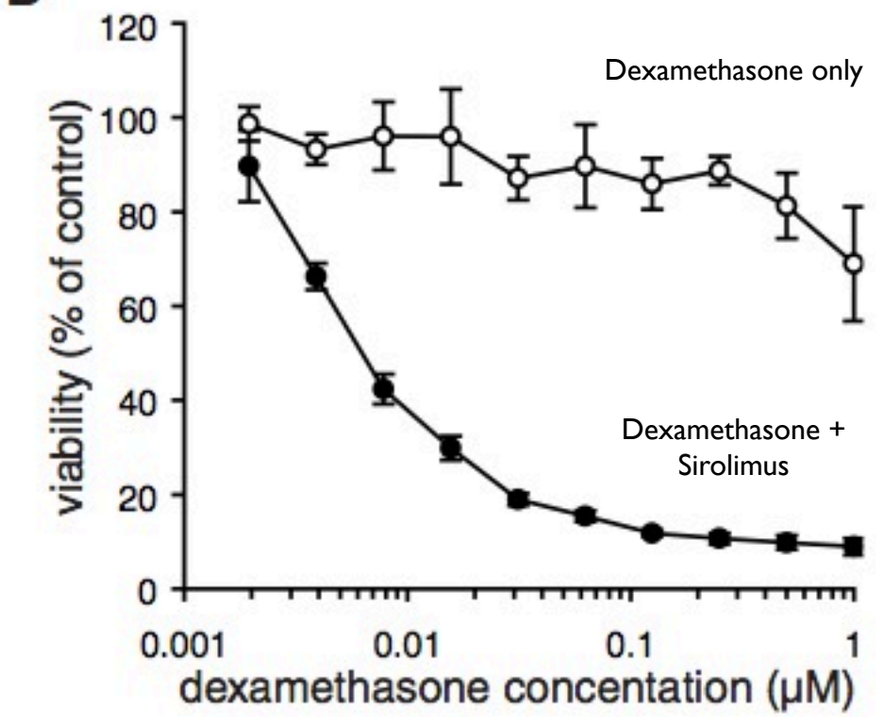
Lamb, J et al. 2006. “The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.” *Science* 313(5795): 1929–1935.

Drug-disease

A

rank	perturbagen	dose	cell	score
1	sirolimus [1059]	100 nM	MCF7	1
2	sirolimus [987]	100 nM	MCF7	0.979
7	sirolimus [1045]	100 nM	MCF7	0.930
11	sirolimus [1022]	100 nM	MCF7	0.897
16	sirolimus [1080]	100 nM	MCF7	0.880
30	sirolimus [1001]	100 nM	MCF7	0.810
35	sirolimus [402]	100 nM	ssMCF7	0.793
39	sirolimus [362]	100 nM	HL60	0.778
64	sirolimus [921]	100 nM	MCF7	0.687
104	sirolimus [326]	100 nM	MCF7	0.530

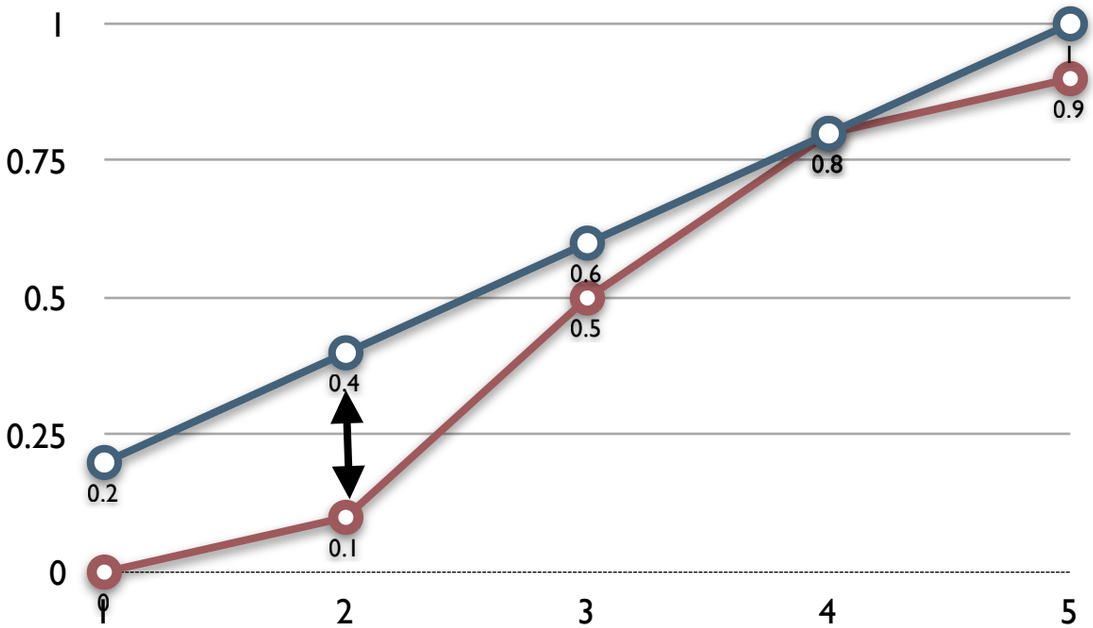
B



Lamb, J et al. 2006. “The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.” *Science* 313(5795): 1929–1935.

Sample Calculation

Kolmogorov-Smirnov statistics



$$a = \max_{j=1}^{t}\left[\frac{J}{t} - \frac{V(j)}{n}\right]$$

$$b = \max_{j=1}^{t}\left[\frac{V(j)}{n} - \frac{(j-1)}{t}\right]$$

- The score captures the maximum derivation from the random distribution.

Sample calculation

213731_s_at
208793_x_at
208711_s_at
208623_s_at
203725_at

200965_s_at
201540_at
215677_s_at
205472_s_at
204802_at

Raw score calculation

detailed results

temporary [16588] (13-AUG-09)

total instances: 6100, *signature*: temporary, *export:* Excel

permuted results | isolate shaded

barview	rank	batch	cmap name		dose	cell	score	up	down
	1	626	tanespimycin		1 µM	MCF7	1	.674	-.608
	2	634	nitrofural		20 µM	HL60	.998	.665	-.615
	3	695	zuclopenthixol		9 µM	MCF7	.988	.644	-.623
	4	613	lynestrenol		14 µM	HL60	.969	.594	-.648
	5	677	risperidone		10 µM	MCF7	.942	.831	-.376
	6	686	PHA-00851261E		10 µM	MCF7	.938	.549	-.653
	7	619	chenodeoxycholic acid		10 µM	HL60	.932	.608	-.588
	8	649	thiamazole		35 µM	HL60	.911	.600	-.568
	9	673	digoxigenin		10 µM	MCF7	.899	.681	-.472
	10	715	rolitetracycline		8 µM	PC3	.888	.741	-.398
	11	671	amrinone		21 µM	MCF7	.885	.585	-.550
	12	703	thiamazole		35 µM	PC3	.876	.697	-.426
	13	649	sulfadimethoxine		13 µM	HL60	.872	.650	-.468
	14	671	bergenin		12 µM	MCF7	.871	.697	-.420
	15	634	amphotericin B		4 µM	HL60	.871	.435	-.681

Done

Table

http://www.broadinstitute.org/cmap/ta

626 - tanespimycin 1 µM MCF7

up tags

jump to down tags

probe_id	rank	score	amplitude
208711_s_at	1134	0.149	0.24
208623_s_at	2095	0.306	0.16
203725_at	2396	0.492	0.15
208793_x_at	2801	0.674	0.13
213731_s_at	11971	0.463	0.0

down tags

jump to up tags

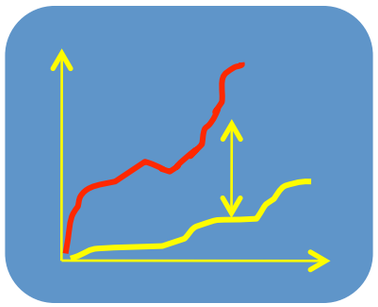
probe_id	rank	score	amplitude
215677_s_at	13542	-0.608	0.0
204802_at	13714	-0.215	0.0
200965_s_at	18125	-0.213	-0.05
201540_at	21387	-0.16	-0.25
205472_s_at	22148	0.0060	-0.5

Done

Score

$$a = \max_{j=1}^{t}\left[\frac{j}{t} - \frac{V(j)}{n}\right]$$

$$b = \max_{j=1}^{t}\left[\frac{V(j)}{n} - \frac{(j-1)}{t}\right]$$



a = 4/5 - 2801/22283 = 0.674

b =13542/22283 - 0/5 = 0.608

Table

http://www.broadinstitute.org/cmap/ta

626 - tanespimycin 1 µM MCF7

up tags

jump to down tags

probe_id	rank	score	amplitude
208711_s_at	1134	0.149	0.24
208623_s_at	2095	0.306	0.16
203725_at	2396	0.492	0.15
208793_x_at	2801	0.674	0.13
213731_s_at	11971	0.463	0.0

down tags

jump to up tags

probe_id	rank	score	amplitude
215677_s_at	13542	-0.608	0.0
204802_at	13714	-0.215	0.0
200965_s_at	18125	-0.213	-0.05
201540_at	21387	-0.16	-0.25
205472_s_at	22148	0.0060	-0.5

Done

Drug ranking

permuted results

temporary [11266] (02-FEB-09)

total instances: 6100 , *signature:* temporary, *export:* Excel

search:

by name | by name and cell line | by ATC code

rank	cmap name	mean	n	enrichment	p	specificity	% non-null	
1	hexetidine	0.644	4	0.892	0.00014	0.0000	100	
2	pivampicillin	-0.575	4	-0.852	0.00092	0.0000	100	
3	H-7	-0.659	4	-0.848	0.00097	0.1591	100	
4	0175029-0000	-0.523	6	-0.718	0.00101	0.0634	83	
5	hycanthone	0.587	4	0.837	0.00109	0.0518	100	
6	methylprednisolone	-0.582	4	-0.839	0.00119	0.0056	100	
7	0173570-0000	0.206	6	0.701	0.00185	0.0229	50	
8	phenazopyridine	-0.540	4	-0.818	0.00203	0.0127	100	
9	dicycloverine	0.305	5	0.750	0.00234	0.0055	60	
10	omeprazole	-0.535	4	-0.809	0.00259	0.0148	100	
11	rimexolone	-0.562	4	-0.808	0.00265	0.0052	100	

Enrichment

rank	cmap name	mean	n	enrichment	p	specificity	% non-null	
1	hexetidine	0.644	4	0.892	0.00014	0.0000	100	
2	pivampicillin	-0.575	4	-0.852	0.00092	0.0000	100	
3	H-7	-0.659	4	-0.848	0.00097	0.1591	100	

$$a = \max_{j=1}^{t}\left[\frac{j}{t} - \frac{V(j)}{n}\right]$$

$$b = \max_{j=1}^{t}\left[\frac{V(j)}{n} - \frac{(j-1)}{t}\right]$$

result detail

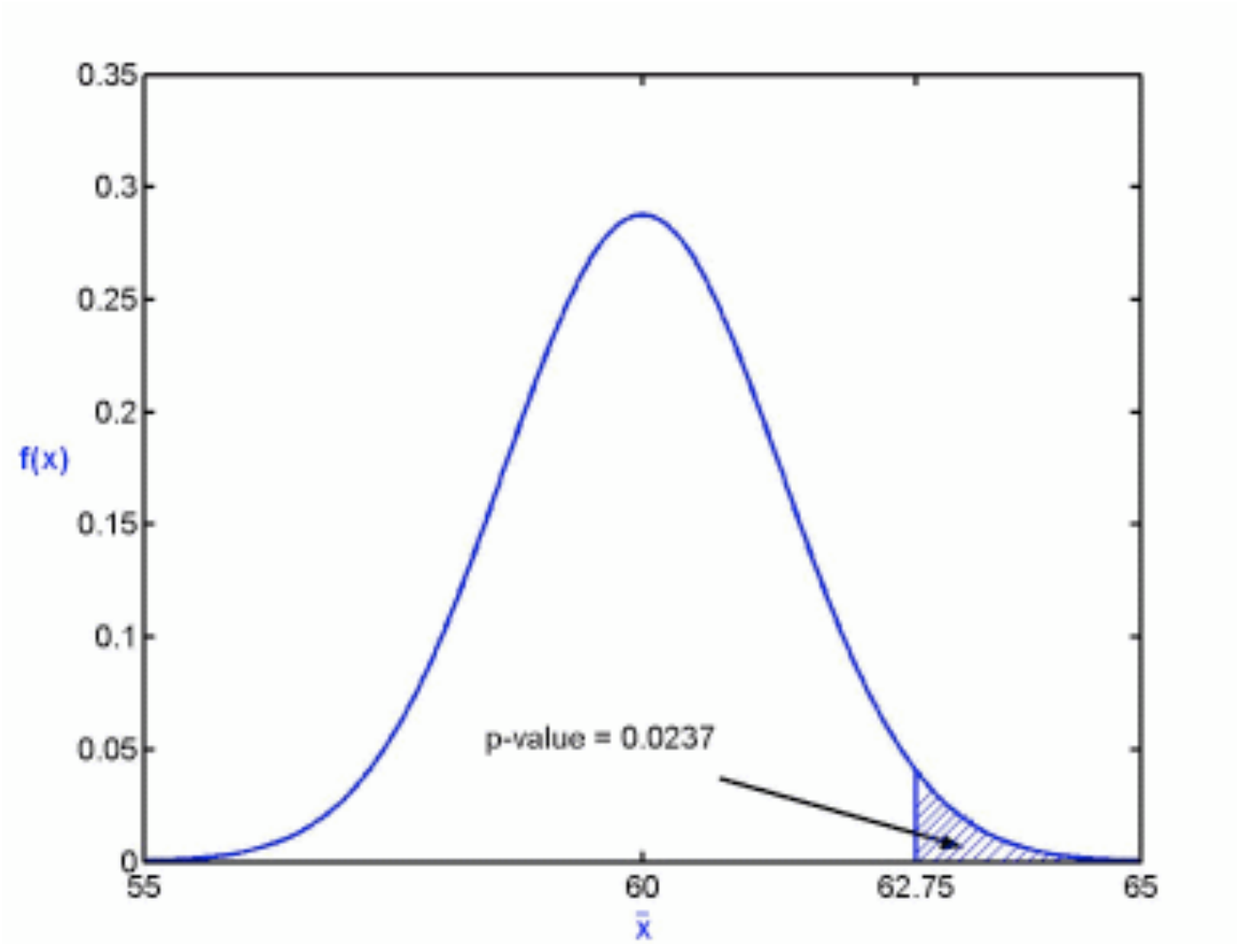
H-7

rank	batch	cmap name		dose	cell	score	up	down	instance_id
5171	1012	H-7		100 µM	MCF7	-.520	-.377	.331	5963
5380	1006	H-7		100 µM	MCF7	-.557	-.486	.271	5936
5967	1007	H-7		100 µM	PC3	-.717	-.517	.457	5941
6084	1013	H-7		100 µM	PC3	-.844	-.525	.623	5968

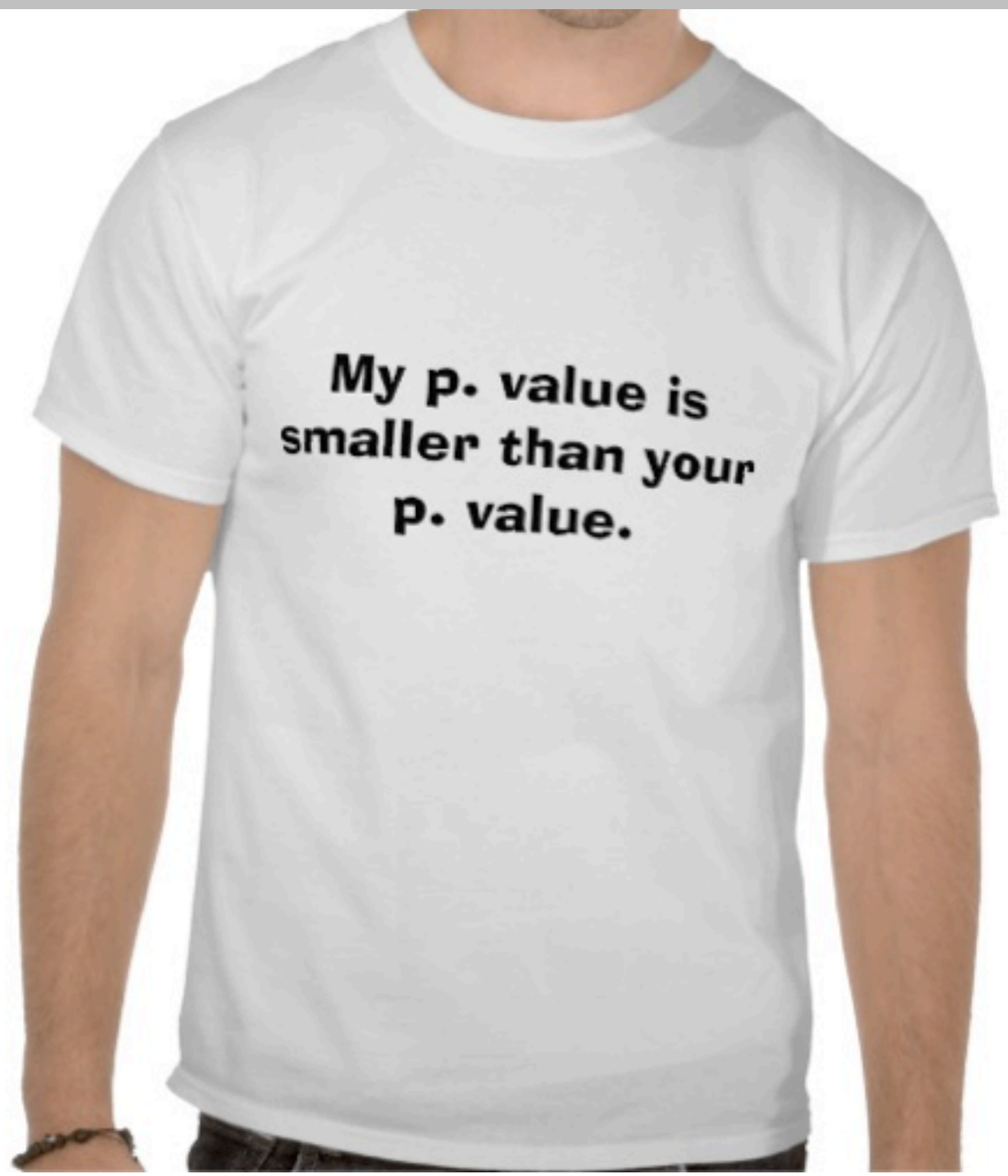
- t: number of instances
- n: total number of instances (6100)

5171/6100 - 0/1 = 0.848

Permutation



- p-value is estimated through permutation analysis by constructing empirical null distribution.



http://rlv.zcache.com/my_p_value_is_smaller_than_your_p_value_tshirt-rdb18d031a5774ce18f1a42edfe212916_804gs_512.jpg

Tutorial

Connectivity Map Tutorials
Cmap教學

0. 準備工作

- 下載檔案: Data.xls
- 這個檔案是某種藥物的基因表現。 Control 是加藥前, Drug A 是加藥後的結果。我們想尋找和這個藥相似的藥物

1. 找出Most Differentially Expressed (DE) genes

- 我們第一要找出最有變化的 gene。通常來說，我們會找 Fold Change (FC) 大於2的 gene。FC=2的意義就是說 Treatment 和 Control 相比，有2倍以上 (up-regulation)，或是1/2=0.5倍 (down-regulation) 的變化。請用 Excel 找出 FC>2 的 up-regulation 和 down-regulation genes
- 答案: up-regulated 28個
- 答案: down-regulated 13 個

2. 存grp檔案

- 我們首先要把剛剛找出來的 gene 存成檔案。 請打開文字編輯器 (e.g. NotePad) 然後每一個 gene 都存成一行
- 請在 Notepad 裡面個別儲存成 up.grp 和 down.grp
- 請在存的時後，在檔名前後打引號"。像是 "up.grp"，這樣 Notepad 才會存檔成 up.grp，而不是 up.grp.txt
- 參考檔案 test_up.grp

3. 執行Cmap

- 先進入 Cmap

- http://acb.csie.ntu.edu.tw/cmap_tutorial/

Thank you

Questions?

roger.w.yang @ gmail.com