QUAPO : Quantitative Analysis of Pooling in High-Throughput Drug Screening

Raghu Kainkaryam

Systems Biology Group
University of Michigan

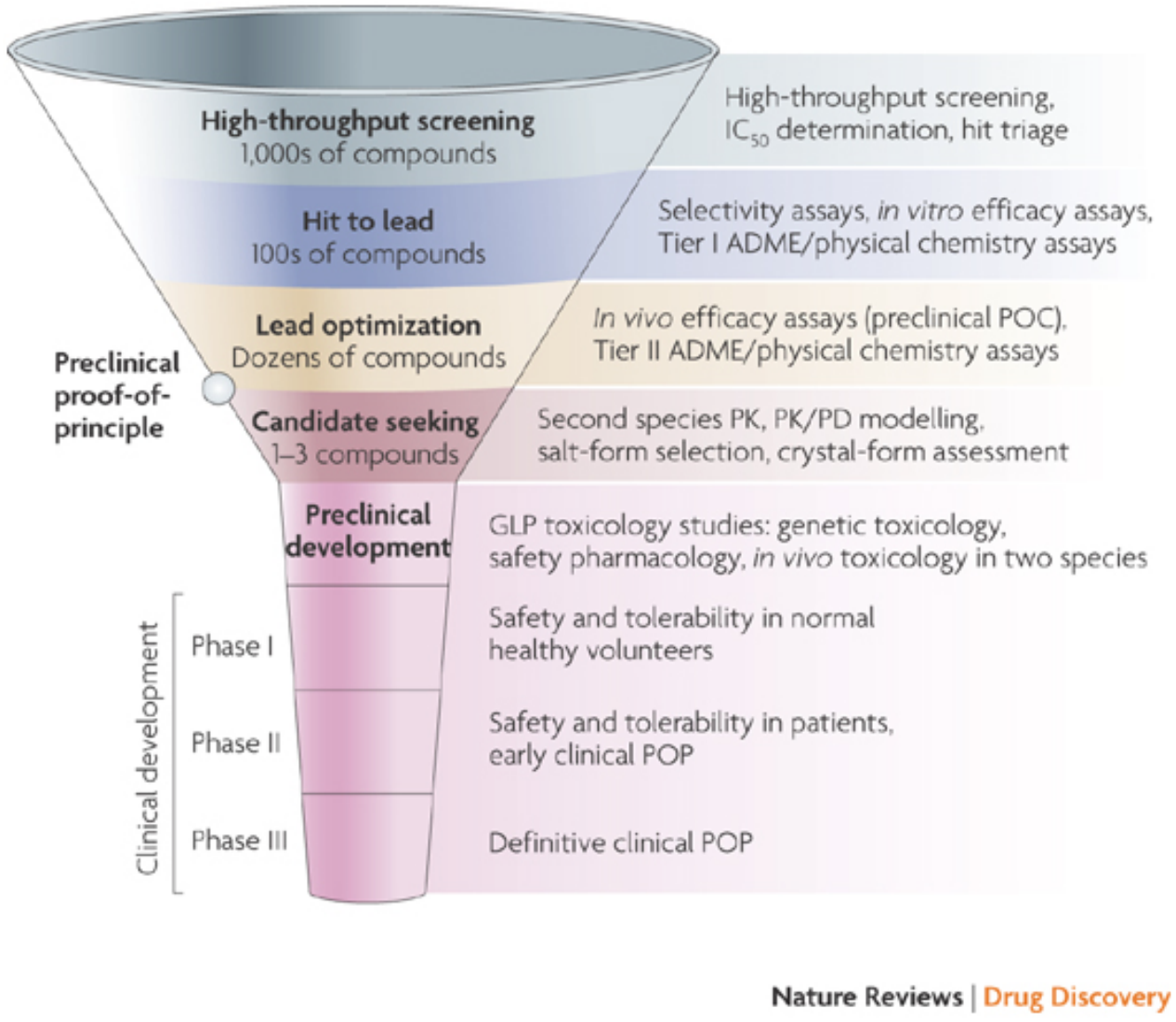
(joint work with Anna Gilbert, Paul Shearer and Peter Woolf)

March 27, 2009

DIMACS/DyDAn Workshop on Streaming, Coding & Compressive Sensing

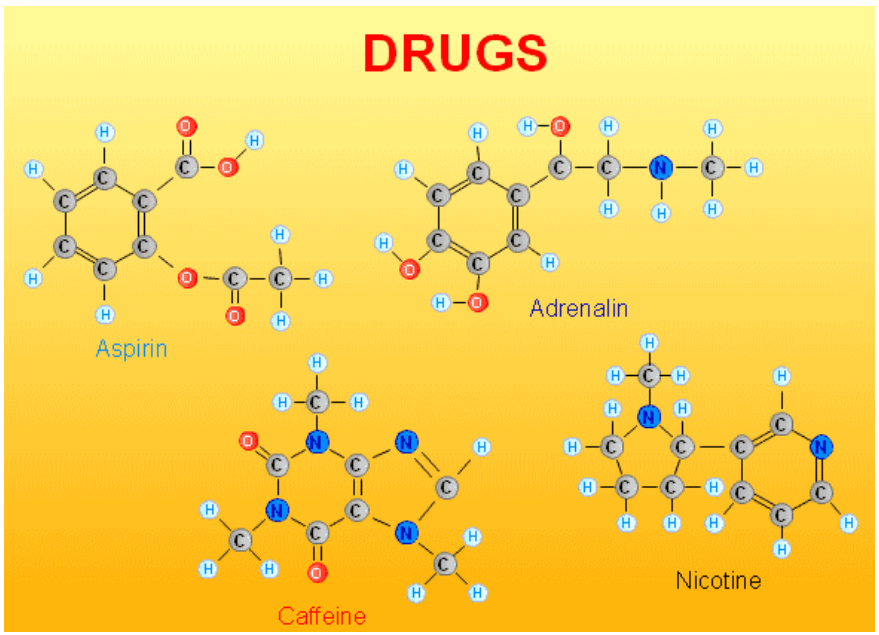
Talk Outline

Drug Discovery Funnel



Drug Discovery Cost

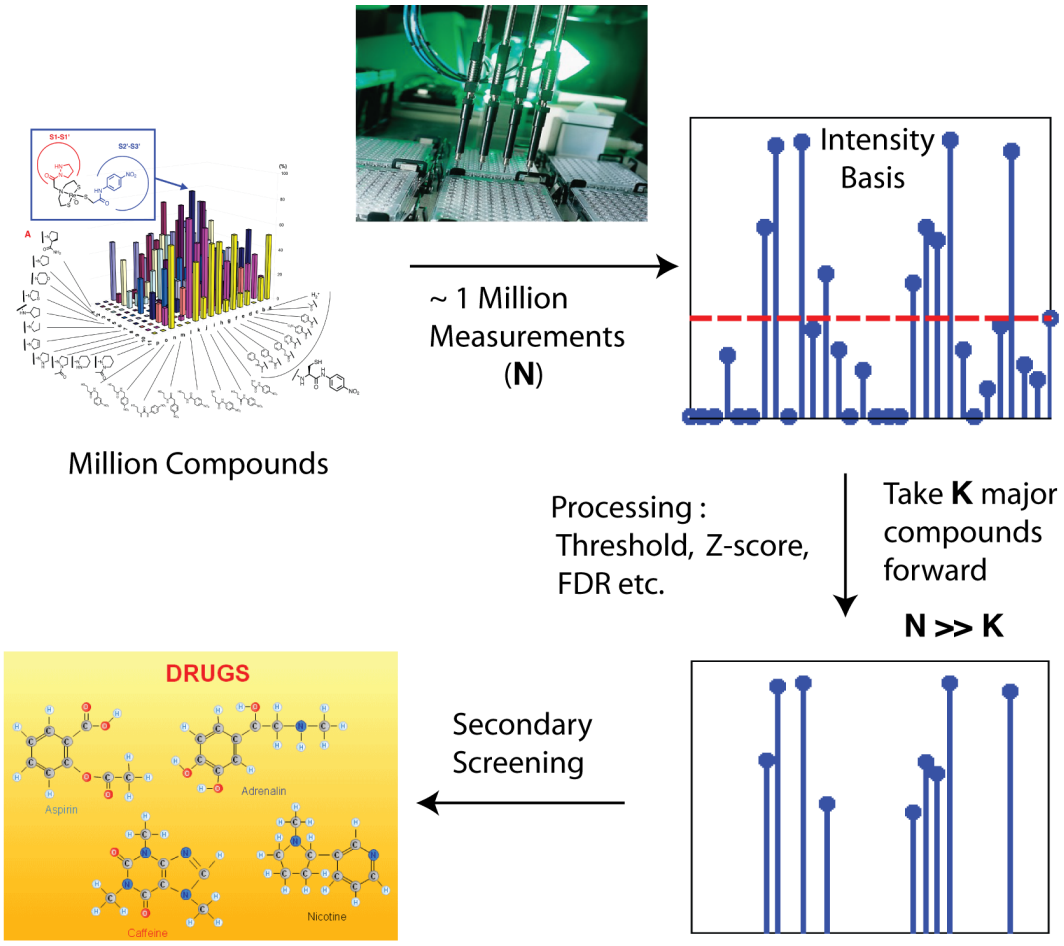
- Approx. Cost $\sim$ $800 million to bring a *new drug* to market[1]
- New drug = New Chemical Entity
- Each year, worldwide, only about 26 such drugs enter the market
- Millions of chemical compounds are tested to find them



[1] includes the cost of all drug development which did not result in a new drug

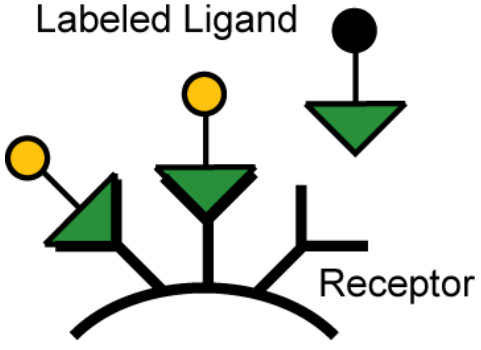
High-Throughput Screening (HTS)

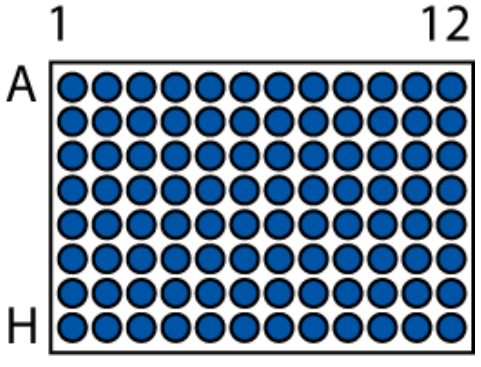
First step in drug discovery is High-Throughput Screening (HTS).



ABC of HTS

- Automation & high-throughput achieved through robotic liquid handling
- Biological Assay – Typically a biochemical binding event detected by an optical signal
- Chemical Library – thousands to millions of chemical compounds, available in pre-configured plates.
- Hit Rate – number of active compounds found in a screen (0.01 – 10%)





Pooling in HTS

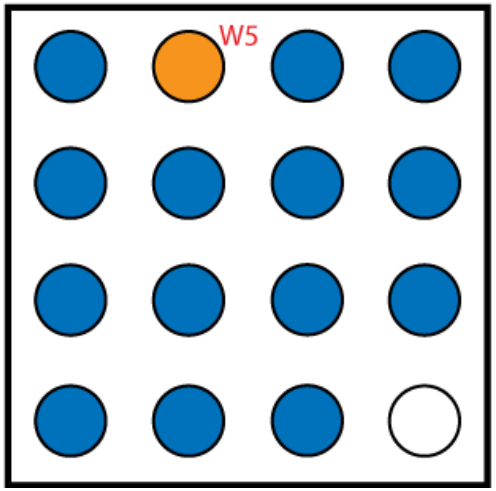
Comparison of *one compound*, *one well* and *pooled* HTS.

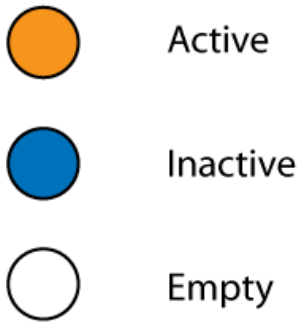
Current HTS

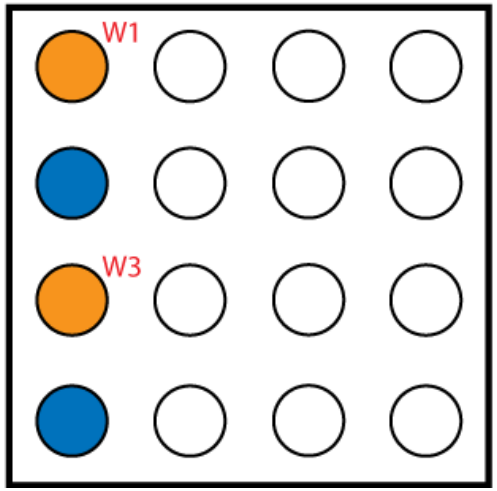
COMPOUNDS	1	2	3	4	5	6	7	8
WELLS	W1	W2	W3	W4	W5	W6	W7	W8
	9	10	11	12	13	14	15	
	W9	W10	W11	W12	W13	W14	W15	

Pooled HTS

WELLS	COMPOUNDS							
W1	1	3	5	7	9	11	13	15
W2	2	3	6	7	10	11	14	15
W3	4	5	6	7	12	13	14	15
W4	8	9	10	11	12	13	14	15



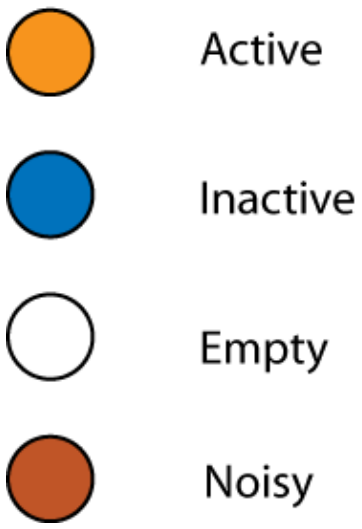


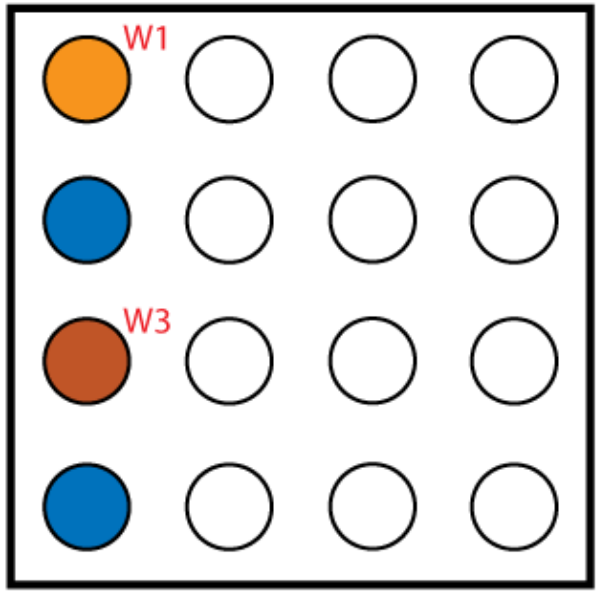


Multiple Items & Noisy Tests

Unique boolean tagging does not work when multiple active compounds or testing errors occur.

WELLS	COMPOUNDS							
W1	**1**	3	**5**	7	9	11	13	15
W2	2	3	6	7	10	11	14	15
W3	4	**5**	6	7	12	13	14	15
W4	8	9	10	11	12	13	14	15





Group Testing

Problem : Create pooling strategy that reduces tests, guarantees identification and corrects errors in testing.

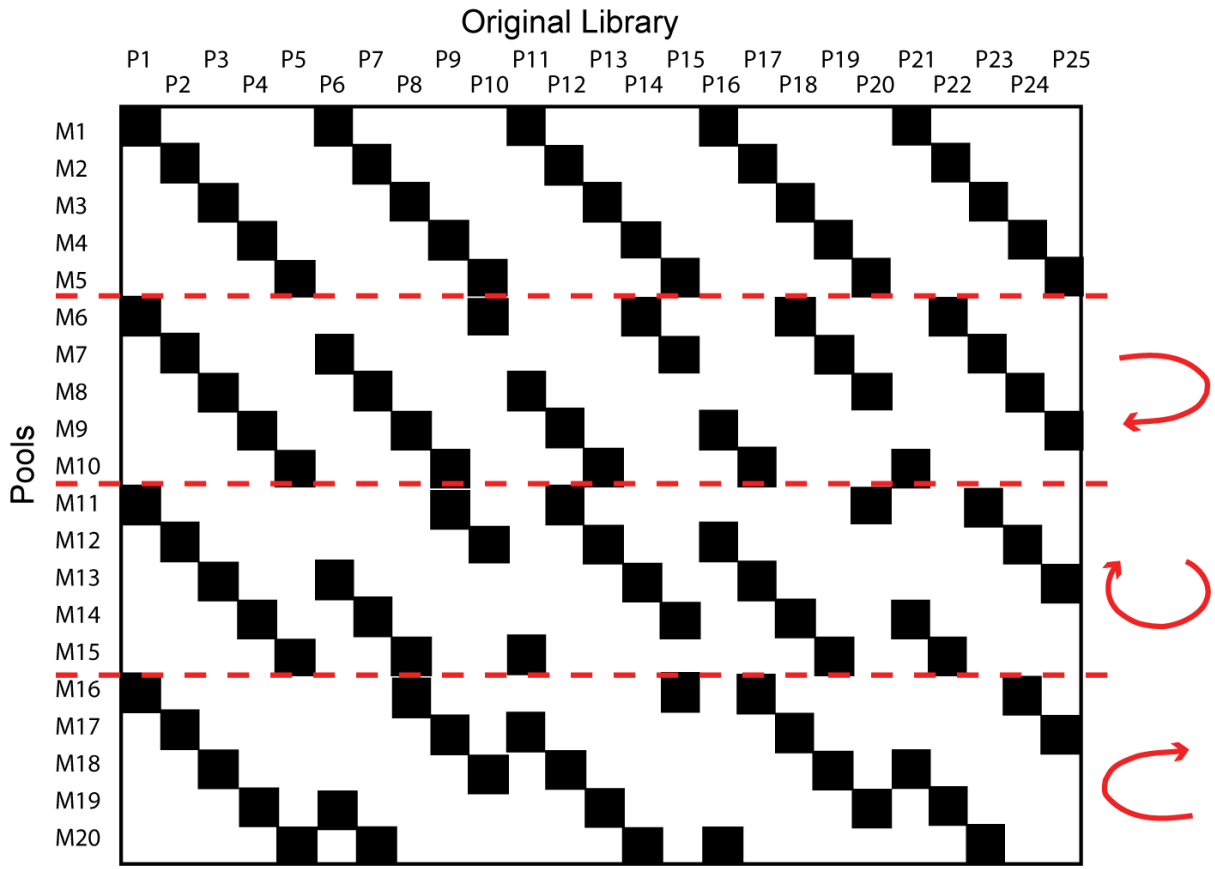
Solution : Group Testing [2]

- For n compound library
- With at most k active
- With at most E testing errors
- Design pooling strategy to guarantee the identification of k actives
- Design a decoding algorithm which works in the presence of E errors

[2] which means Compressive Sensing is around the corner

Pooling Design

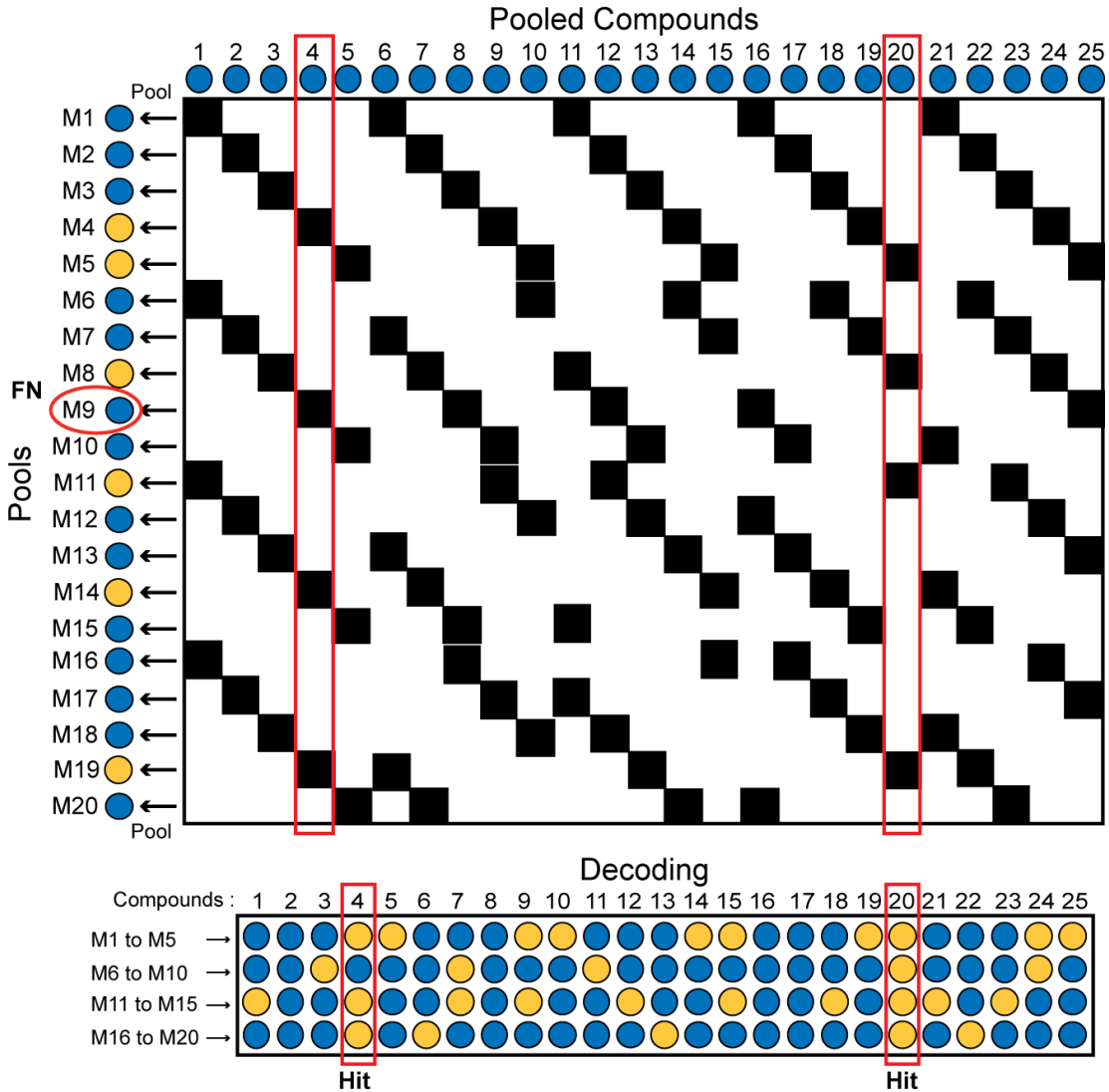
Example: Shifted Transversal Design (STD) of N.Thierry Mieg [3] for $n = 25$, $k = 2$, $E = 1$.



[3] shown to be equivalent to R. DeVore's *Deterministic Construction* (2007)

Decoding Algorithm

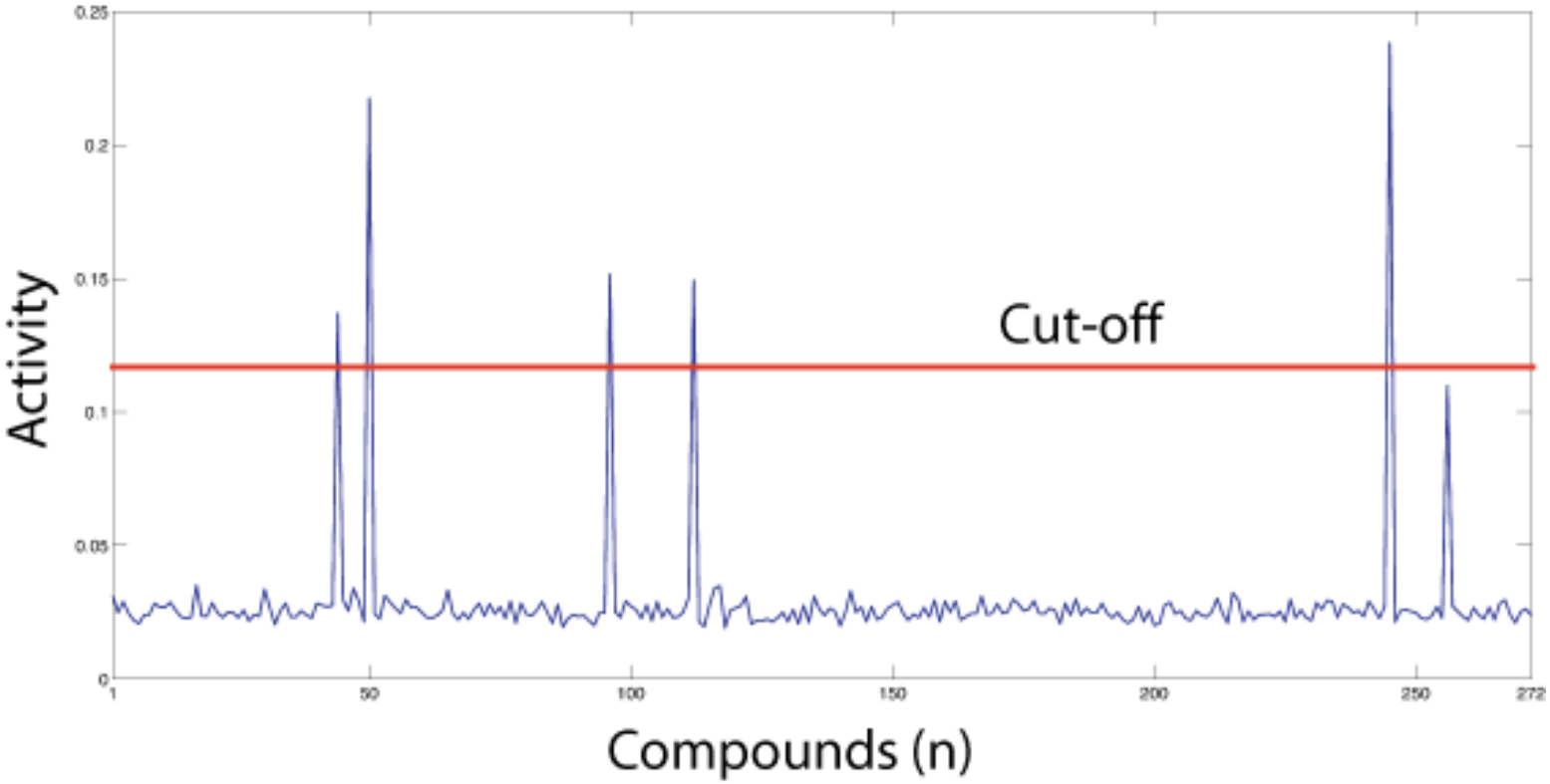
Choose a cut-off to reduce measurements to binary (*hit* or *miss*).[4]



[4] figures from **K. & Woolf**, *Curr. Op. in Drug Disc. & Dev*, *in press* 2009

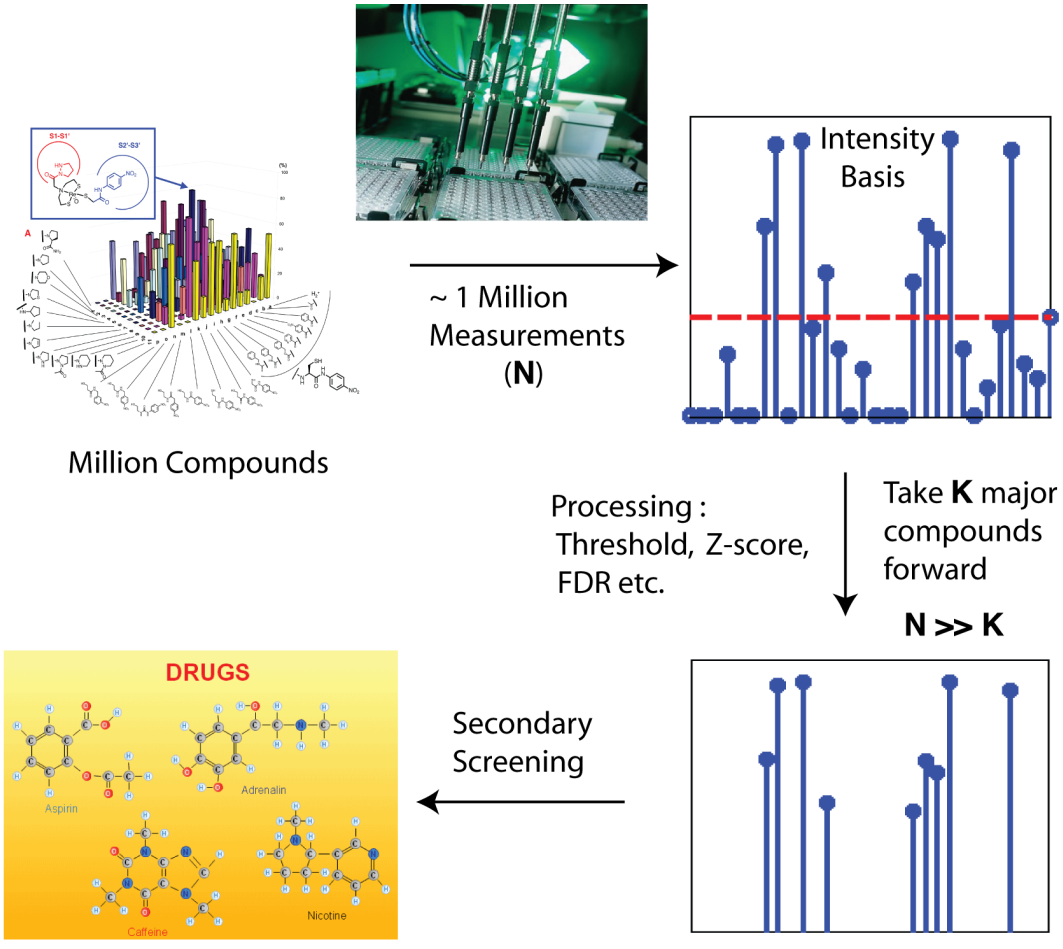
Quantitative Analysis of Pooling

- Quantitative information is present in measurements.
- Binary binning of data introduces false positive and false negative testing errors.
- Hard to choose cut-off for pooled measurements.



Compressive Sensing in HTS

Quantitative Analysis of Pooling is possible via Compressive Sensing.

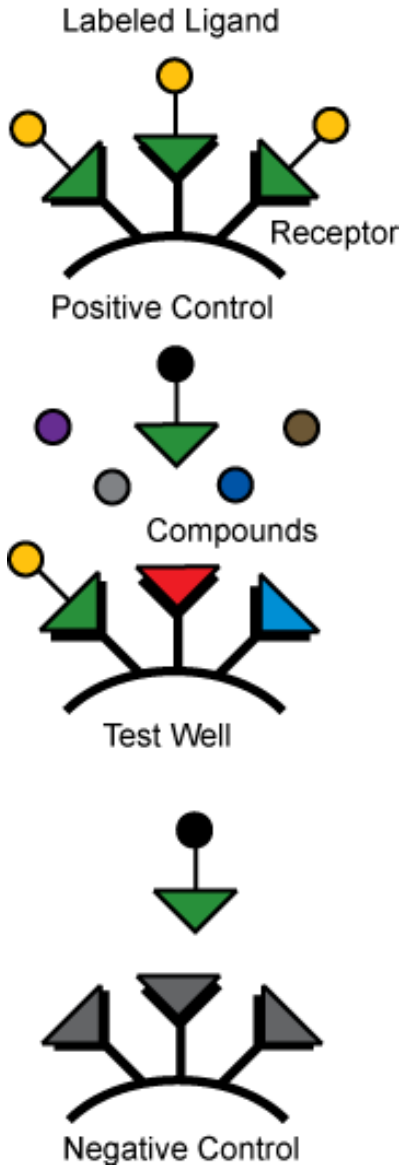


It is sparse but is it linear?

Biochemical Model for Pooling

Competitive binding assay.

- $R + D_1 \underset{k-1}{\overset{k1}{\rightleftharpoons}} C_1$
- $\vdots$
- $R + D_i \underset{k-i}{\overset{ki}{\rightleftharpoons}} C_i$
- $\frac{I_{\text{Test}} - I_{\text{NC}}}{R_{\text{Tot}}} \propto \frac{K_a[L]}{1 + K_a[L] + K_{a1}[D_1] + \ldots}$
- % Inhibition $= \frac{I_{\text{PC}} - I_{\text{Test}}}{I_{\text{PC}} - I_{\text{NC}}} \times 100$
  $= \frac{K_{a1}[D_1] + \ldots + K_{ai}[D_i]}{1 + K_a[L] + K_{a1}[D_1] + \ldots + K_{ai}[D_i]} \times 100$
- Assume : All drugs present in equal & excess conc.



Linear Model for Activity

$$y = \frac{(1 + K_a[L])}{[D]} \frac{\%I}{100 - \%I} = \sum_i K_i$$

y – modified measured quantity. K_a, $[L]$ and $[D]$ are known.

Linear Algebra Problem : $y = MK$

QUAPO : Quantitative Analysis of Pooling in HTS

QUAPO

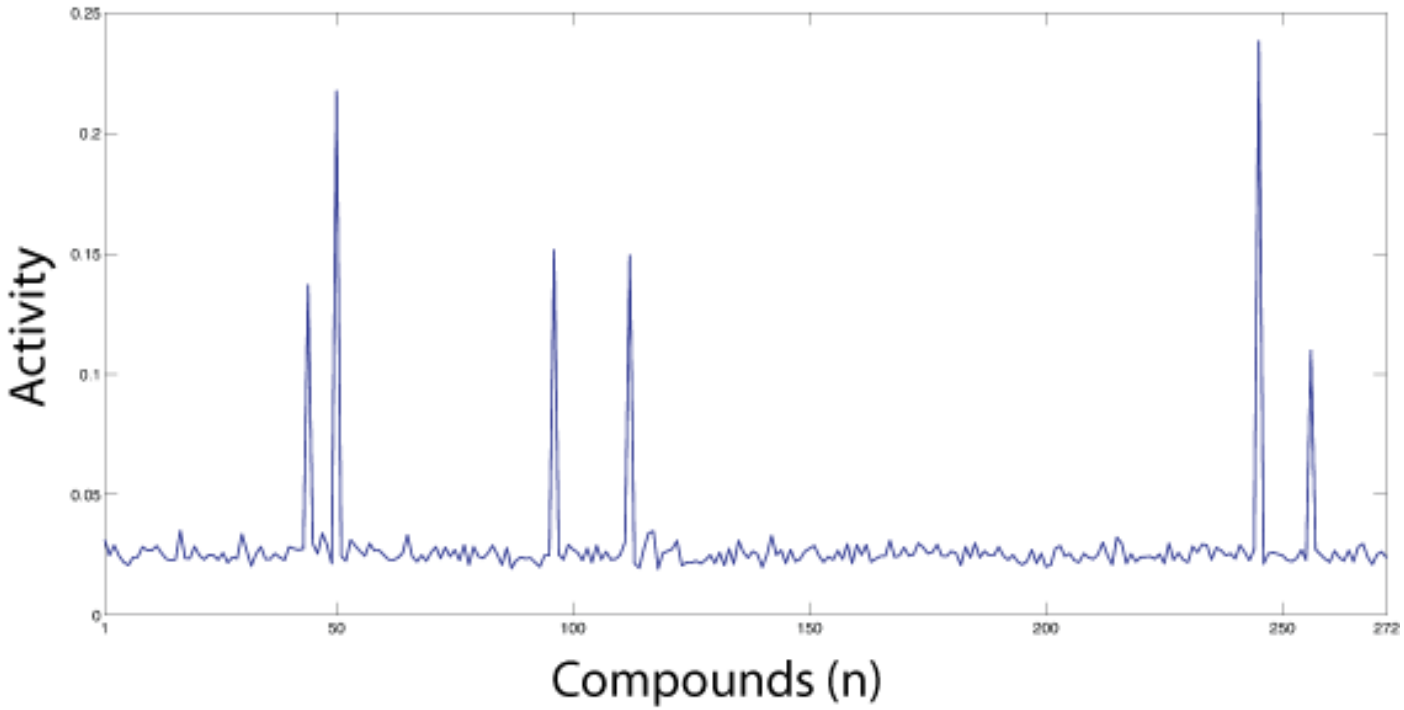
- Sparsity : Most compound activities (K_a's are close to zero (inactive).
- Linearity : Measured quantity maps linearly to compounds activity (with reasonable approximations).
- Solve

$$\min_x ||x||_1 \text{ subject to } ||\Phi x - y||_2 \leq \epsilon$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{pmatrix}_{t\times 1} \propto \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 1 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ & \vdots & & \vdots & \vdots & \\ 1 & 0 & \dots & 1 & \dots & 0 \end{bmatrix}_{t\times n} \times \begin{pmatrix} K_{a1} \\ K_{a2} \\ \vdots \\ K_{ai} \\ \vdots \\ K_{an} \end{pmatrix}_{n\times 1}$$

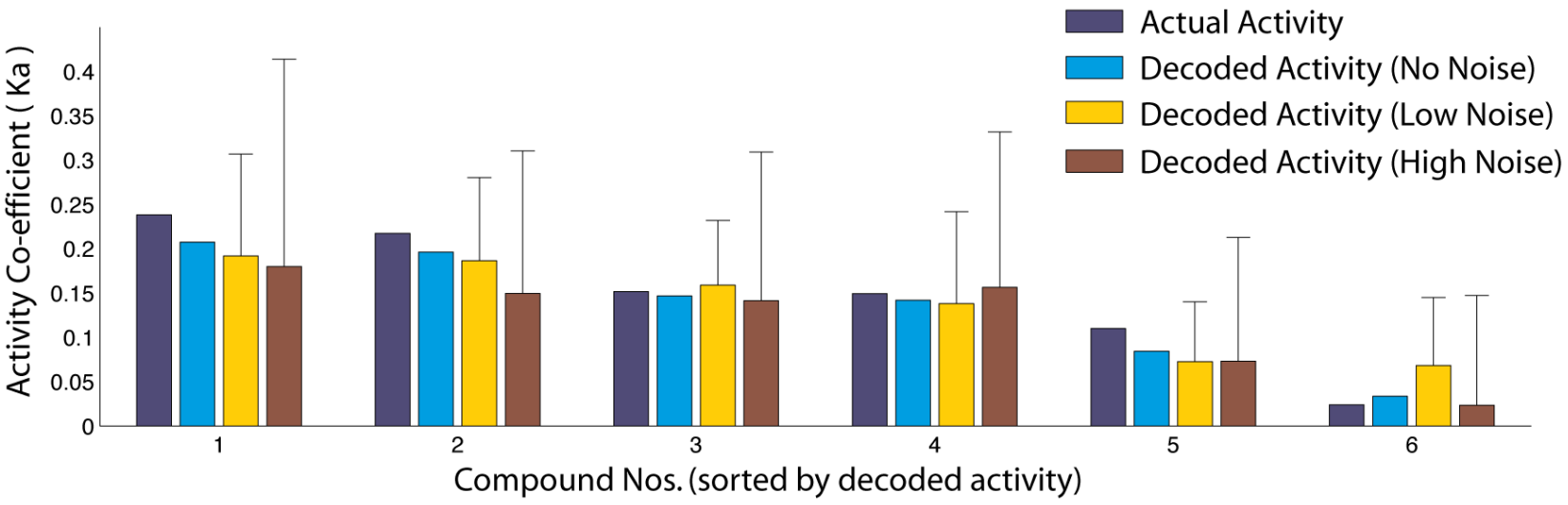
Small Library Simulation

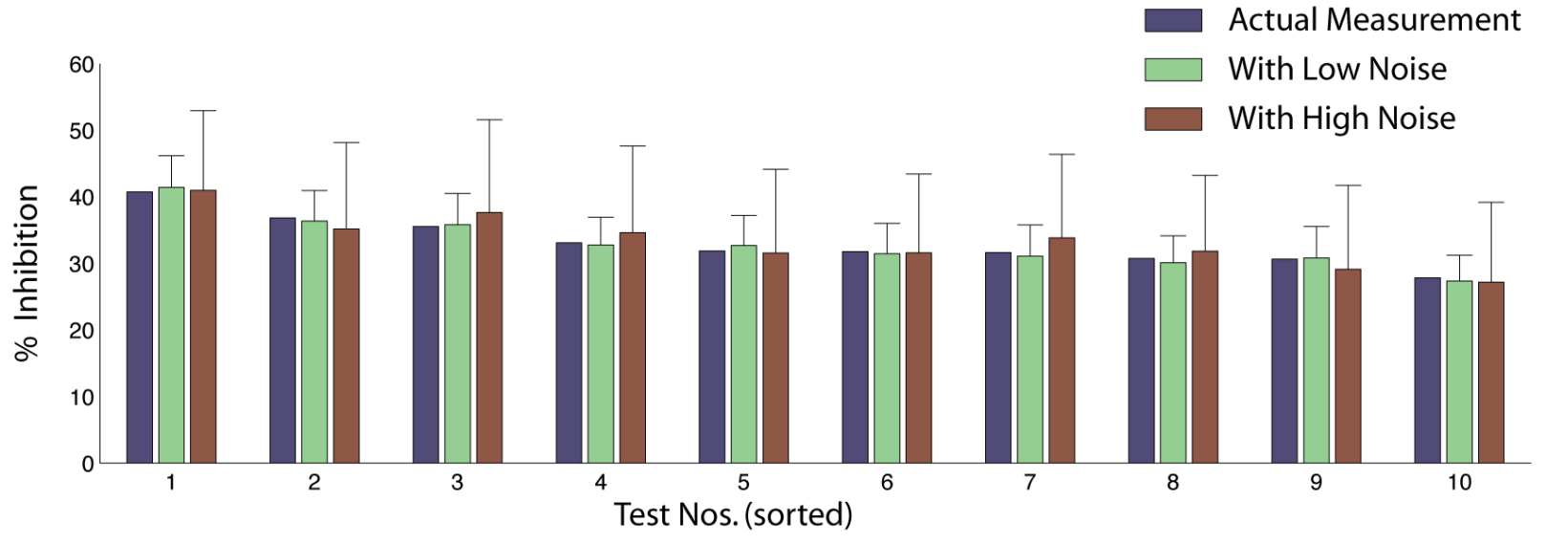
- Synthetic Screen : small molecule ligands for formylpeptide receptor (FPR) with 6 showing activity.[5]
- STD($n = 272, d = 3, e = 0\%, r = 10$) required $m = 116$ tests.
- $y = \frac{(1+K_a[L])}{[D]} \frac{\%I}{100-\%I} = \sum_i K_i$
- $[L] = 1.5\mu$M, $1/K_a = 3\mu$M and $[D] = 1.5\mu$M



[5] Edwards *et. al.*, Nature Protocols (2006)

Small Library : QUAPO Result





Challenge 1 : Pooling Design (Φ) Constraints

With existing HTS technology, easiest to use Sparse Binary Matrices (STD/DeVore matrix) or Expander Graphs.

Mixing Constraint

- Compound concentration must be detectable in physiological range.
- Ionic strength of mixture must be low to prevent precipitation or changes to biological target.
- The assay must be reasonably simple to physically construct.

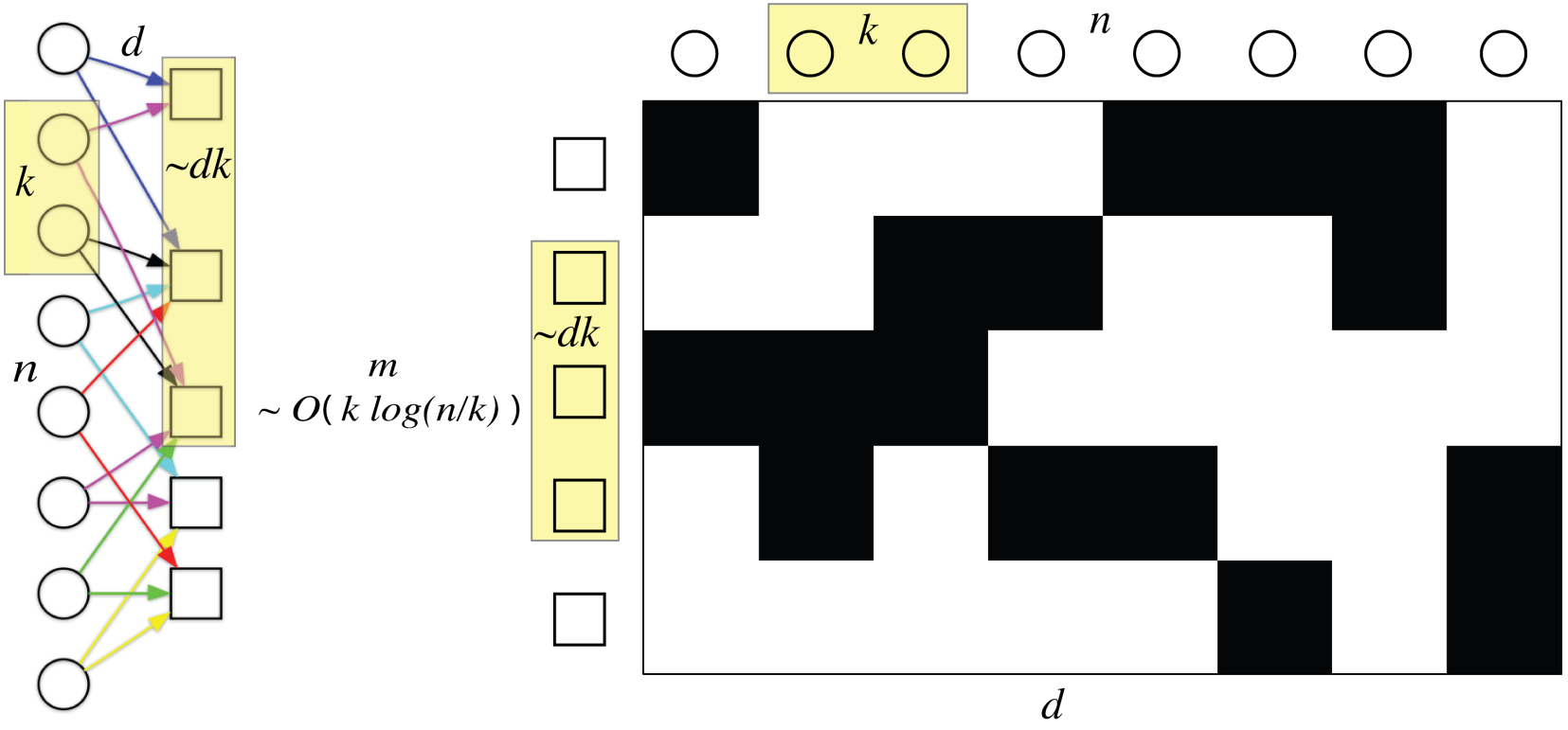
Challenge 1

Row weight of Φ is tightly capped.

Simple Heuristic : Not more than ~ 10 compounds can be pooled in a test.

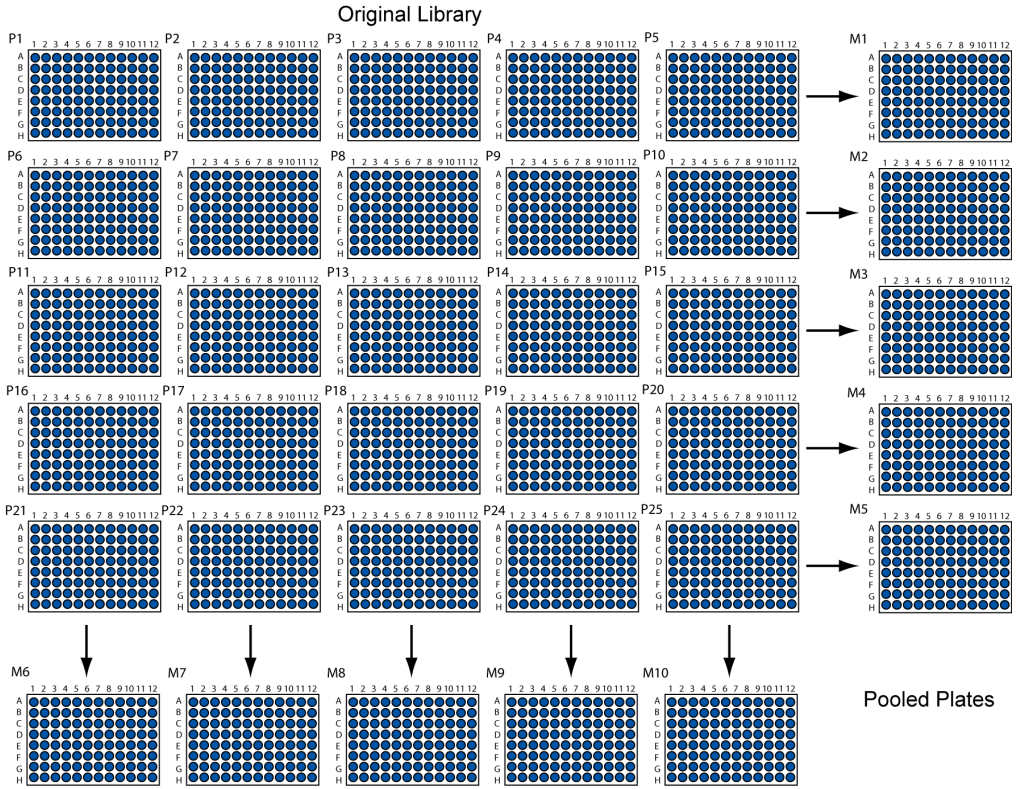
Really Sparse Matrices

Row weight cap implies that limited compression can be achieved.



Challenge 2 : Liquid Handling Issue

Pooling at the level of individual compounds is hard and/or costly.

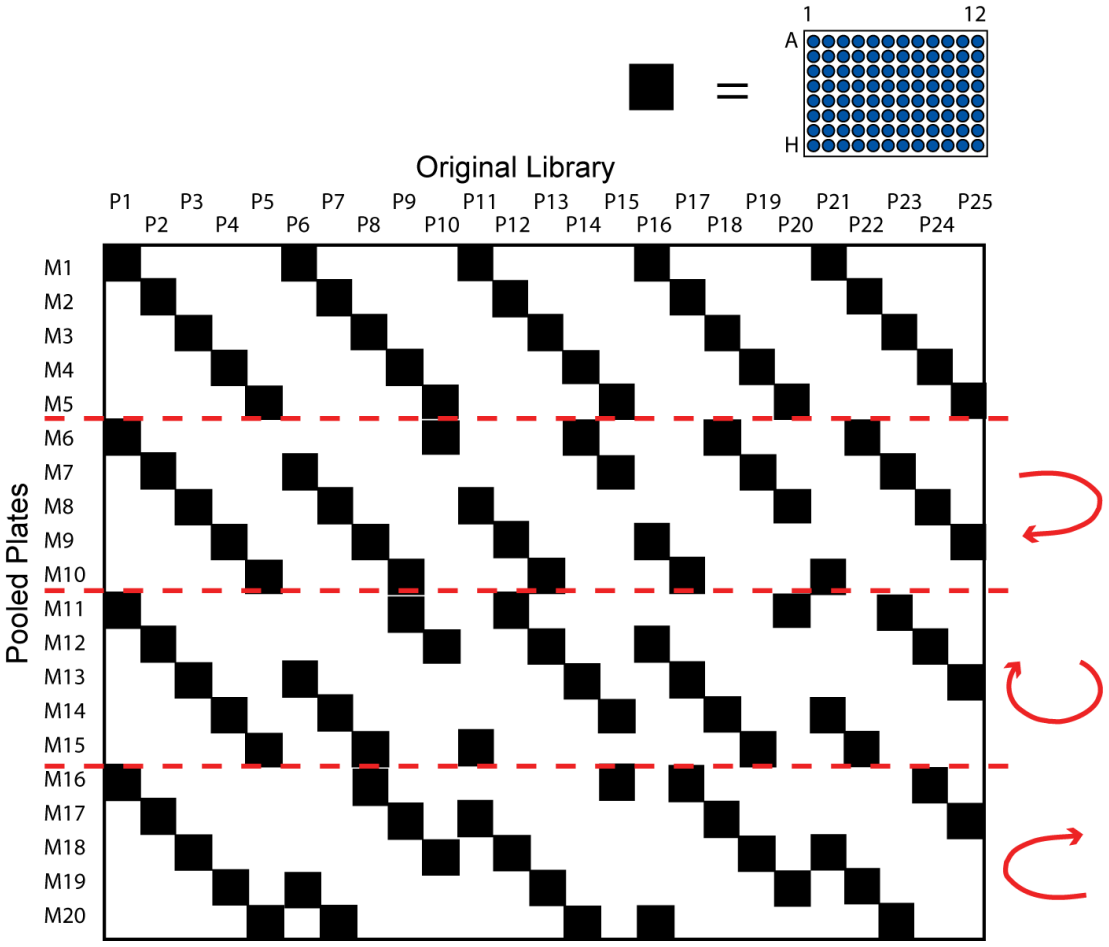


Challenge 2

Original Library is subdivided into mutually exclusive blocks.

Challenge 2 : A Simple Solution

Φ must be designed for smaller $\hat{n}$ and repeated in blocks on whole library n



Challenge 3 : Measurement Error

- CS algorithms promise to handle additive noise.
- Small volumes and automation mean erasures are possible.
- Given Challenges 1 & 2, promising compression *and* error-correction might be difficult.

Challenge 3

Erasures of measurements are possible

Challenge 4 : Non-additive behavior

- Synergy : pooled compounds react or aggregate to produce a *hit*
- Antagonism : pooled compounds block each other out

Solution: Challenges can be treated as bugs or features.

- Bug : make designs more robust to these *errors*
- Feature : ability to detect mutli-compound drugs or drug cocktails

Challenge 4

Algorithms to handle non-additive behavior

Advances in Pooling

Theme [6]

- Use *chemical structure information* about compounds while designing pools
- Simulations to *predict* probabilities of synergy or antagonism
- Simulations to evaluate *average-case* pooling design properties (theorems give worst-case bounds)
- Bayesian Decoders to evaluate various scenarios of compound interaction

[6] Will take more (compute) time

Summary

Take Home Points

- Current HTS strategies have hit a wall.
- Ever increasing compound collections and explosion of biological targets from genomics need a new approach.
- Age of *multi-compound*, *multi-target* therapeutics requires a paradigm shift in HTS.
- Pooling designs have the potential to be that change.
- Compressive Sensing can help make HTS quantitative (QUAPO).
- Lots of interesting (theory) problems need to be solved to make this approach practical.
- Currently implementing experimental validation at HTS facility in Univ. of Michigan.

Acknowledgments

At the University of Michigan :

- Peter Woolf
- Anna Gilbert
- Paul Shearer
- Systems Biology Group
- Center for Chemical Genomics

Questions . . . Comments . . . Suggestions

Thank You