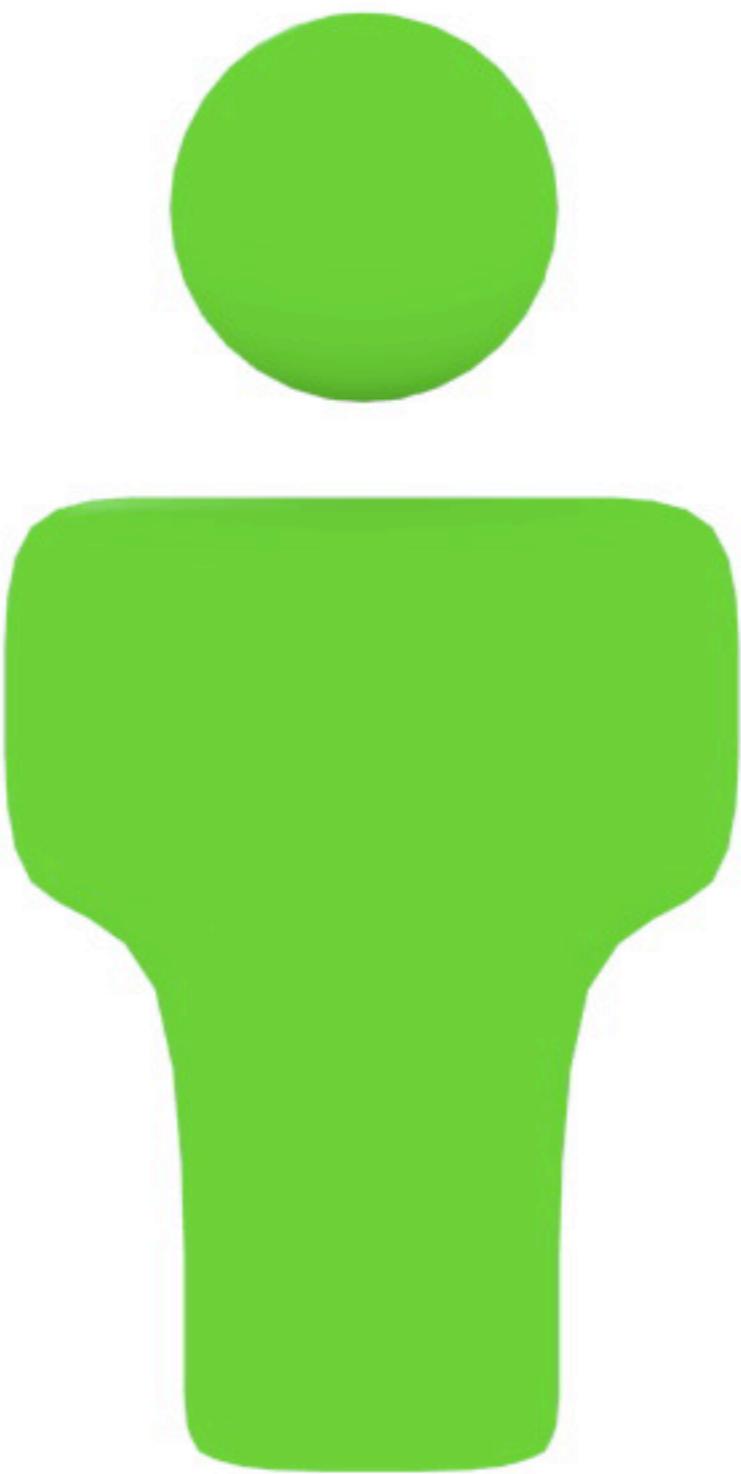
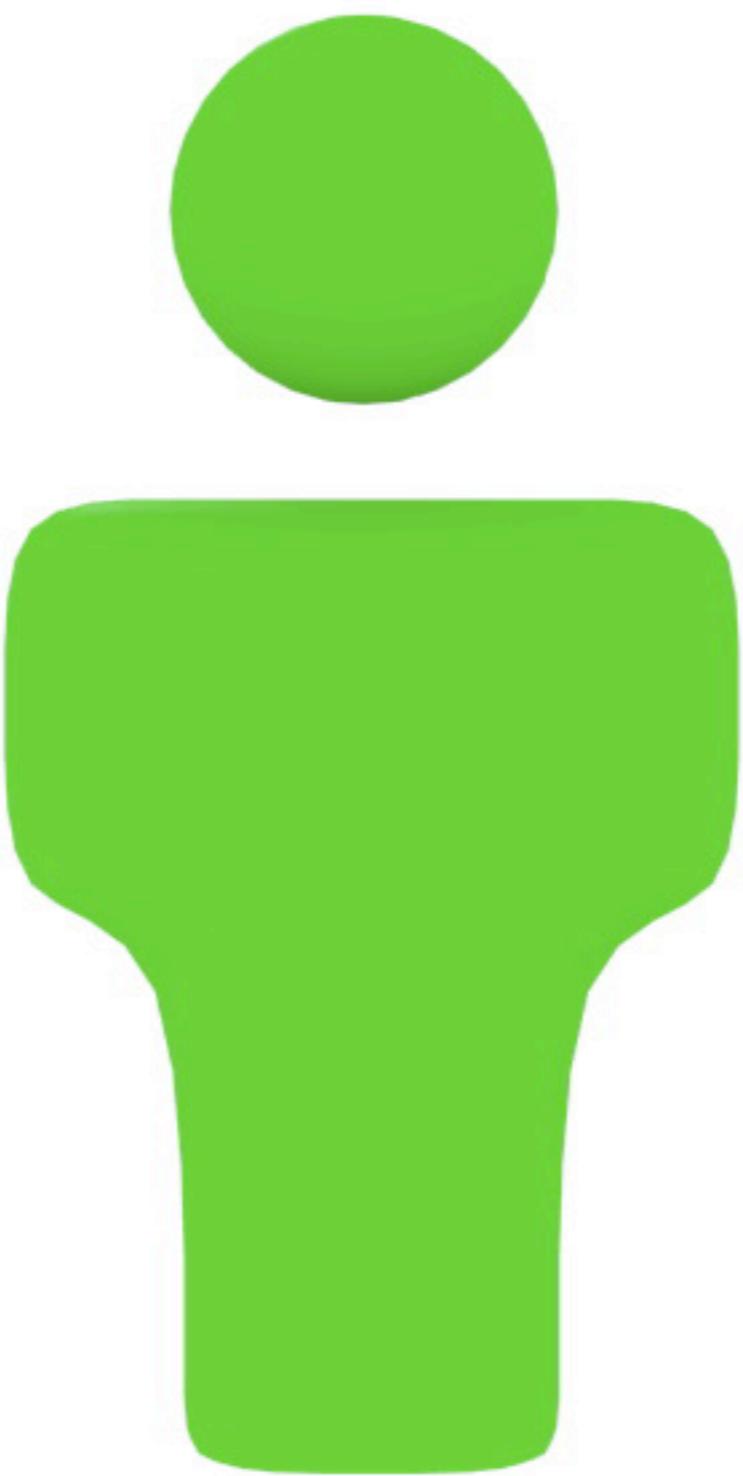

Connectivity Map

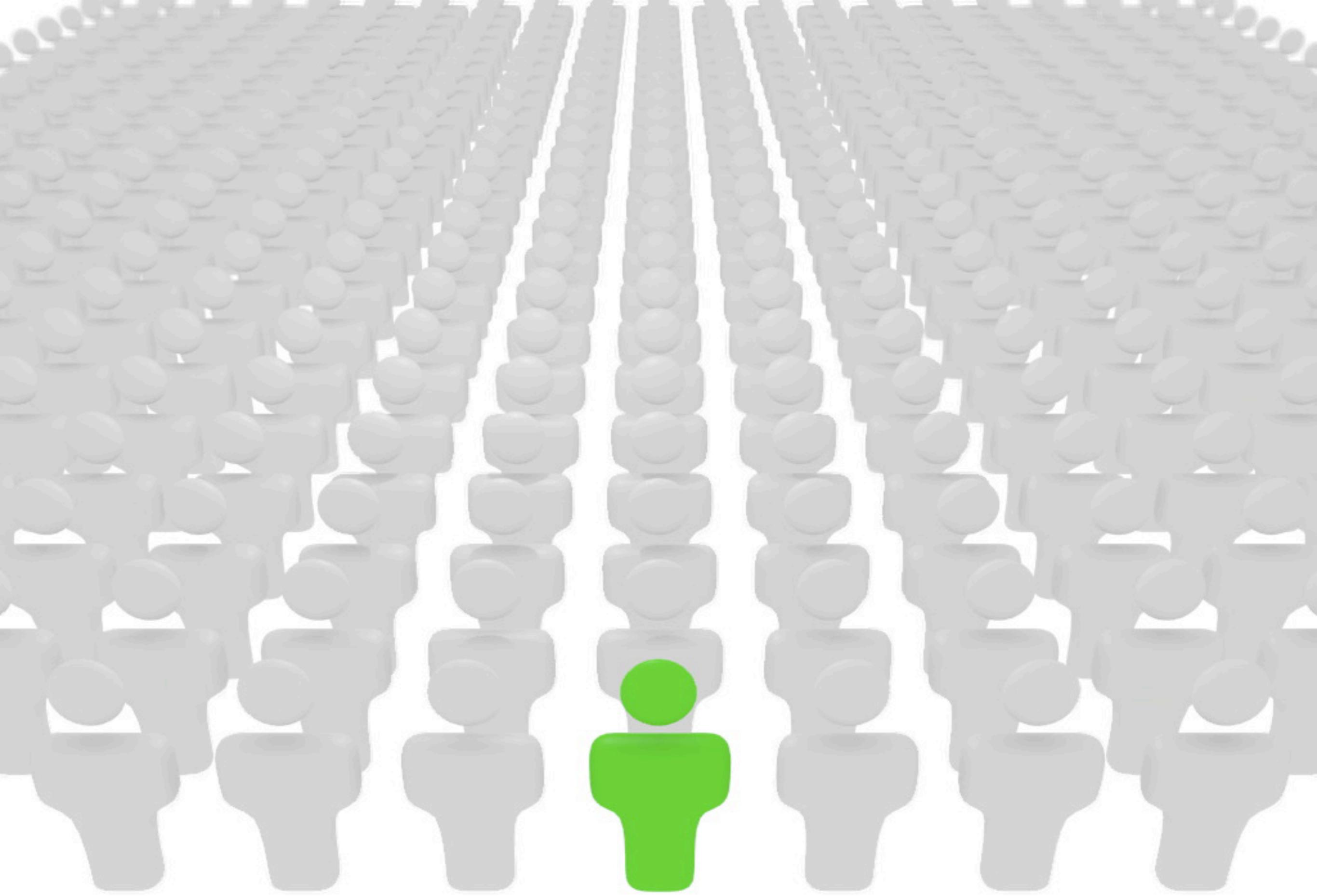
Gene Expression at Scale

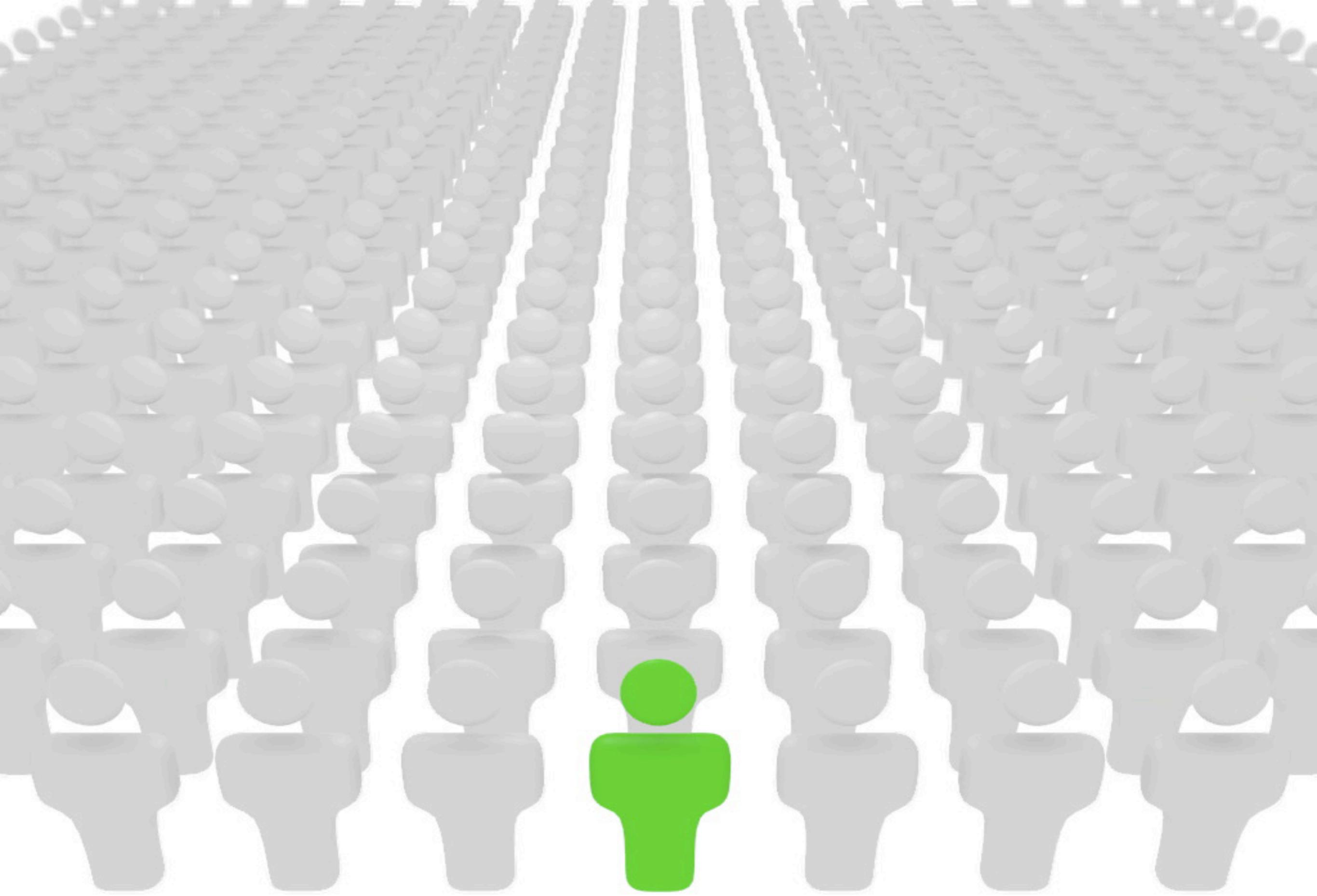


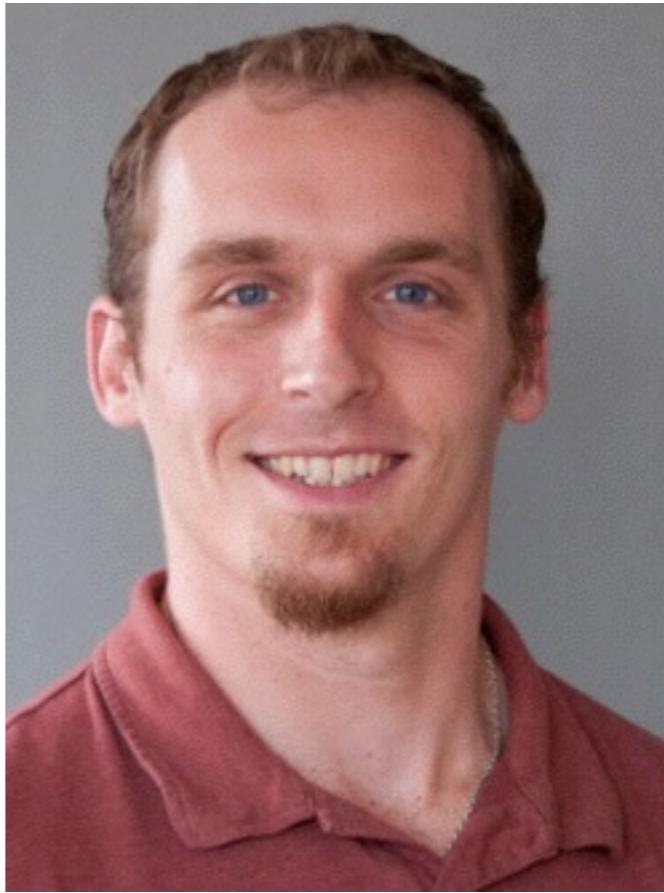












Corey Flynn

cflynn@broadinstitute.org

@CoreyJFlynn

Gene Expression

a common language for biology

Gene A B C D E



Disease



Normal



Drug



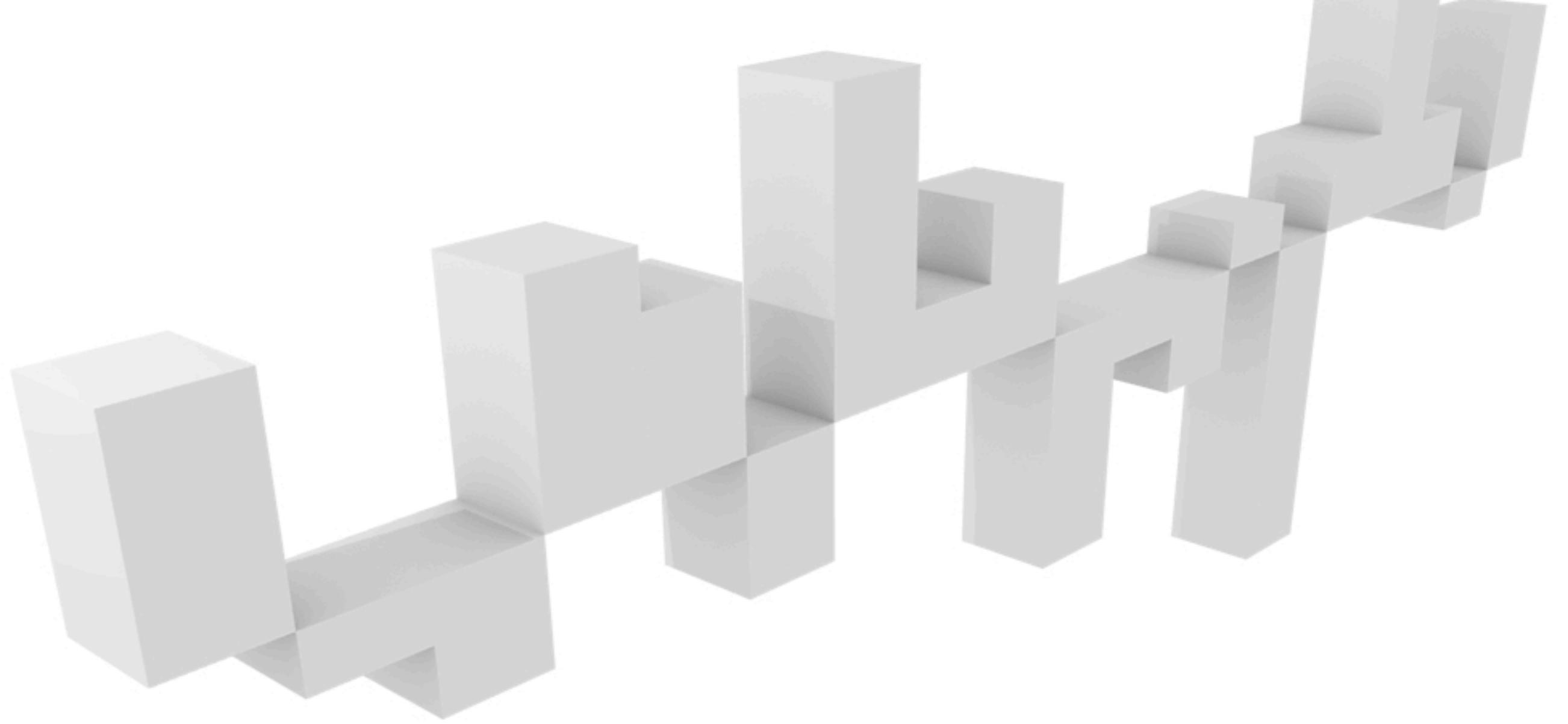
Disease

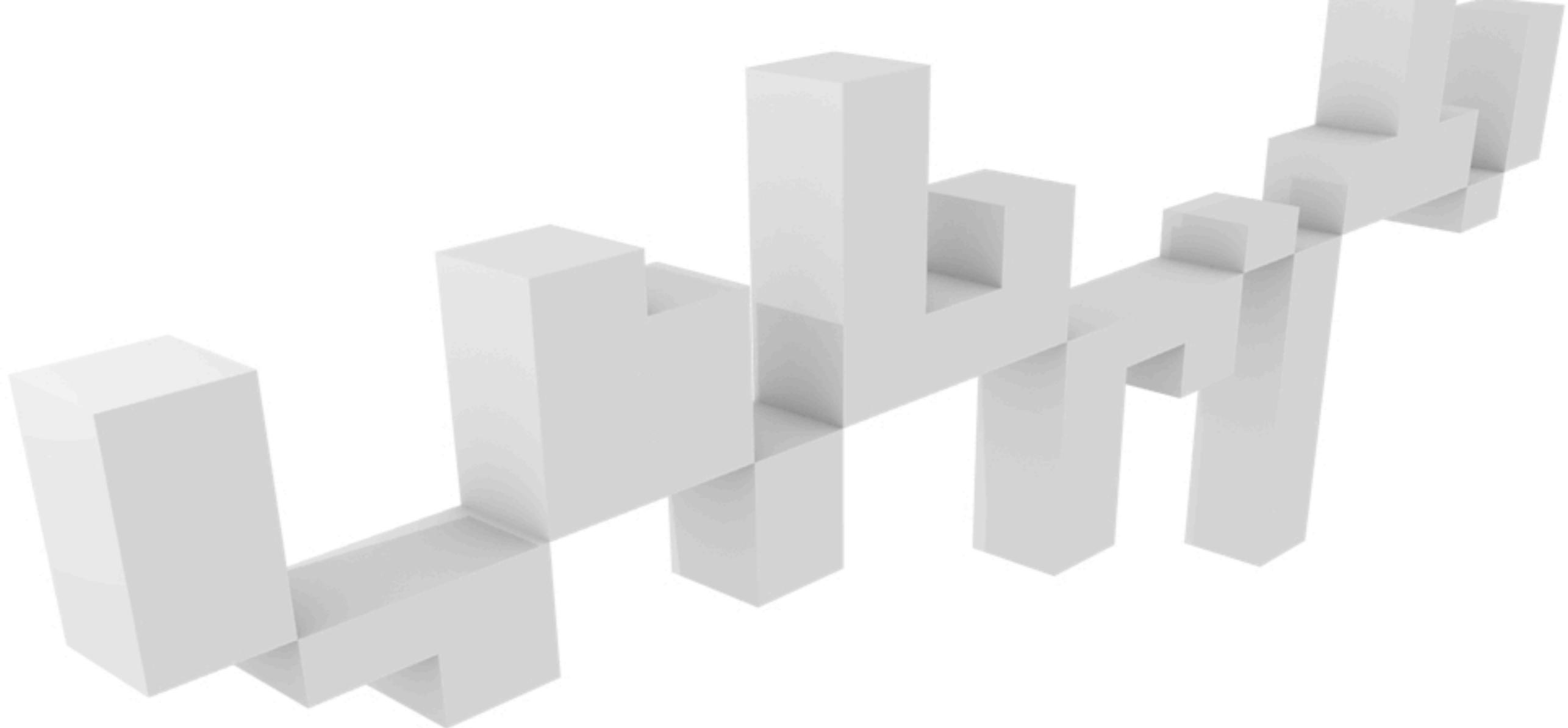


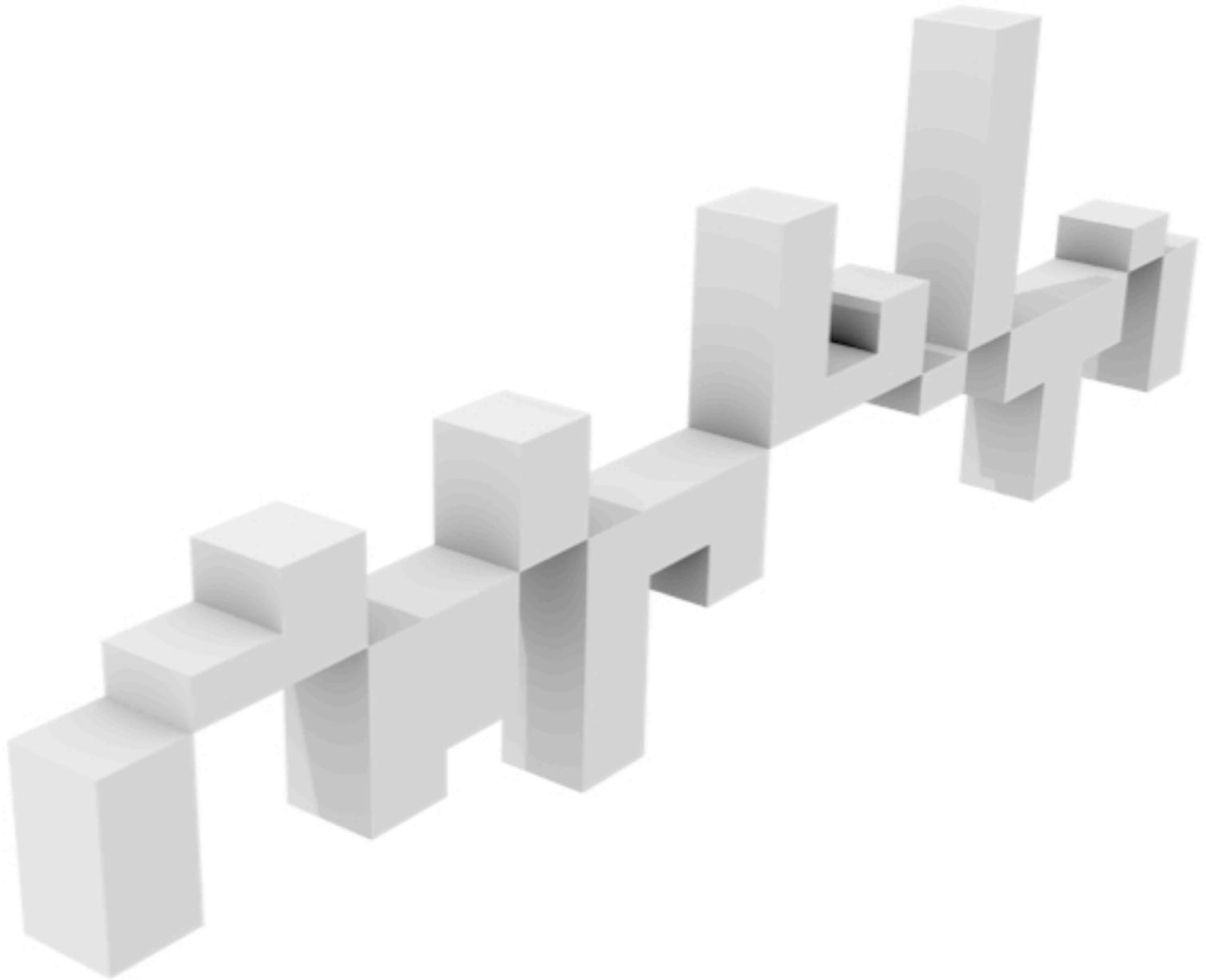
Normal

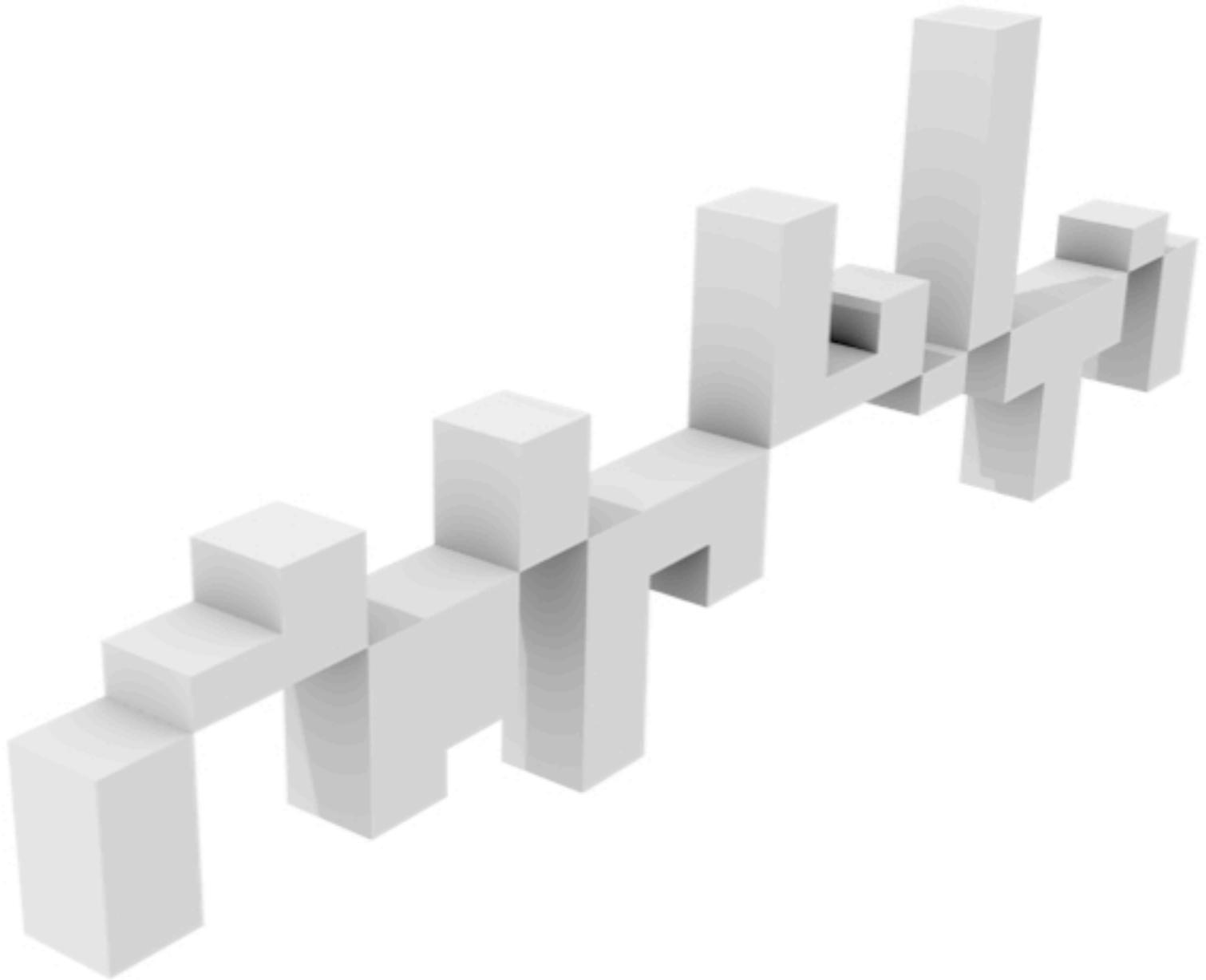


Drug









2006



**~7,000 experiments
Over 19,000 registered users
Cited by over 1,200 scientific reports**

2006



2014

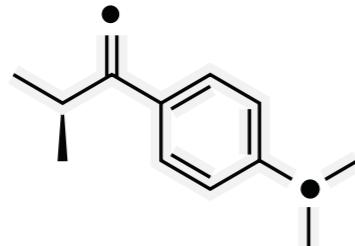






Connectivity Map Dataset

1.4 million gene expression signatures



12,488 Compounds

- FDA approved drugs
- Bioactive tool compounds
- Screening hits



3,800 Genes (shRNA & cDNA)

- Targets/pathways of approved drugs
- Candidate disease genes
- Community nominations



15 Cell types

- Banked primary cell types
- Cancer cell lines
- Primary hTERT-immortalized
- Patient-derived iPS cells
- Community nominated

Connectivity Map (CMap)

a resource of perturbational profiles

Connectivity Map (CMap)

a resource of perturbational profiles

Aim 1

Construct a **comprehensive reference database** of perturbational signatures

Connectivity Map (CMap)

a resource of perturbational profiles

Aim 1

Construct a **comprehensive reference database** of perturbational signatures

Aim 2

Develop **algorithms and software** to make these data and results accessible to the entire biomedical community

Connectivity Map (CMap)

a resource of perturbational profiles

Aim 1

Construct a **comprehensive reference database** of perturbational signatures

Aim 2

Develop **algorithms and software** to make these data and results accessible to the entire biomedical community

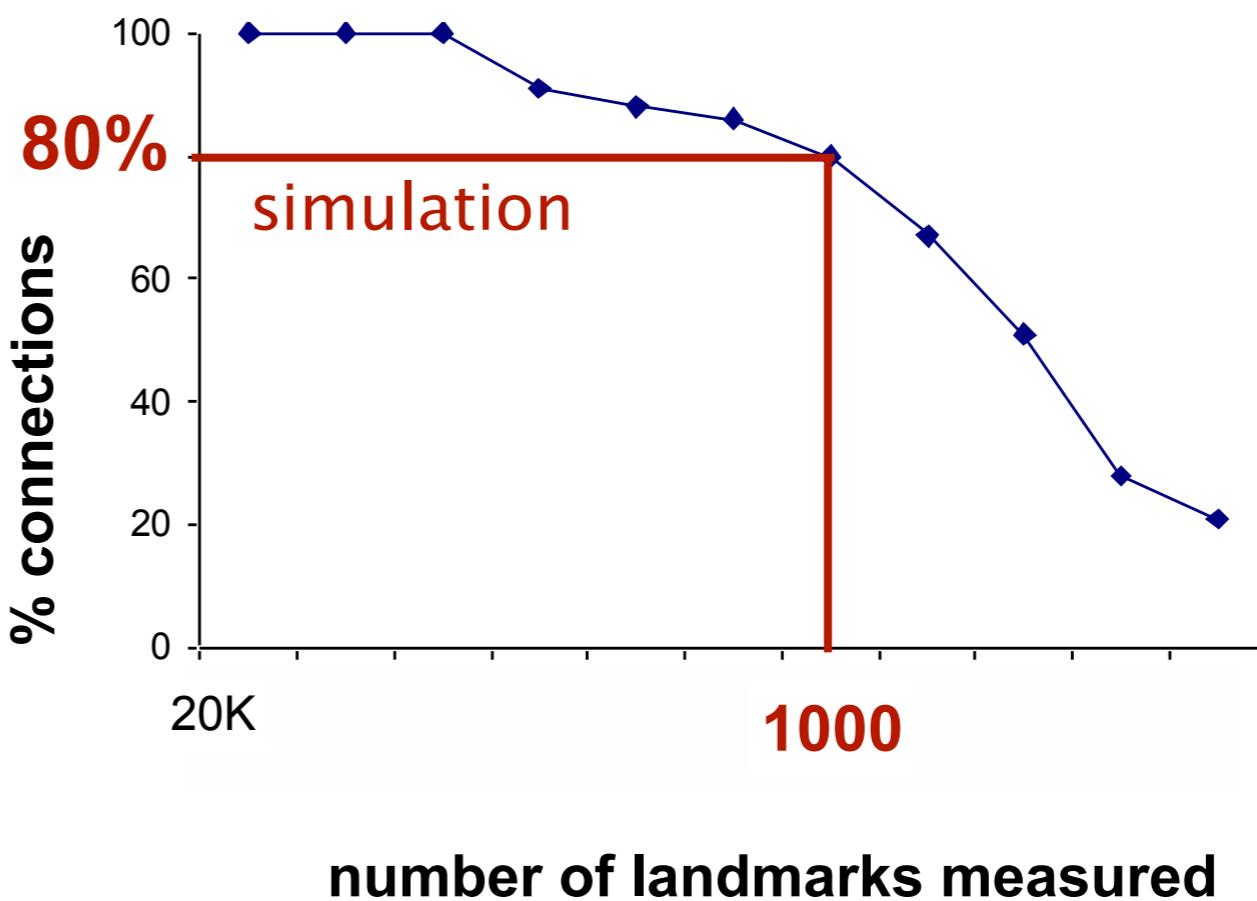
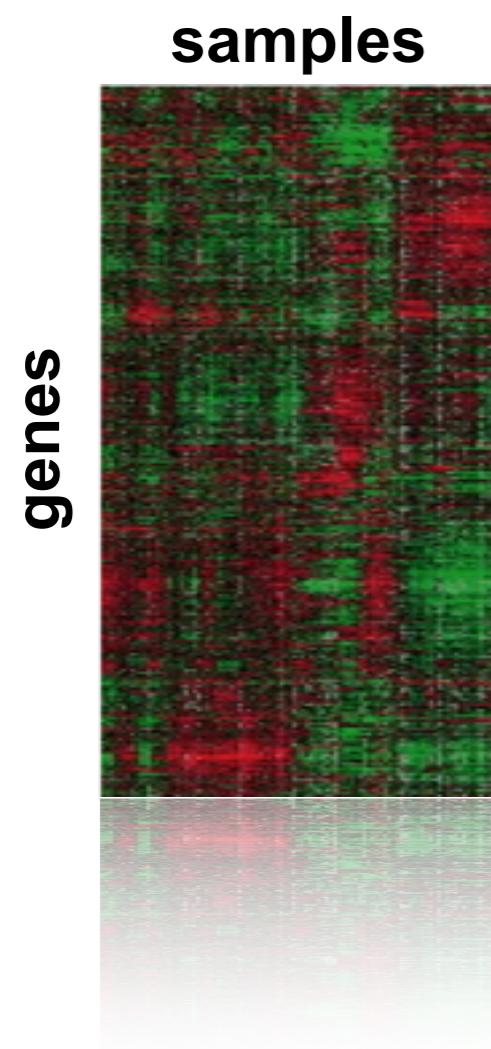
Aim 3

Utilize these resources to make **biological and therapeutic discoveries**

L1000 ASSAY

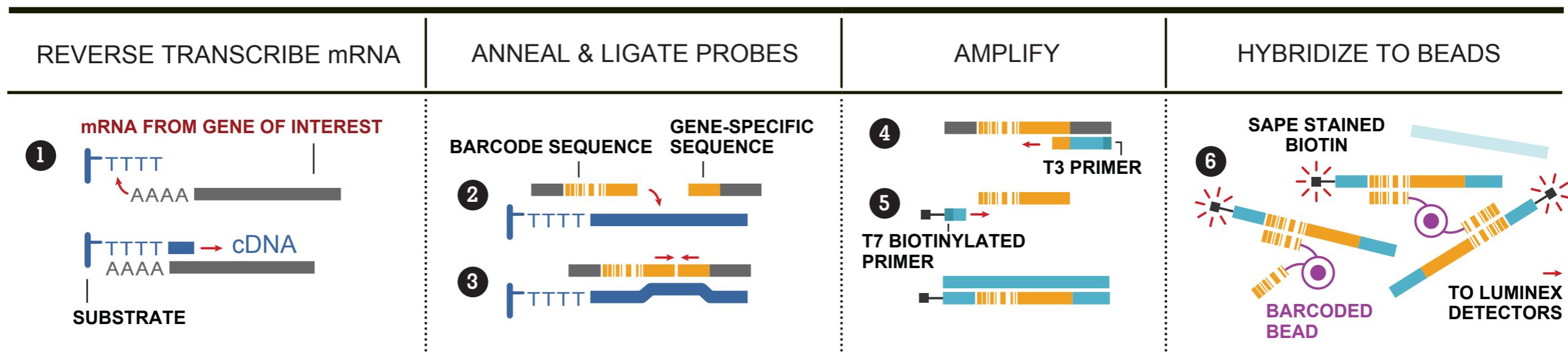
Reduced representation of transcriptome using 'landmarks' to infer the rest

Reduced representation
of transcriptome
“Landmark genes” + Computational inference model = Genome wide expression



The L1000 Assay

ligation-mediated 1000-plex amplification

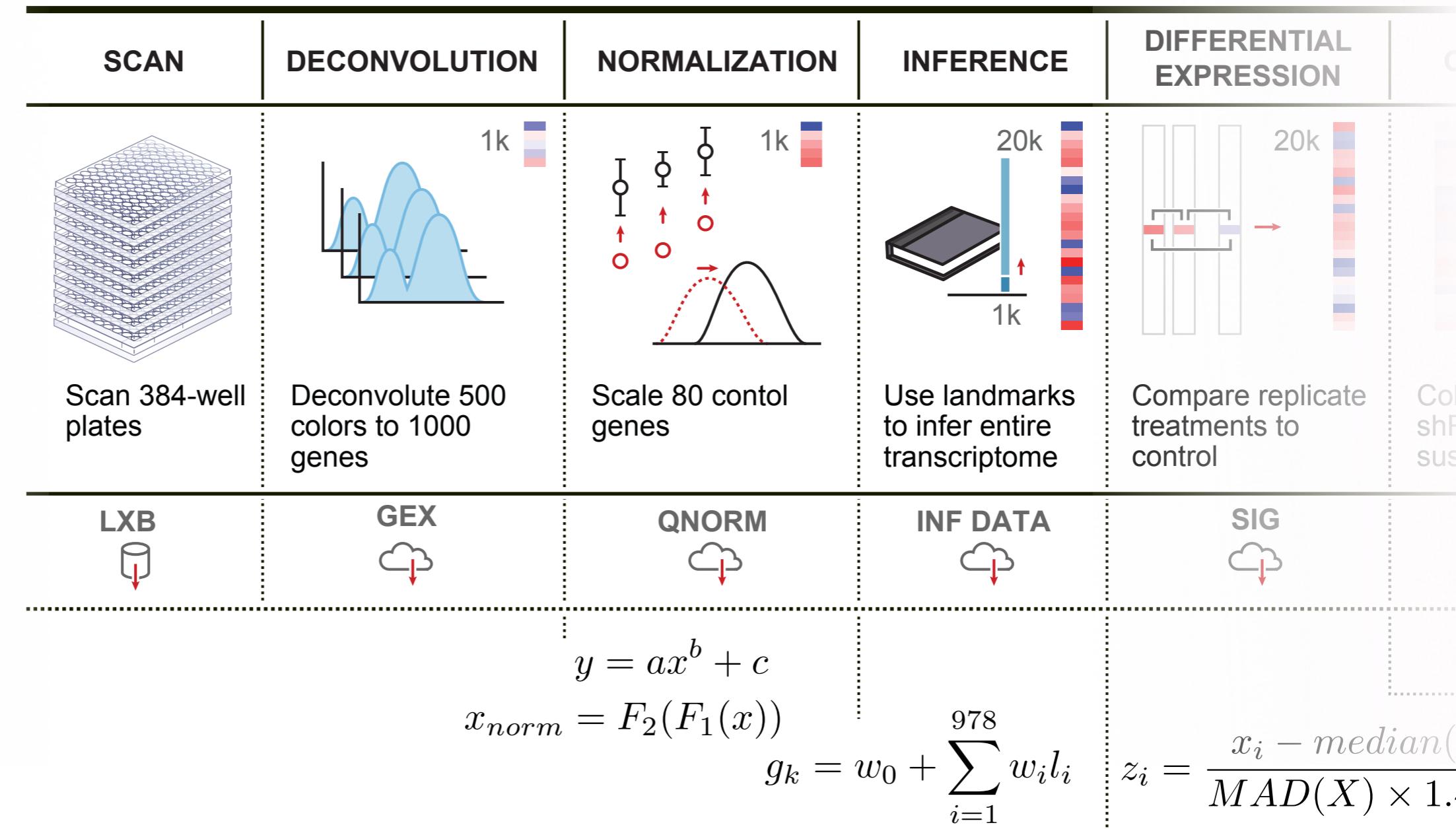


Measure 1,000 transcripts with 500-color Luminex
beads

384-well plate format compatible with HTS
Cell lysates (not purified RNA)
Dramatic cost reduction

Processing L1000 data

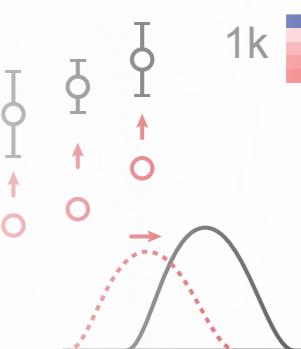
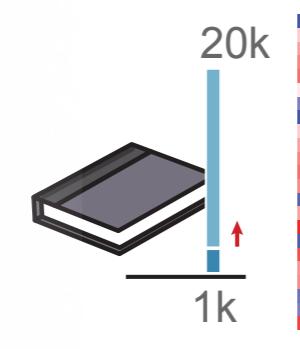
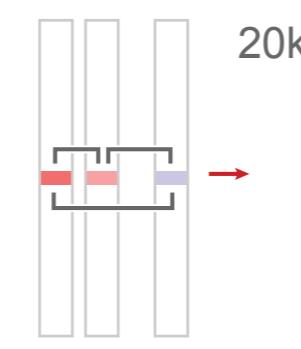
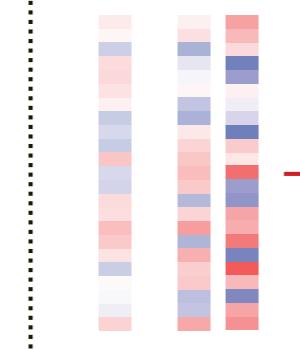
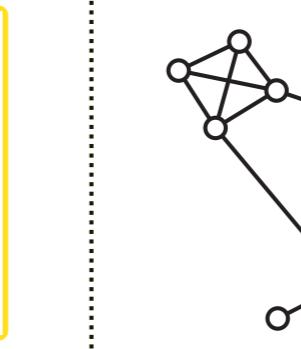
Data flow: from scans to signatures



$$MODZ = \frac{w_a \cdot z_a + w_b \cdot z_b}{w_a + w_b + w_c}$$

Processing L1000 data

Data flow: from scans to signatures

N	NORMALIZATION	INFERENCE	DIFFERENTIAL EXPRESSION	CONSENSUS	RELATIONSHIPS
	 <p>Scale 80 control genes</p>	 <p>Use landmarks to infer entire transcriptome</p>	 <p>Compare replicate treatments to control</p>	 <p>Collapse multiple shRNA to consensus signatures</p>	 <p>Compute correlation across entire signature data set</p>
	QNORM INF DATA	SIG CGS			GRAPH
$y = ax^b + c$ $rm = F_2(F_1(x))$ $g_k = w_0 + \sum_{i=1}^{978} w_i l_i$		$z_i = \frac{x_i - \text{median}(X)}{\text{MAD}(X) \times 1.4826}$			$P_{hit}(S, i) = \sum_{\substack{g_j \in S \\ j \leq i}} \frac{ r_j ^p}{N_R}$ $P_{miss}(S, i) = \sum_{\substack{g_j \notin S \\ j \leq i}} \frac{1}{N - N_H}$ $P_{hit} - P_{miss}$

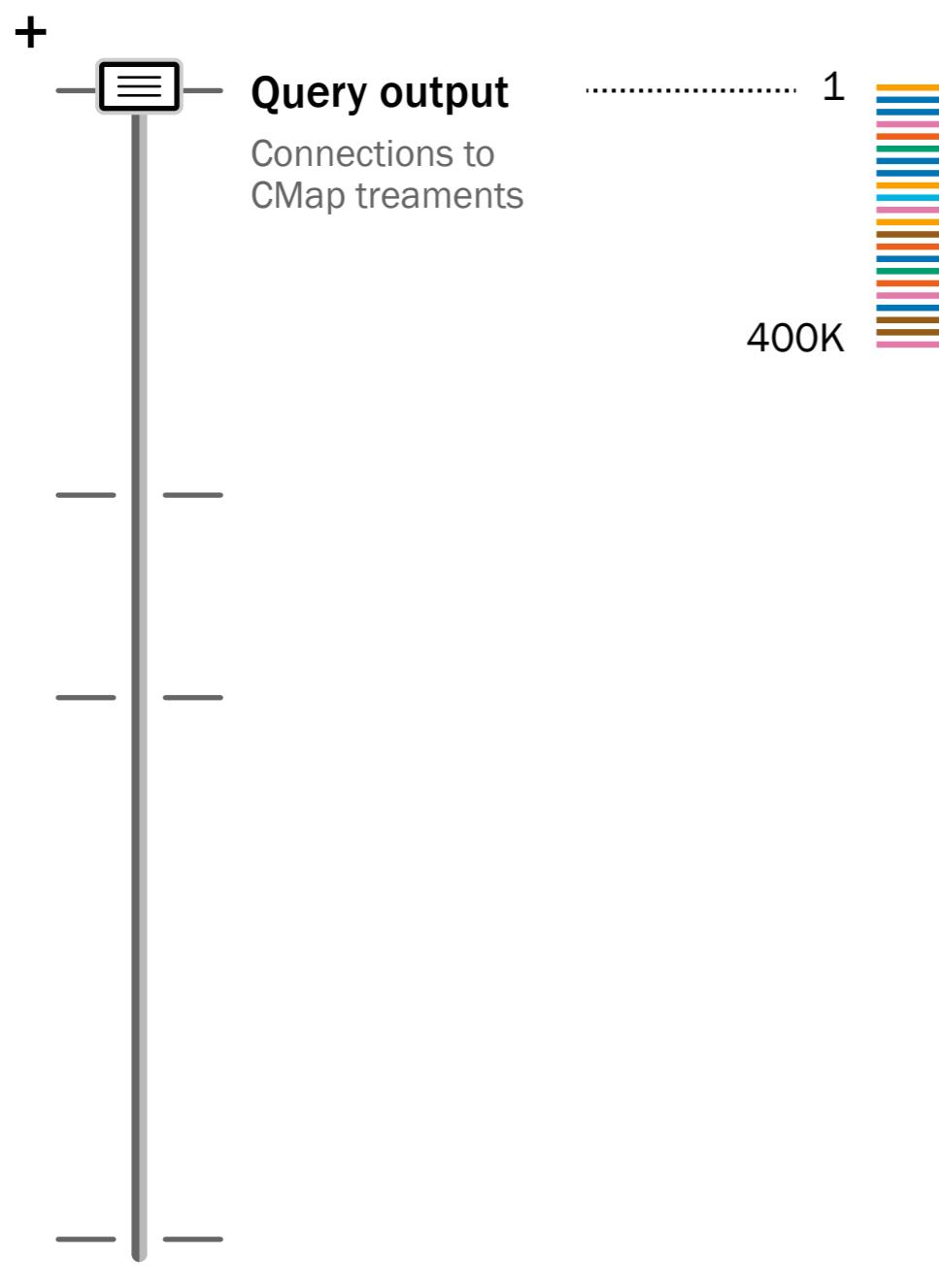
ANALYSIS METHODS

Design goals of CMap data and analytics

- Comprehensive
- Information-rich readouts (not optimized to particular questions)
- Easy to look things up (like Google)
- Easy to compare to non-CMap data
- Accessible to biologists and computationalists

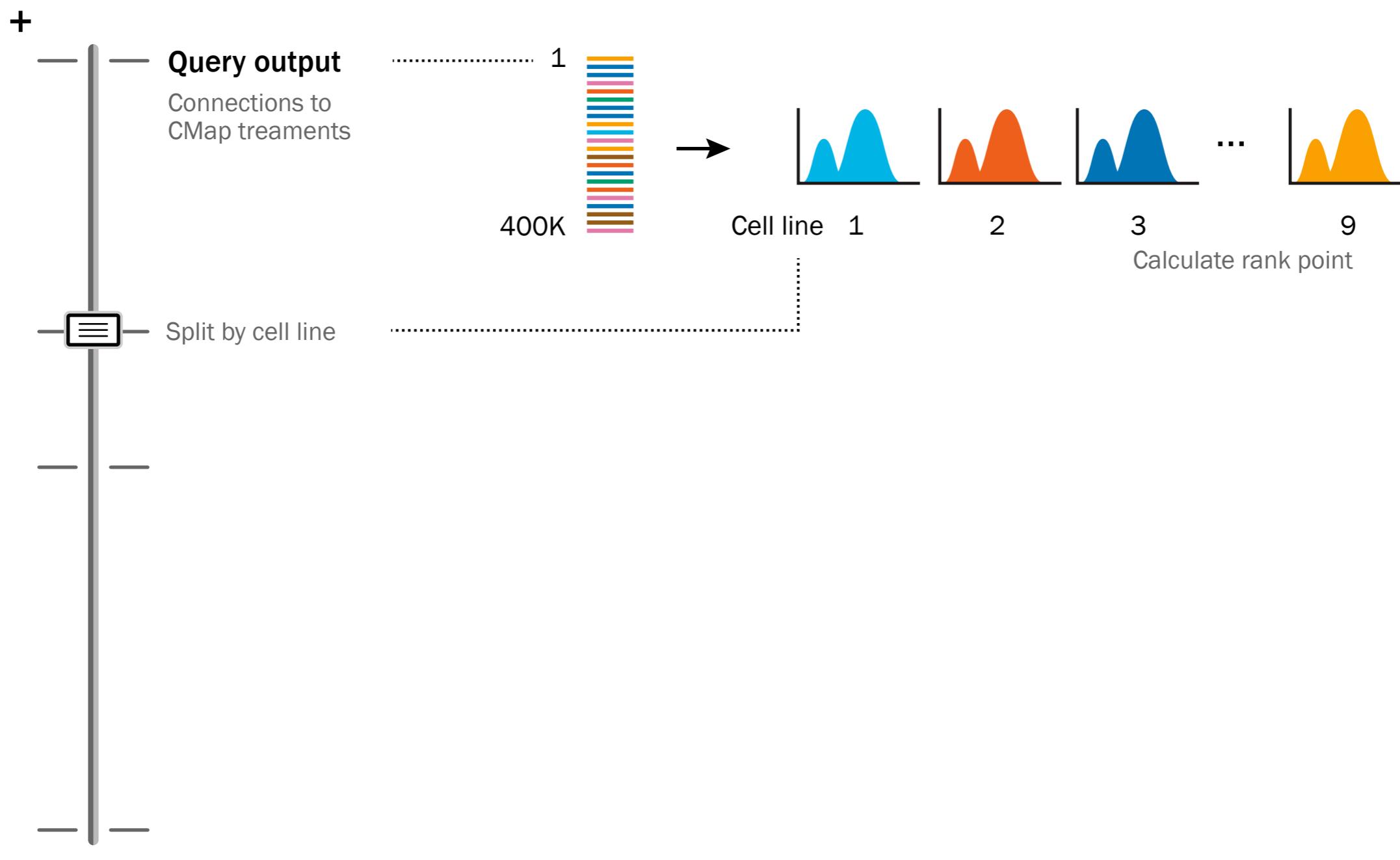
CMap result analysis

Many views of a single query



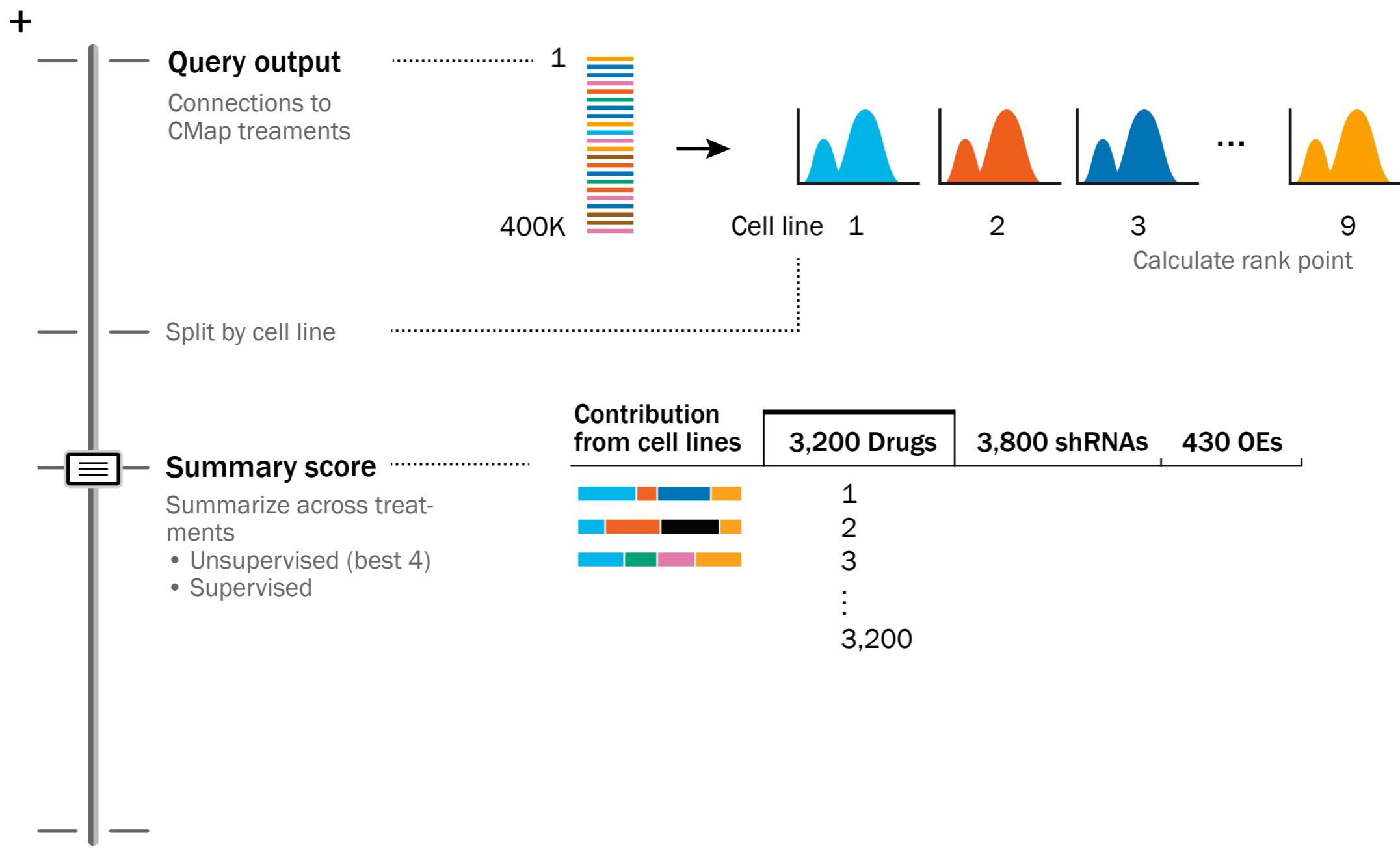
CMap result analysis

Many views of a single query



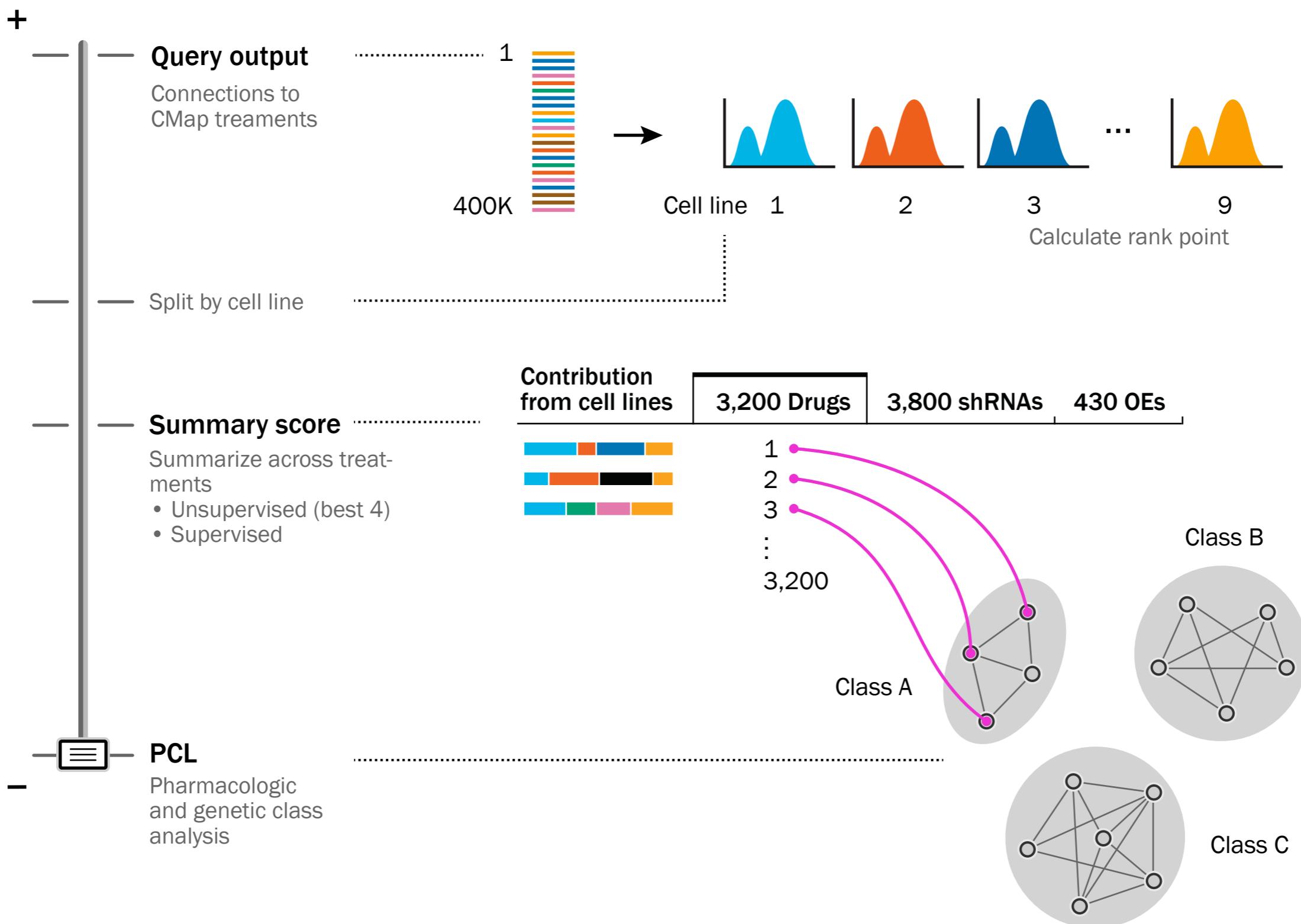
CMap result analysis

Many views of a single query



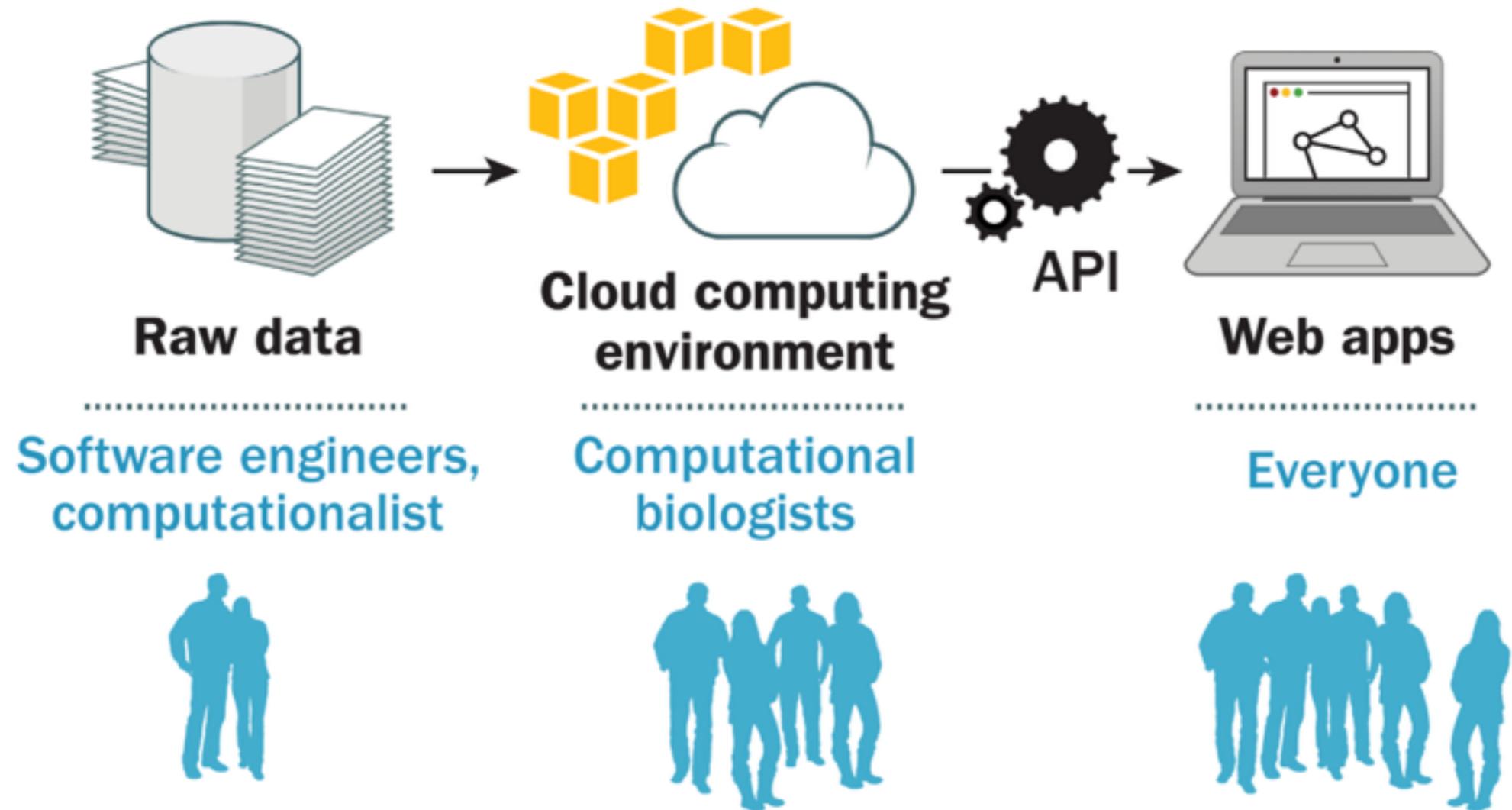
CMap result analysis

Many views of a single query

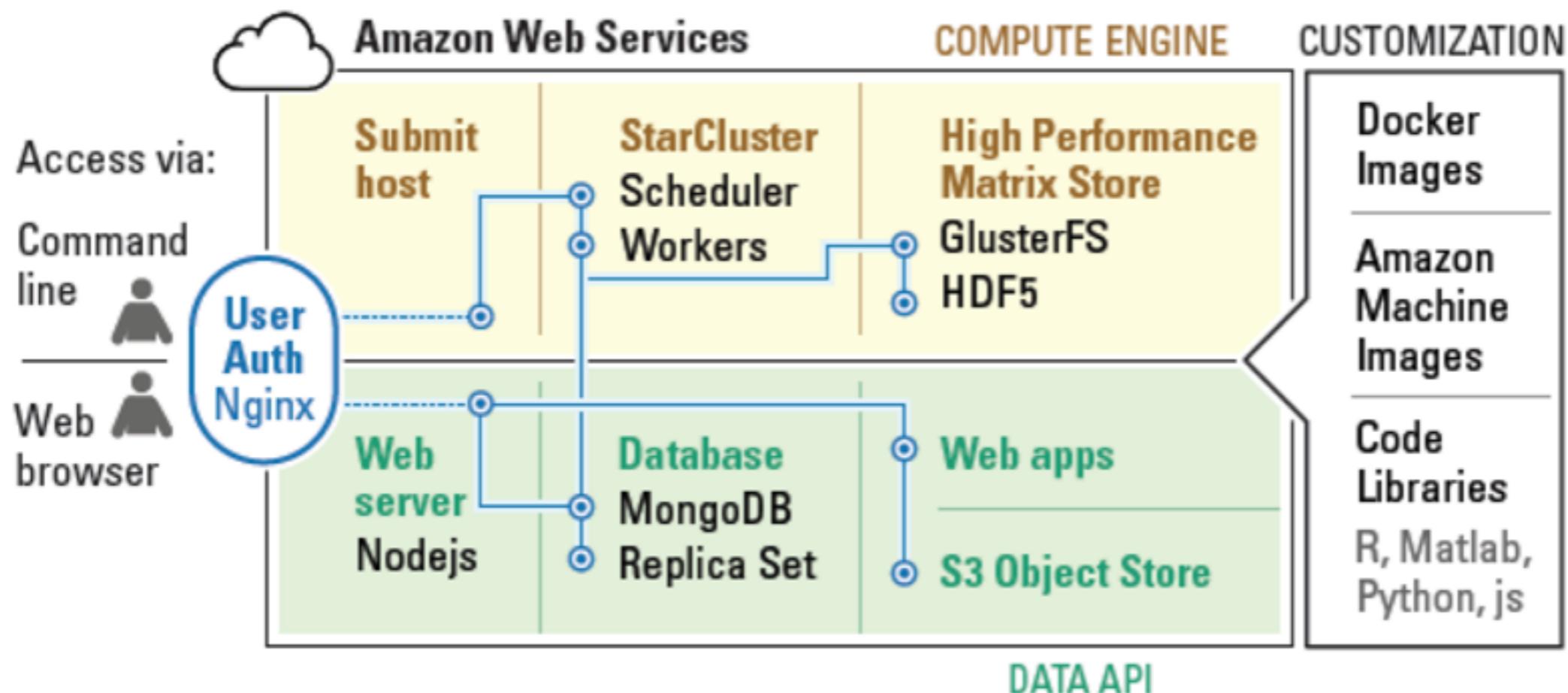


CLUE

Connectivity Map and LINCS User Environment



technology stack





Apps



Command Line



Command Line



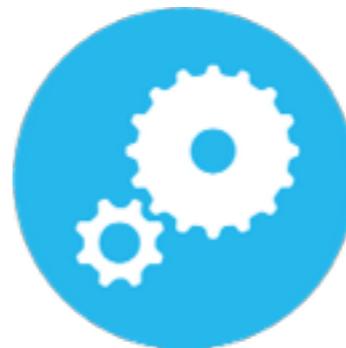
Command Line



Database



Command Line



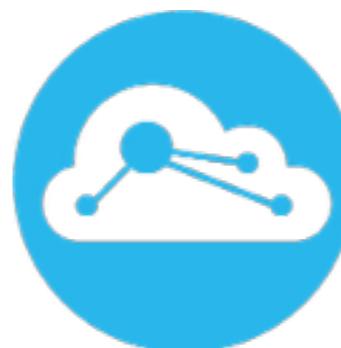
API



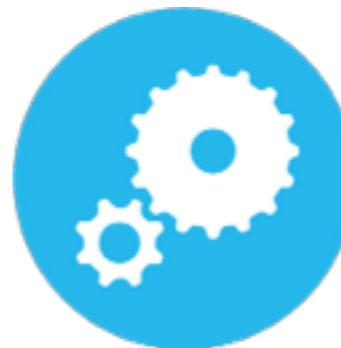
Database



Command Line



Virtual Machines



API



Database



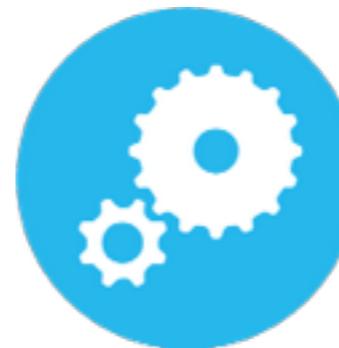
Command Line



Command Line



Virtual Machines



API



Database



Apps



Apps



Database



Apps



API



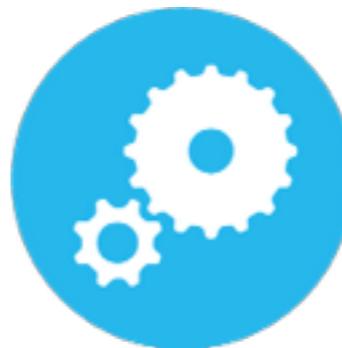
Database



Apps



Components



API



Database



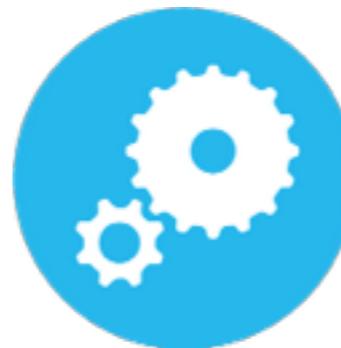
Apps



Apps



Components



API



Database

bookmarks window users Help

lincscloud.org www.lincscloud.org

cflynn ↗

lincscloud

☰⚙️📄❓

ACCESS THE DATA

- Cell Types Profiled
- Perturbagens Assayed
- Gene Expression Data
- Phosphoproteomics Data
- Imaging Data

Download the Data

About Team

About For Biologists For Developers

Our goal is to develop comprehensive signatures of cellular states and tools to analyze them in an effort to understand protein function, small-molecule action, physiological states, and disease characteristics.

LINCS

The Library of Integrated Cellular Signatures (LINCS) is an NIH program which funds the generation of perturbational profiles across multiple cell and perturbation types, as well as read-outs, at a massive scale.

LINCS CLOUD

This website, lincscloud.org, brings together datasets and tools from the LINCS consortium.

The Challenge

How can we make LINCS data accessible to researchers of all types so that it can help accelerate biomedical discovery?

NIH LINCS PROGRAM

Coupled with analytical tools, the vision is to, someday, make it possible for researchers to simply "look up" any cellular response in a genome-scale library of cellular signatures.

To date, LINCS has generated over 1 billion data points of perturbational profiles spanning small-molecules and genetic gain- and loss-of-function across multiple cell types.

bookmarks window users Help

lincscloud.org www.lincscloud.org

cflynn ↗

lincscloud

☰⚙️📝❓

ACCESS THE DATA

- Cell Types Profiled
- Perturbagens Assayed
- Gene Expression Data
- Phosphoproteomics Data
- Imaging Data

Download the Data

About Team

About For Biologists For Developers

Our goal is to develop comprehensive signatures of cellular states and tools to analyze them in an effort to understand protein function, small-molecule action, physiological states, and disease characteristics.

LINCS

The Library of Integrated Cellular Signatures (LINCS) is an NIH program which funds the generation of perturbational profiles across multiple cell and perturbation types, as well as read-outs, at a massive scale.

LINCS CLOUD

This website, lincscloud.org, brings together datasets and tools from the LINCS consortium.

The Challenge

How can we make LINCS data accessible to researchers of all types so that it can help accelerate biomedical discovery?

NIH LINCS PROGRAM

Coupled with analytical tools, the vision is to, someday, make it possible for researchers to simply "look up" any cellular response in a genome-scale library of cellular signatures.

To date, LINCS has generated over 1 billion data points of perturbational profiles spanning small-molecules and genetic gain- and loss-of-function across multiple cell types.



1. fish /Users/cflynn (fish)

wm398-ae4 [~]:



1. fish /Users/cflynn (fish)

wm398-ae4 [~]:

Query



apps.lincscloud.org/query



Query

cflynn ➔

Match user-defined gene sets to L1000 signatures

[take a tour](#)

Name your query

• Enter Up-regulated genes

Enter one gene symbol or
Affymetrix U133A probe ID per line
or drag and drop a plain text file
here.

• Enter Down-regulated genes

Enter one gene symbol or
Affymetrix U133A probe ID per line
or drag and drop a plain text file
here.

Query



apps.lincscloud.org/query



Query

cflynn ➔

Match user-defined gene sets to L1000 signatures

[take a tour](#)

Name your query

• Enter Up-regulated genes

Enter one gene symbol or
Affymetrix U133A probe ID per line
or drag and drop a plain text file
here.

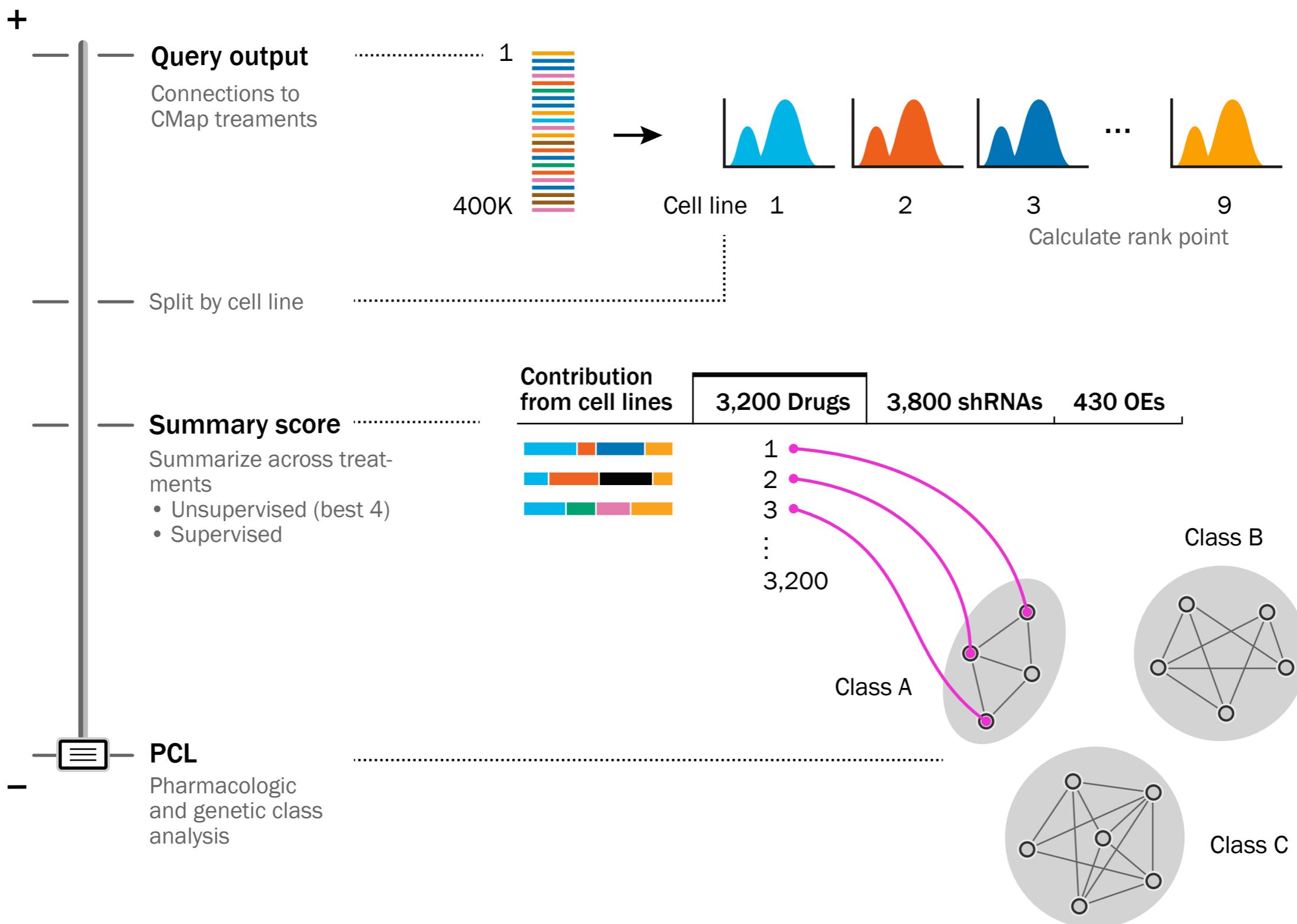
• Enter Down-regulated genes

Enter one gene symbol or
Affymetrix U133A probe ID per line
or drag and drop a plain text file
here.

Results

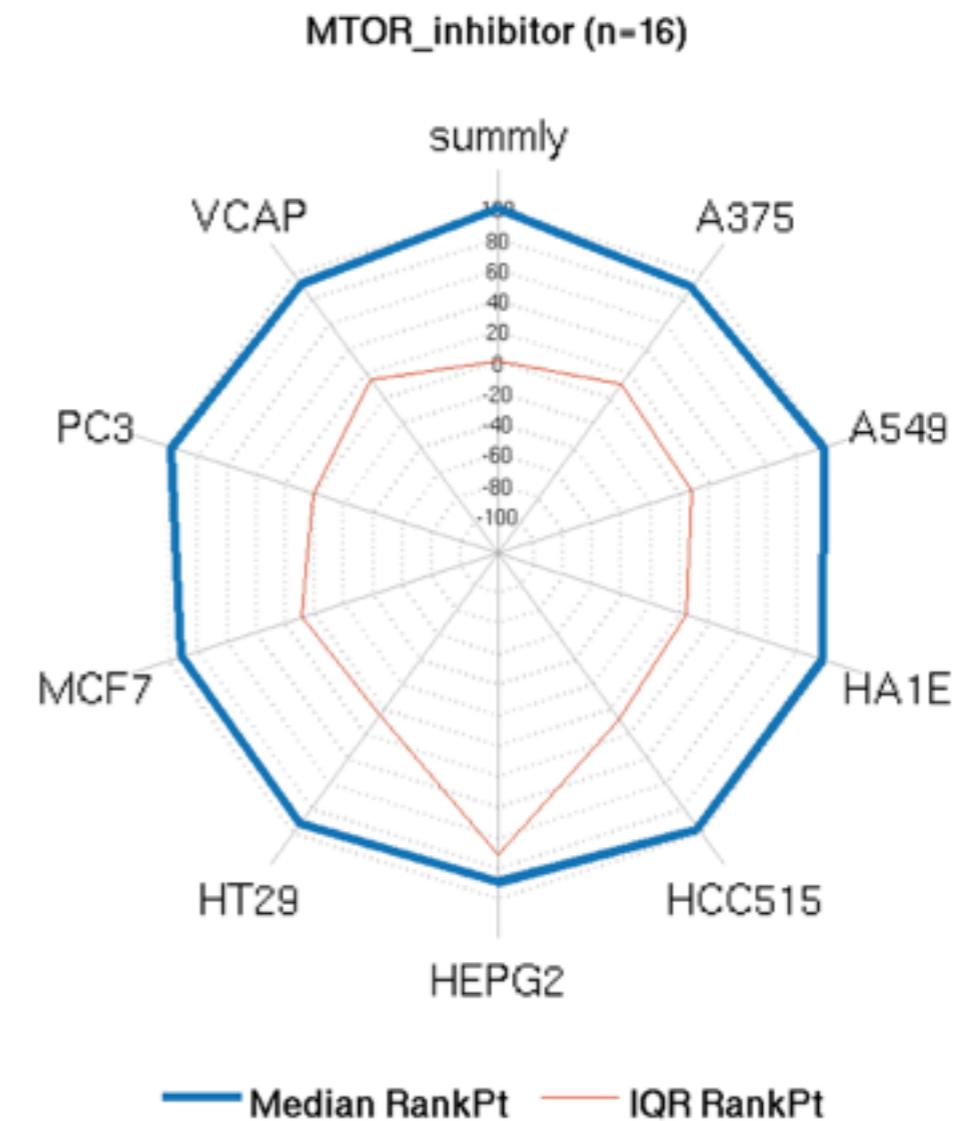
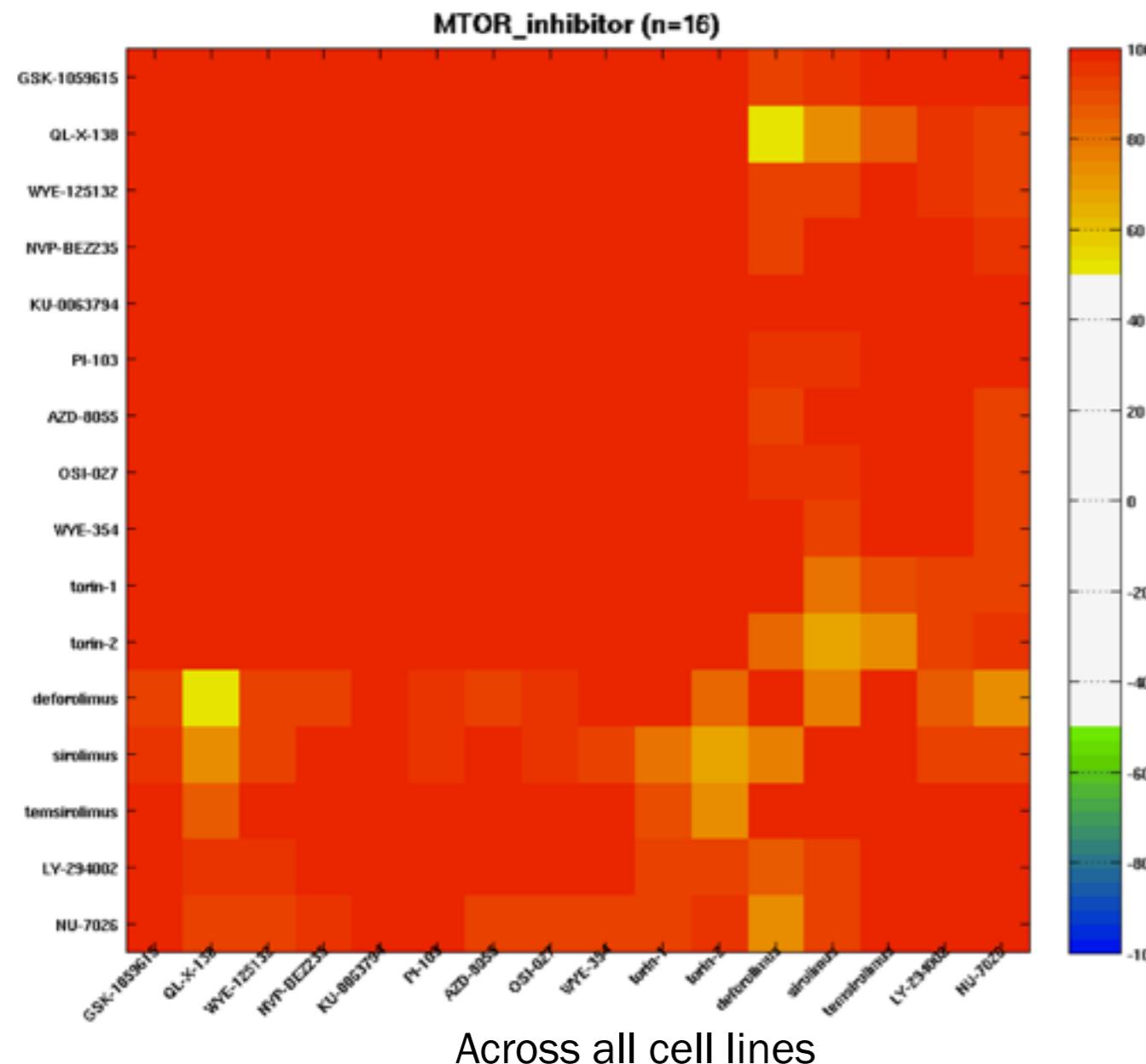
Long tail problem

Lots of connections: PCLs are one solution



PCL example

MTOR inhibitors



Successful mechanism prediction

Compounds not yet annotated in CMap

Compound	PCL	Confirmation
BRD-A66861218	Glucocorticoid agonist	Structural similarity
BRD-K72462751	ROCK inhibitor	Structural similarity
chlorambucil	Ribonucleotide reductase	Known mechanism
z-leu3-VS	Proteasome inhibitor	Known mechanism
sarmentogenin and digoxigenin	Sodium-potassium ATPase	Known mechanism
PF-562271	CDK inhibitor group 2	Known mechanism
HG-5-88-01	IGFR1R inhibitor	Kinomescan
WH-4-025	SRC inhibitor	Kinomescan
WH-4-025	KIT inhibitor	Kinomescan
JW-7-24-1	PI3K/MTOR inhibitor	Kinomescan

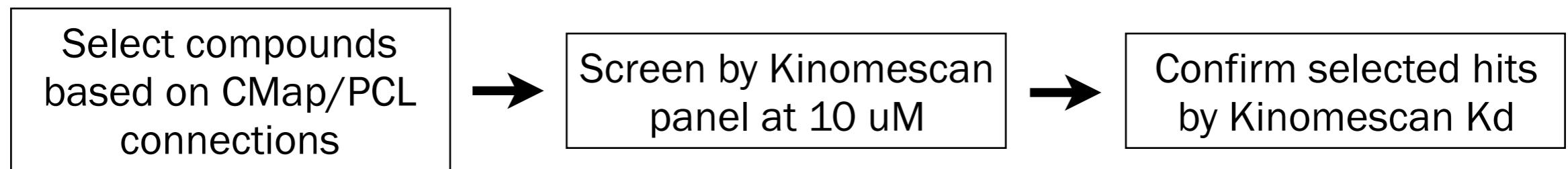
CMap Gene – PCL connections

Strong connections across multiple pathways

PCL	shRNA	OE
CDK inhibitor grp 1	CDK9	
EGFR inhibitor	GRB2	
GSK3 inhibitor	GSK3B	
HDAC inhibitor	EZH2	
KIT inhibitor	SRC	
MEK inhibitor	MAP2K1	RAF1 (anti-correlated)
MTOR inhibitor	MTOR, RPTOR	
P38 MAPK inhibitor	MAPK14/P38-alpha	
PARP inhibitor	PARG	
PKC activator	PRKCE (anti-correlated)	PRKCE
Proteasome inhibitor	PSM genes	
Protein synthesis inhibitor	RPL/RPS genes	
Ribonucleotide reductase inhibitors	RRM1, POLE2	
SRC inhibitor	GRB2	SRC (anti-correlated)
Tubulin inhibitor	ACTB	
TGF-beta inhibitor	SMAD4	

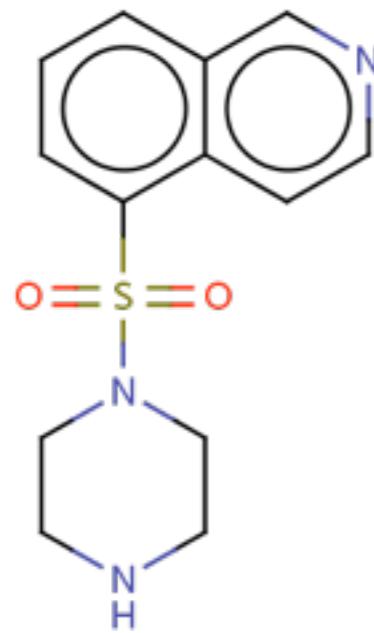
Predicted novel kinase inhibitors

Experimental overview



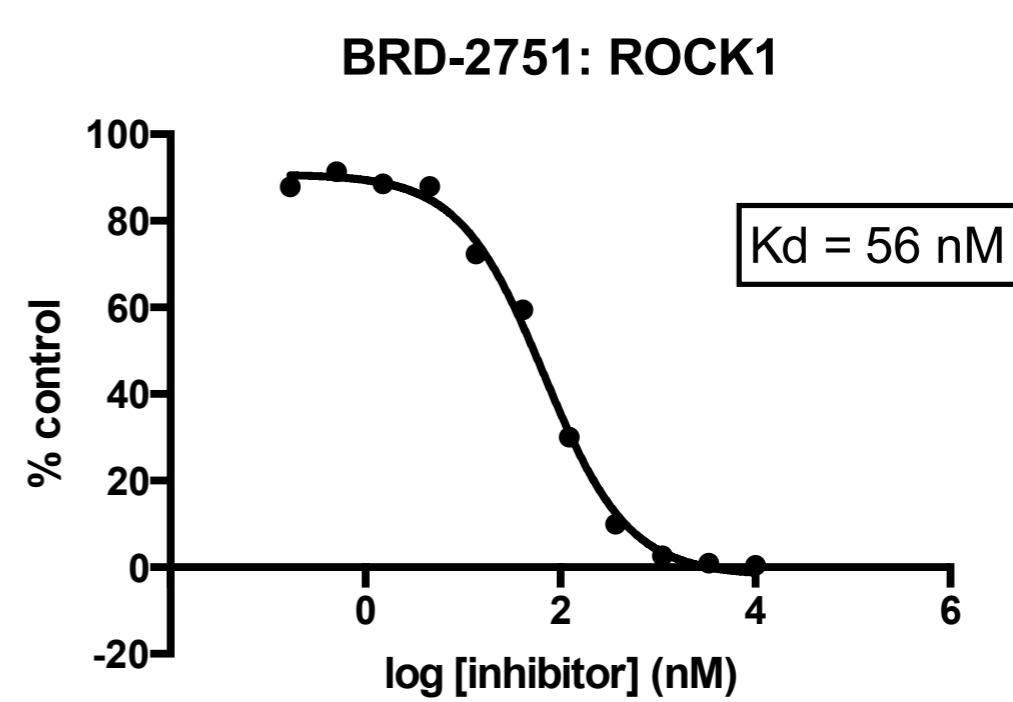
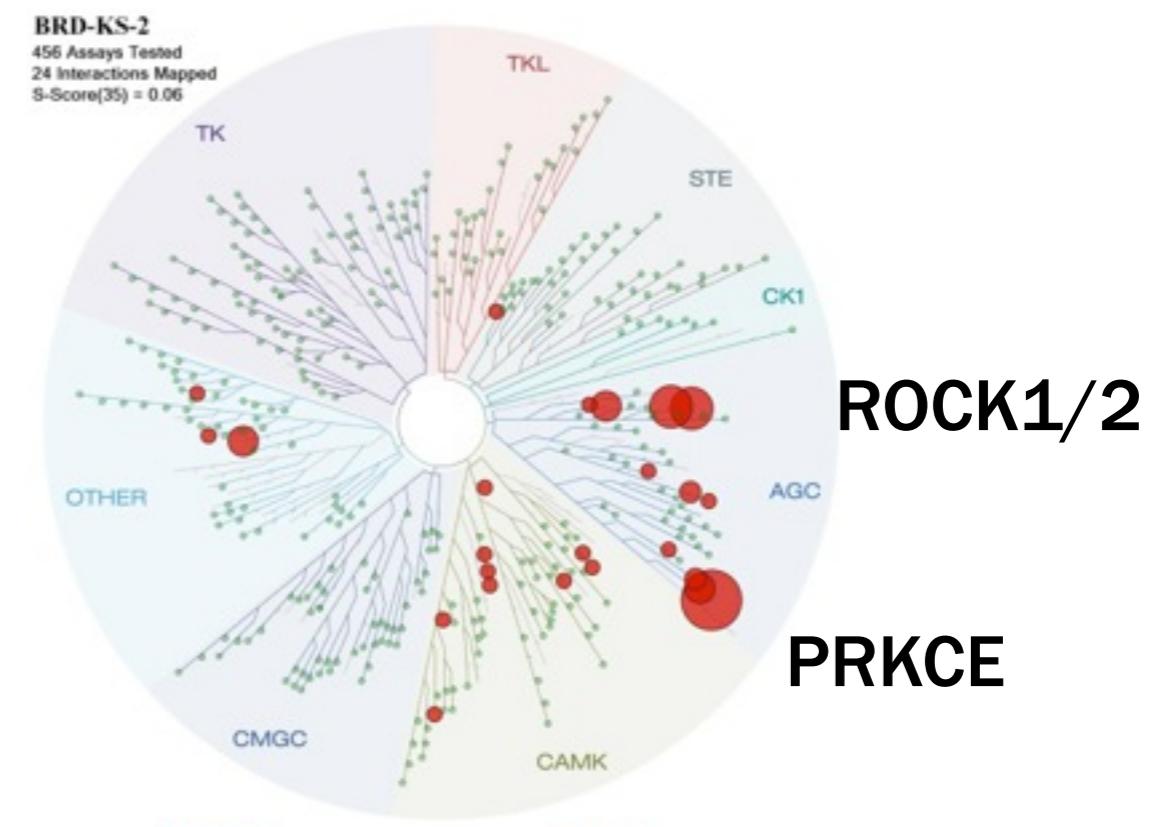
BRD-2751

Confirmed ROCK1/2 inhibitor



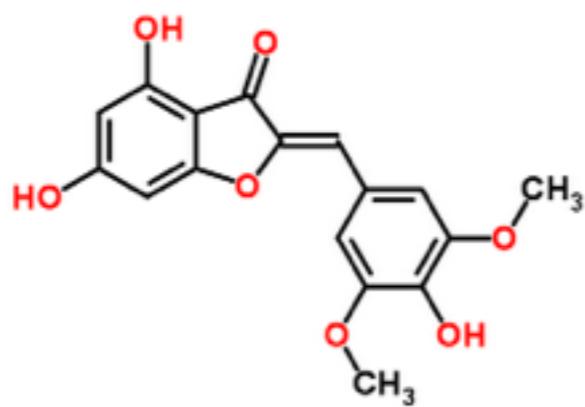
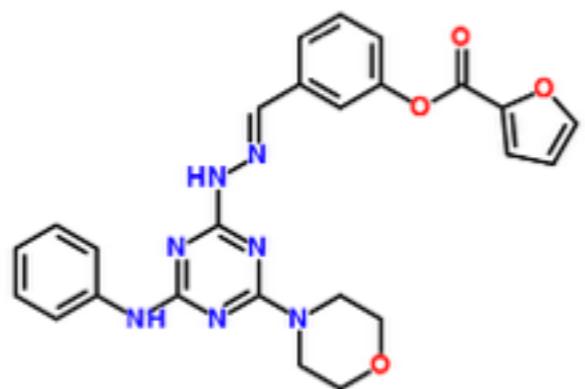
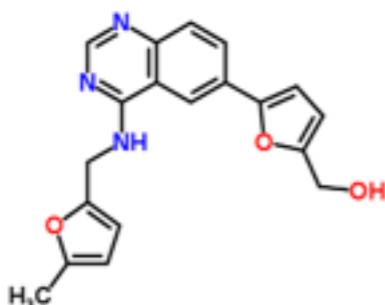
Class	CP	CP	JAK	INHIBITOR
Type	CP	CP	PKA	INHIBITOR
	CP	CP	GSK3	INHIBITOR
	CP	CP	CDK	INHIBITOR
	CP	CP	ROCK	INHIBITOR

Legend: CP (white), JAK (light red), PKA (medium red), GSK3 (dark red), CDK (orange), ROCK (yellow).



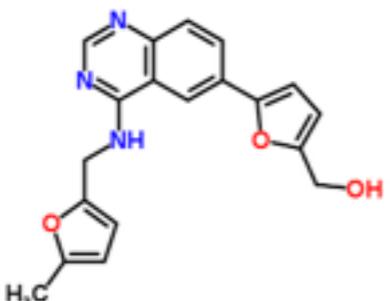
Predicting Drug Function

Diverse structures, common activities

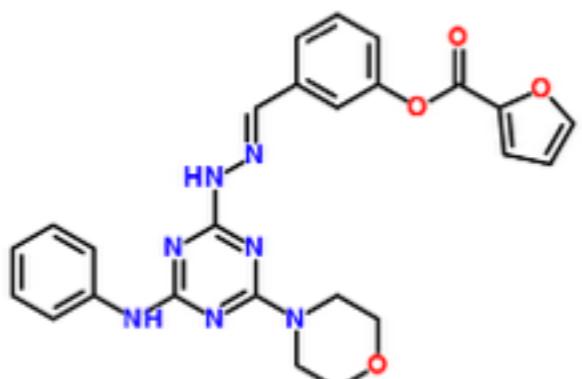


Predicting Drug Function

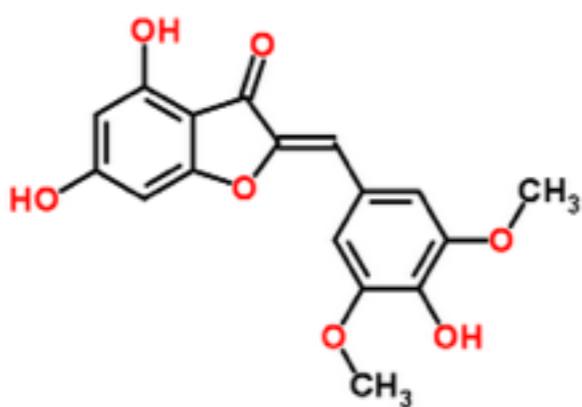
Diverse structures, common activities



✗ VEGFR inhibitor



✗ PPARG agonist



✓ PI3K/MTOR inhibitor

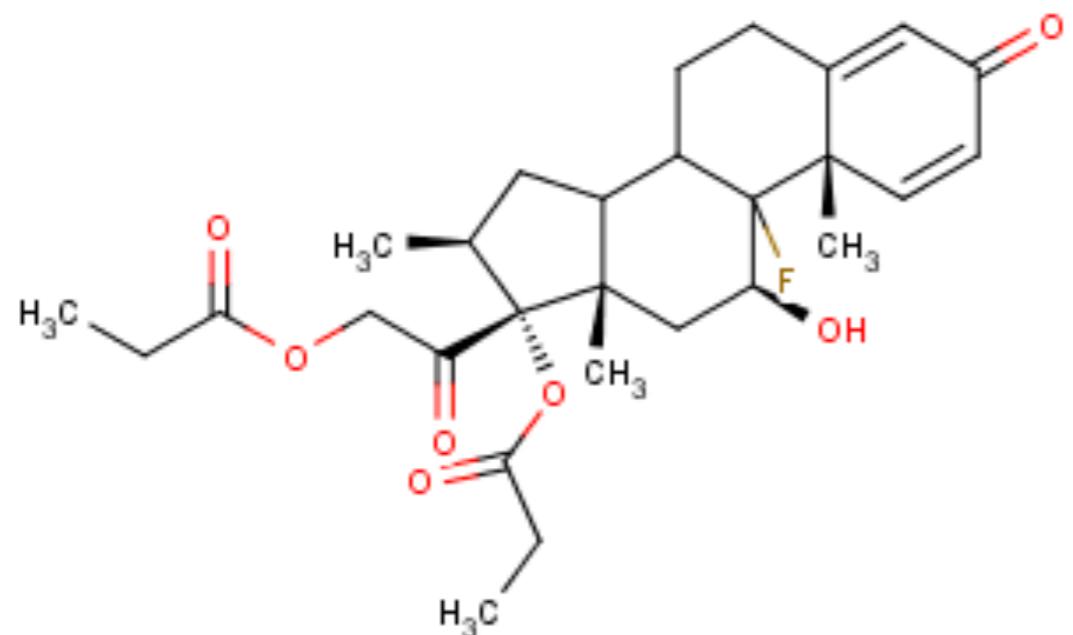
✗ ROCK inhibitor

✗ Estrogen agonist

Finding Novel Drug Targets

Repurposing failed drugs

✓ Original target

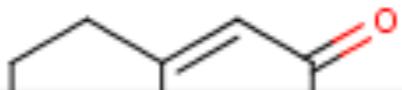


Finding Novel Drug Targets

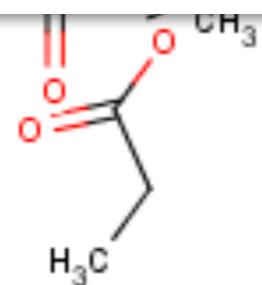
Repurposing failed drugs



Original target

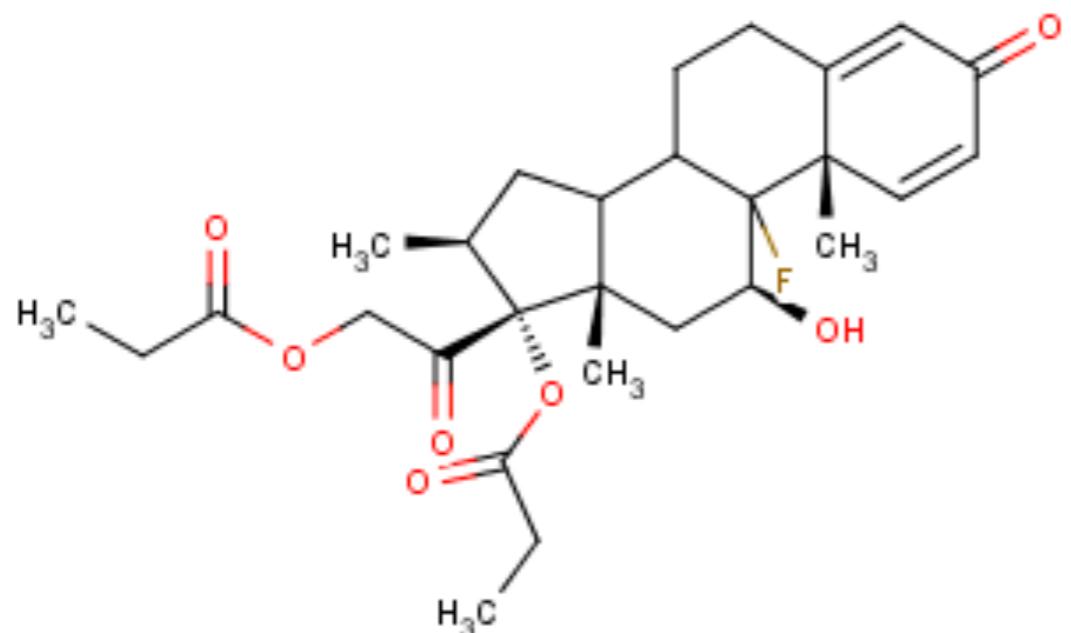


Failed in Phase 2 clinical trial due to lack of efficacy



Finding Novel Drug Targets

Repurposing failed drugs



✓ Original target

✗ Novel Target A

✓ Novel Target B

✓ Novel Target C

✗ Novel Target D

