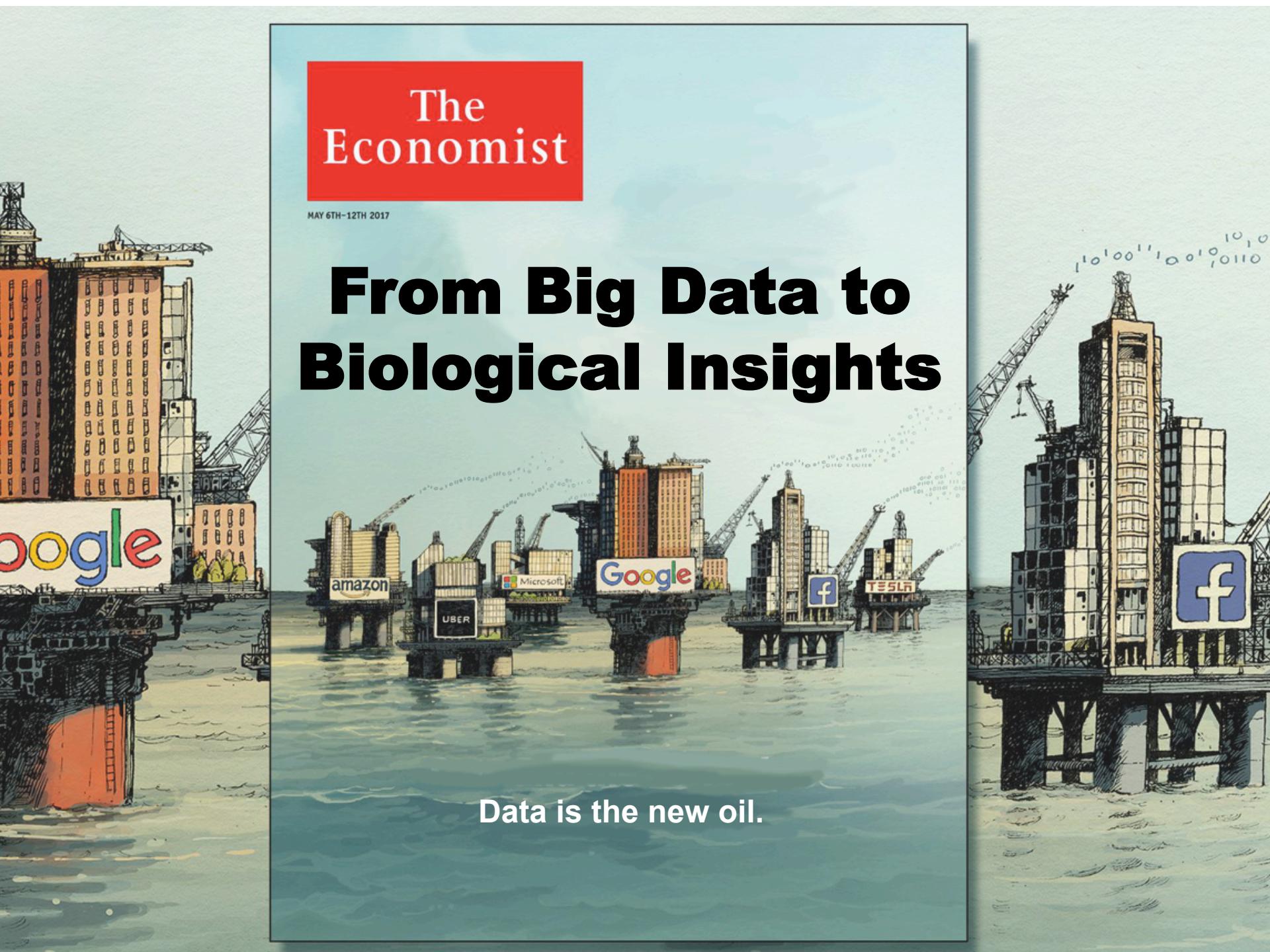


The
Economist

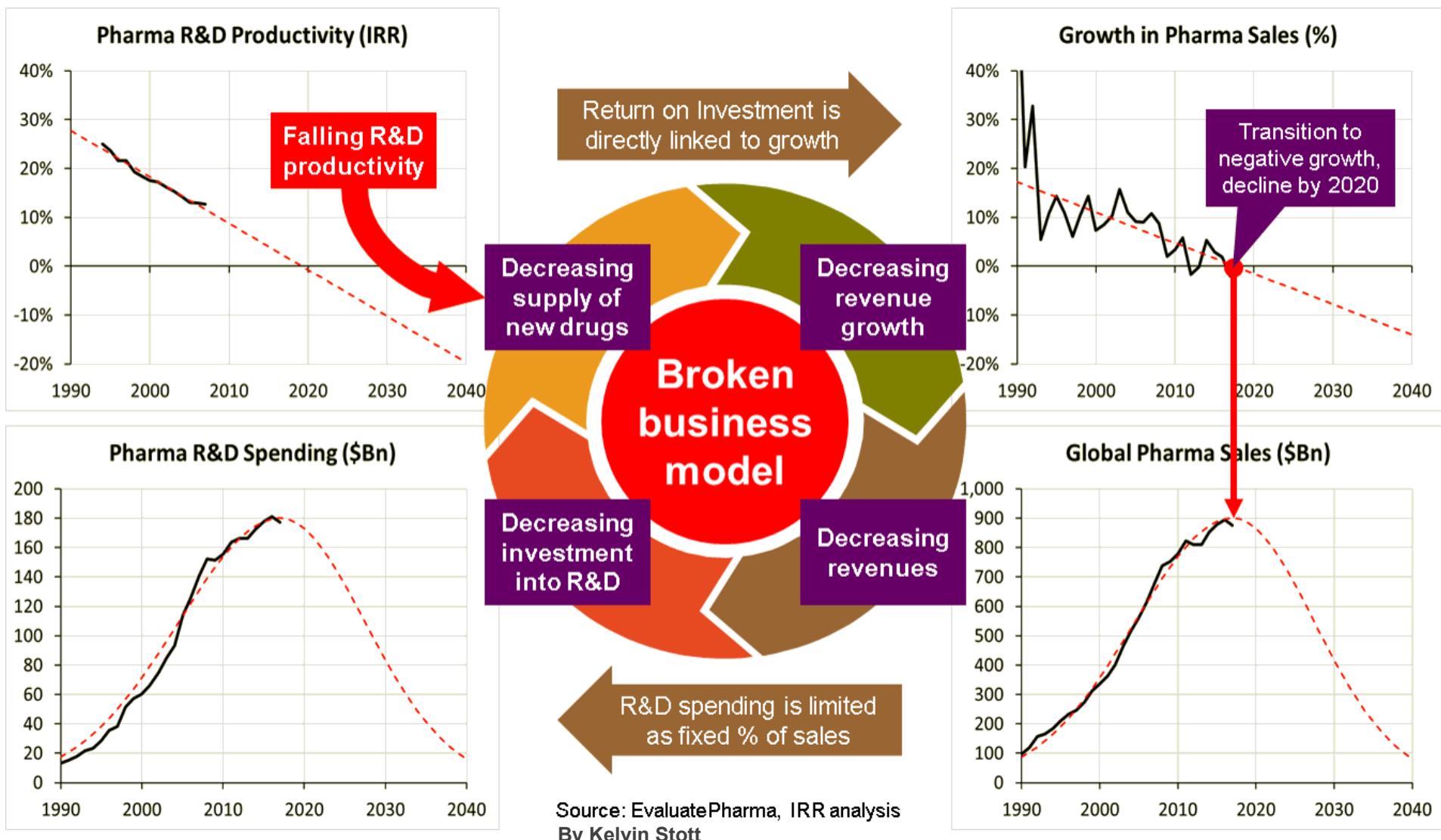
MAY 6TH-12TH 2017

From Big Data to Biological Insights

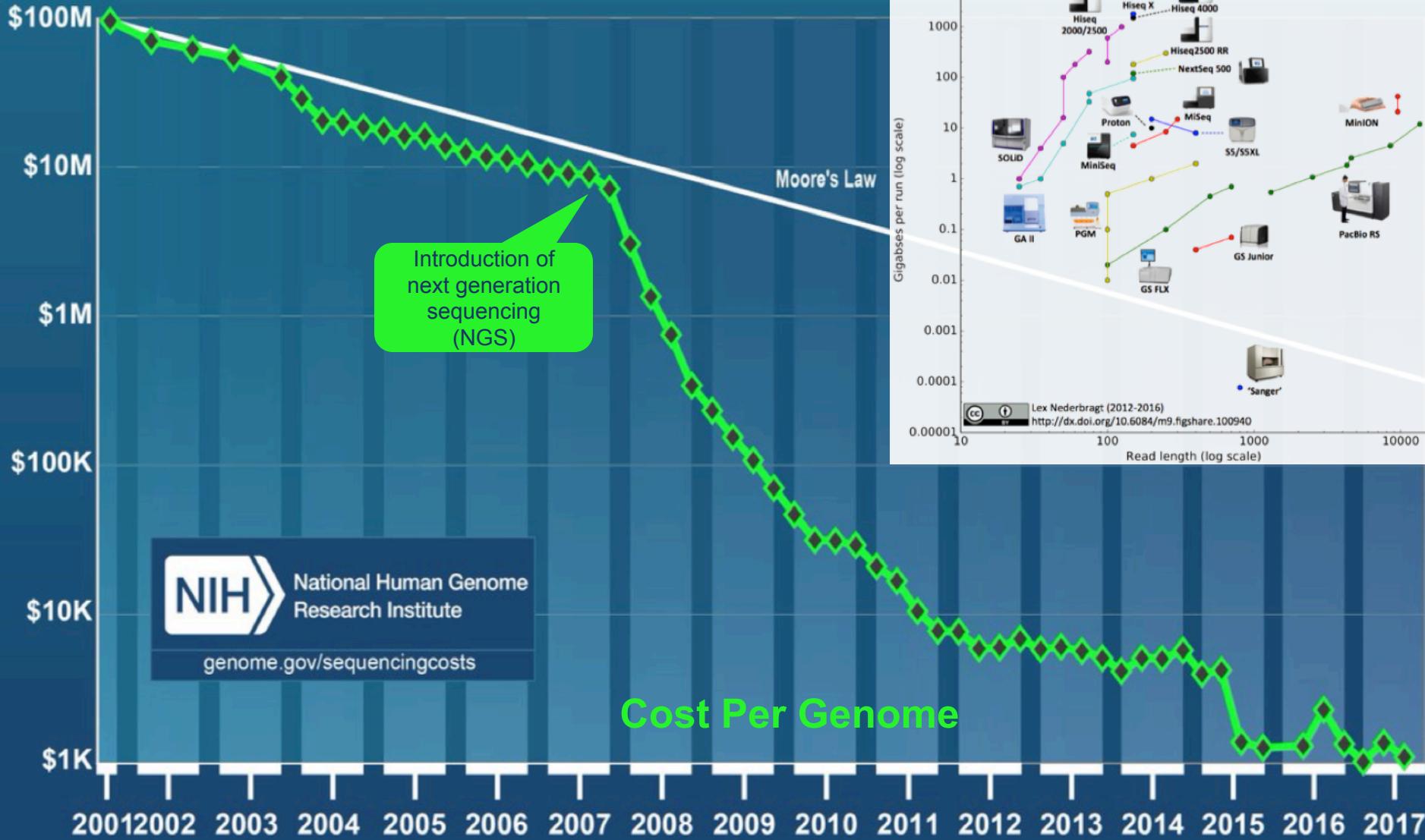


Data is the new oil.

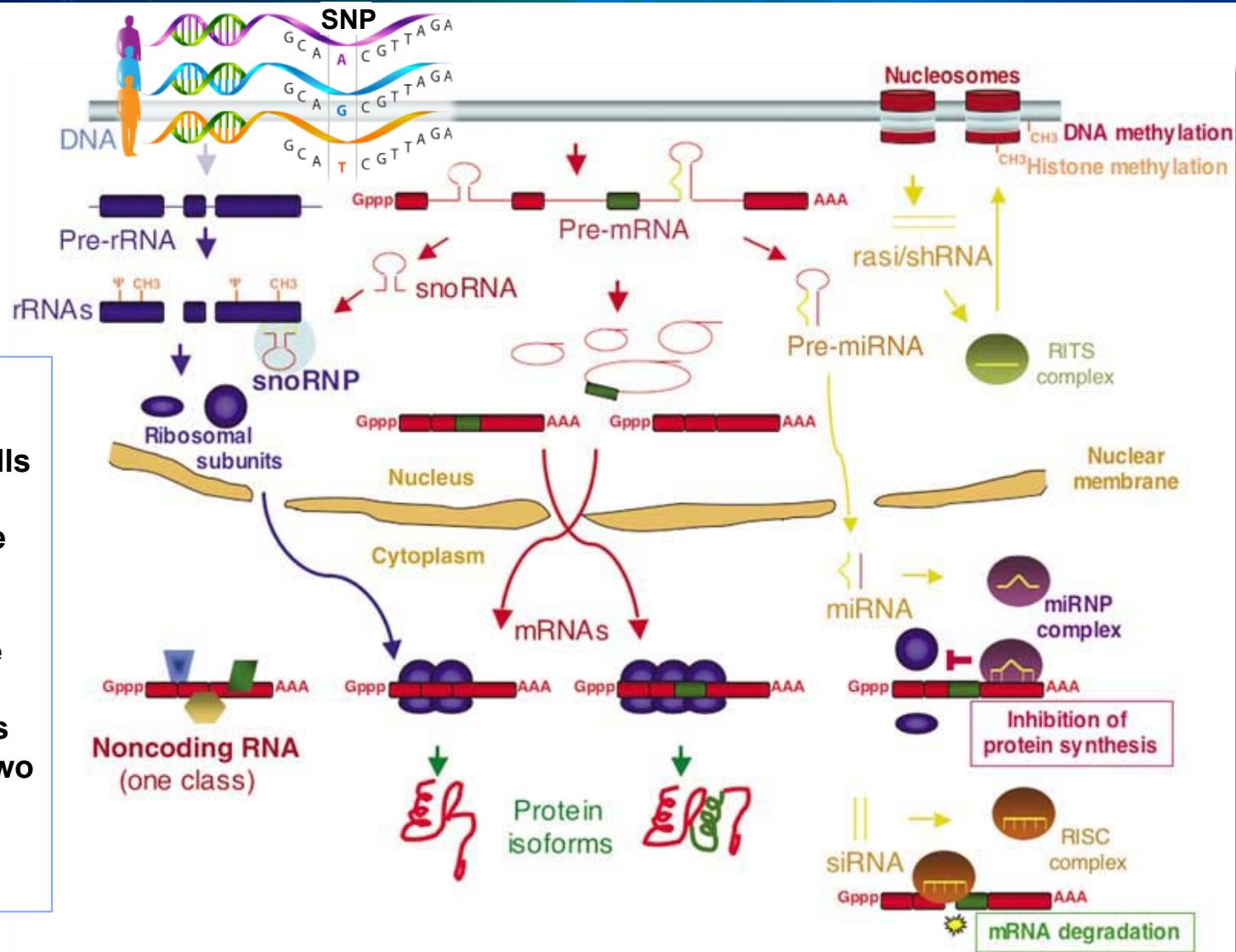
Pharma's Current Broken Business Model



Genomics Revolution: More Data Coming Cheaply



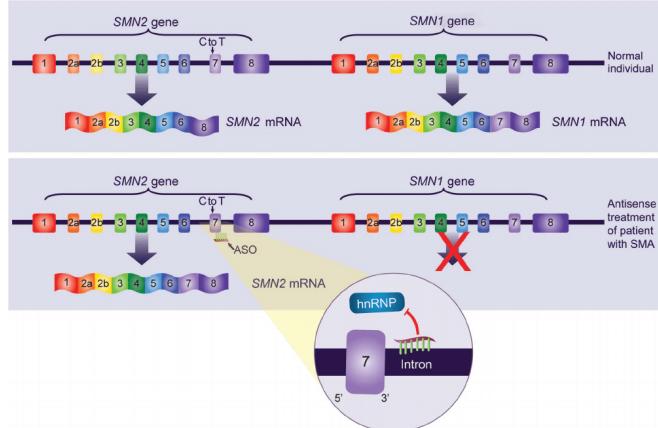
DNA, Genes, Transcripts, Exons and Alternative Splicing



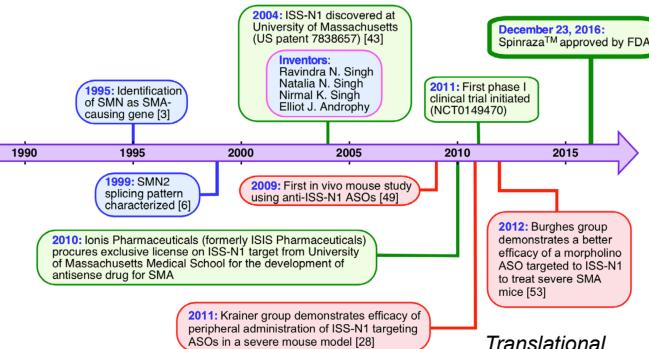
- A gene can encode multiple transcripts that consists of different sets of exons.
- Isoforms are biological entities, and they are **tissue and disease specific**.

From Fundamental Science to Life Saving Medicine

Mechanism of Action

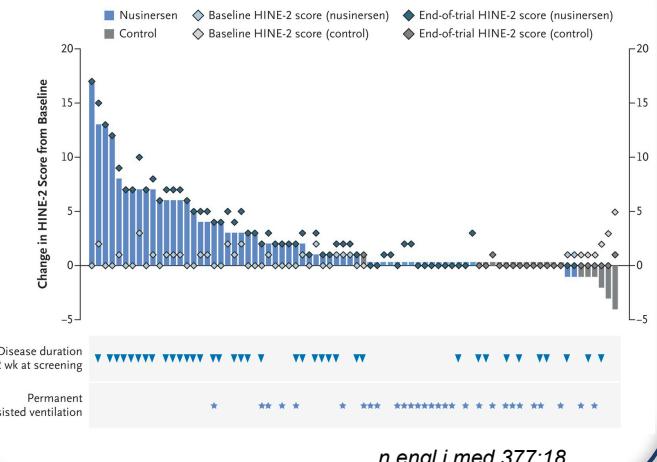


Speedy Development Timeline



Translational
Neuroscience
8 • 2017 • 1-6

Superior Clinical Benefits

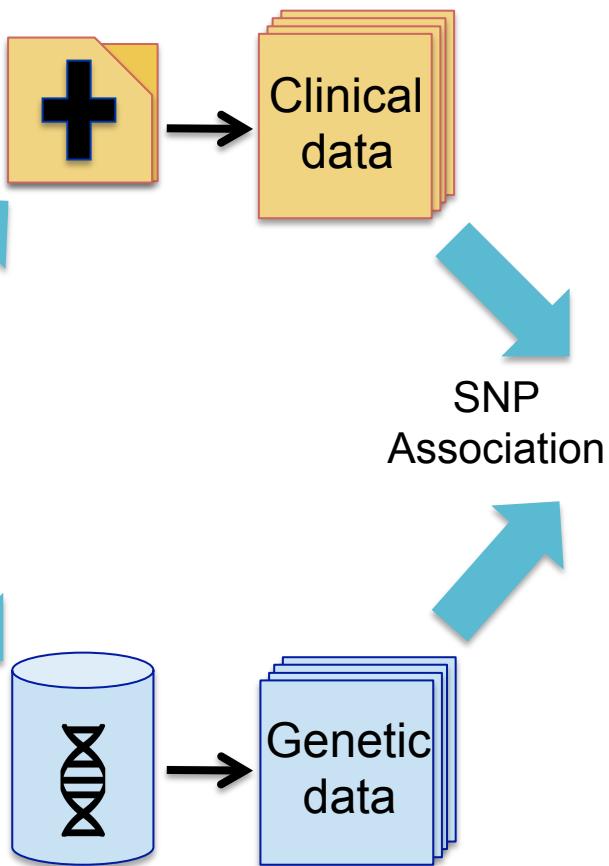


Phenome-wide Association Studies (PheWAS)

Disease-agnostic cohort

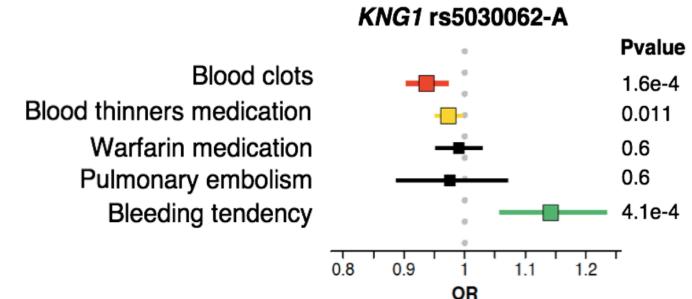
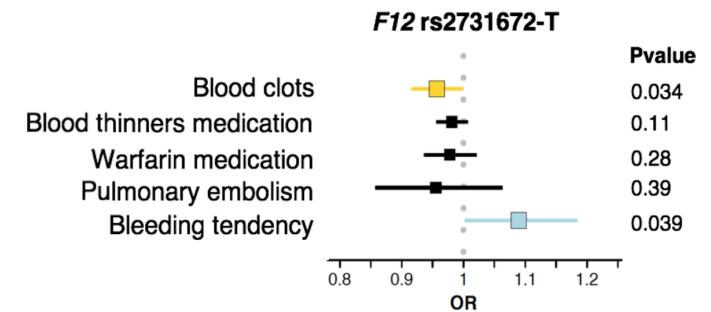
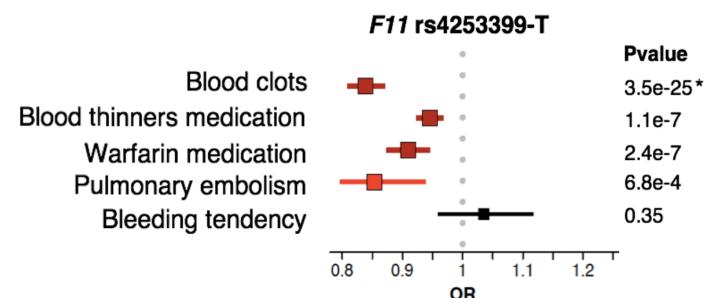


EHRs, Claims,
Questionnaires, etc.



GWAS, exome, whole genome sequencing, etc.

Target Prioritization in Coagulation Pathway



Human Genetics Based Drug Development Platform

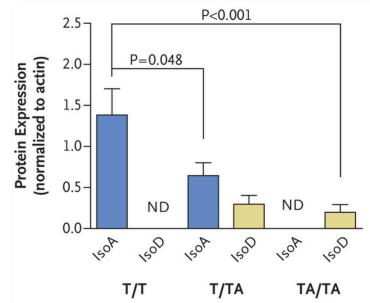
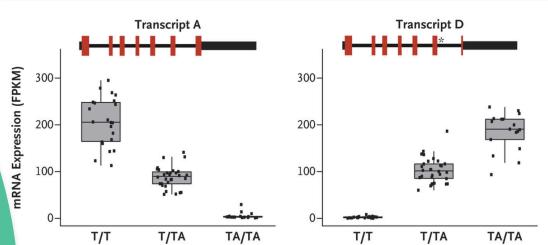
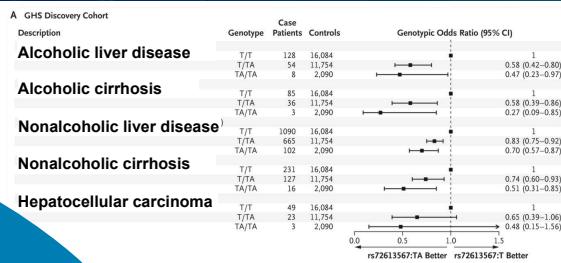
Exome,
EHR
(n=46,544)

RNAseq
Isoform
(n=69)

Western
Blot
(n=69)

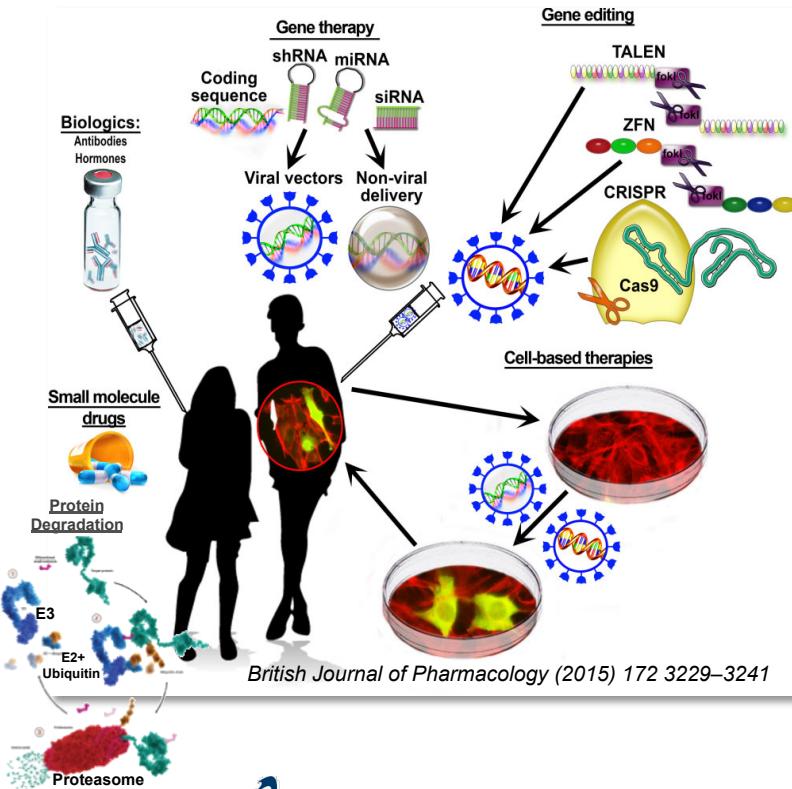


N Engl J Med 2018; 378:1096-1106



Fast
Translation

A loss-of-function variant in HSD17B13 is associated with a reduced risk of chronic liver diseases and of progression from steatosis to steatohepatitis

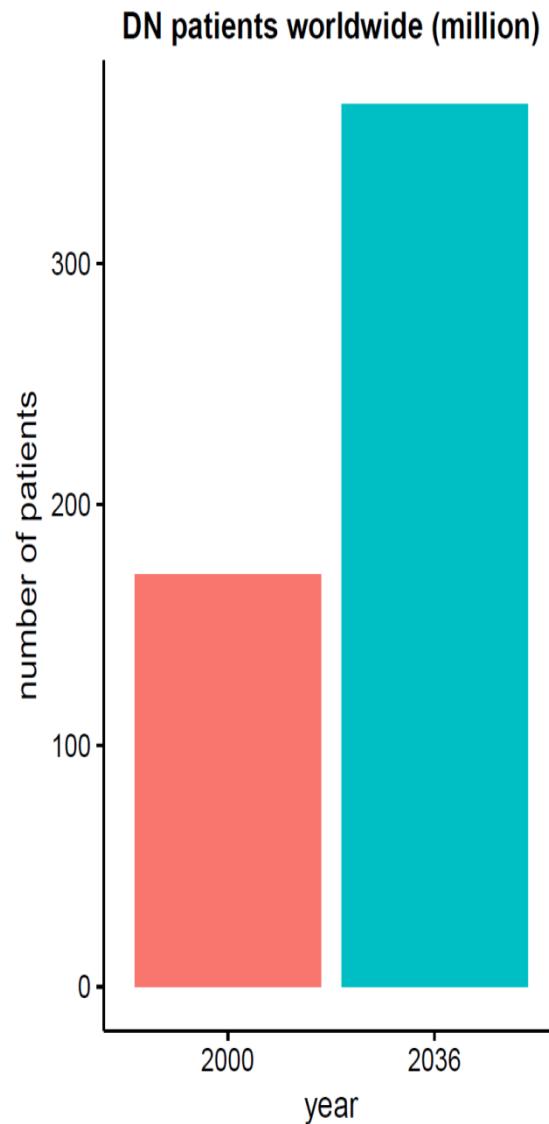


Alnylam[®]
PHARMACEUTICALS

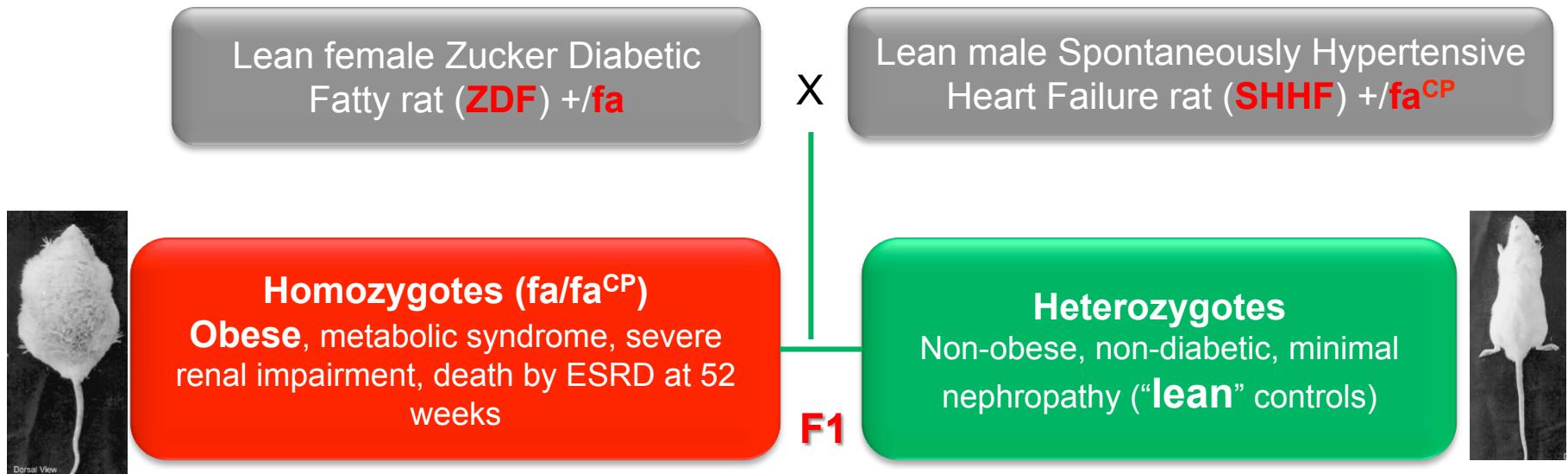
Find the right modality from modulation tool box, ranging from small molecules, biologics, gene therapy, protein degradation to cell therapy

Diabetic Nephropathy: Unmet Medical Need

- **Diabetic nephropathy (DN)**
 1. Leading cause of morbidity and mortality in patients with diabetes
 2. Accounts for **30-40%** of all **end stage renal disease (ESRD)** in the US
- Medicare expenditures for ESRD were an estimated **\$31 billion** in 2013, accounting for over **7%** of Medicare-paid claims costs

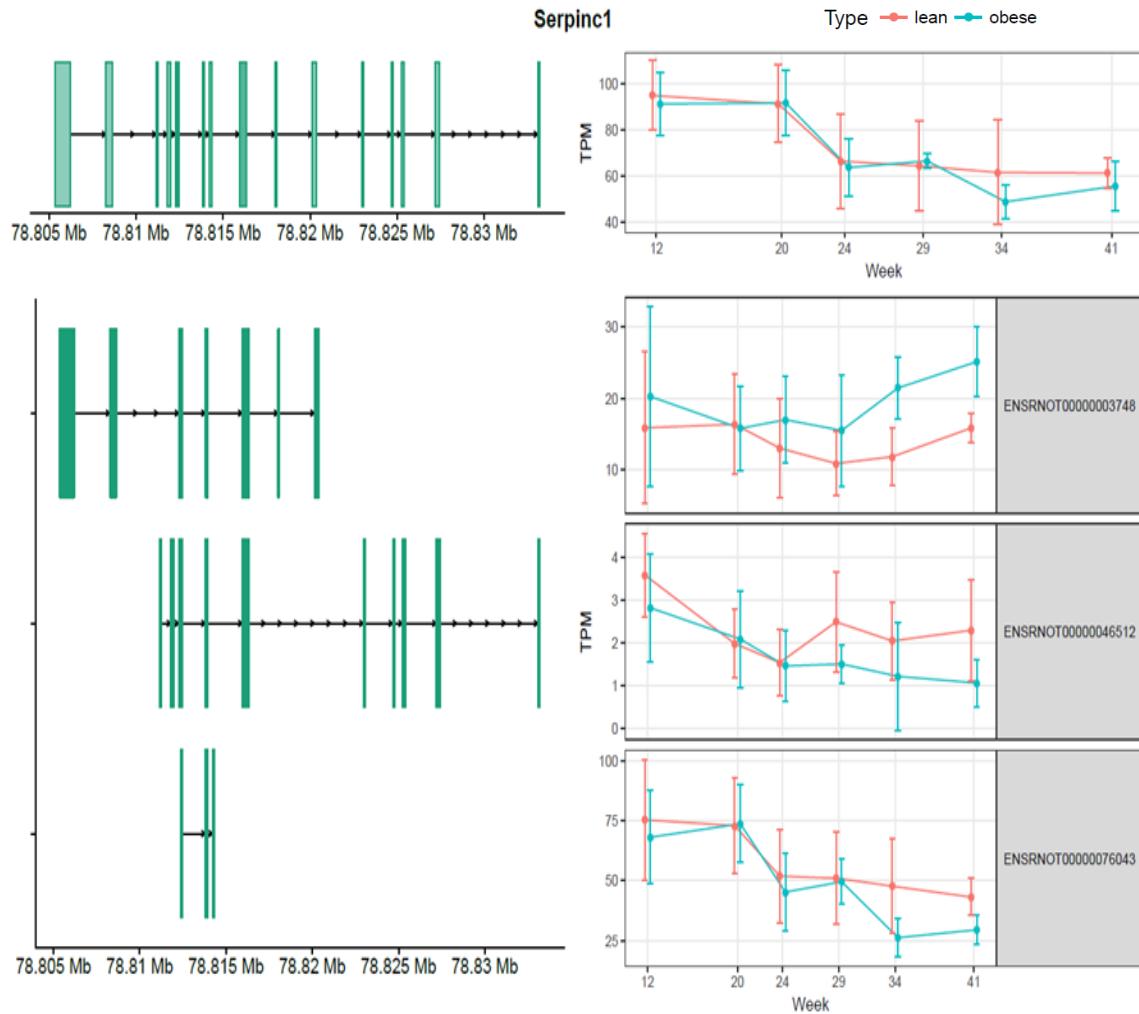


ZSF1 Rats: a Preclinical Model for Type 2 DN



- Lean female ZDF rat (+/fa) and lean male SHHF rat (+/fa^{CP}): both strains have mutated leptin receptors, but differently.
- F1 cross of rat strains ZDF and SHHF and generate rats of two phenotypes
 - **Obese** (Homozygotes): fa/fa^{CP}
 - **Lean** (Heterozygotes): +/fa; +/fa^{CP}, +/+

Expressions of Serpinc1 (serpin family C member 1) and its Isoforms

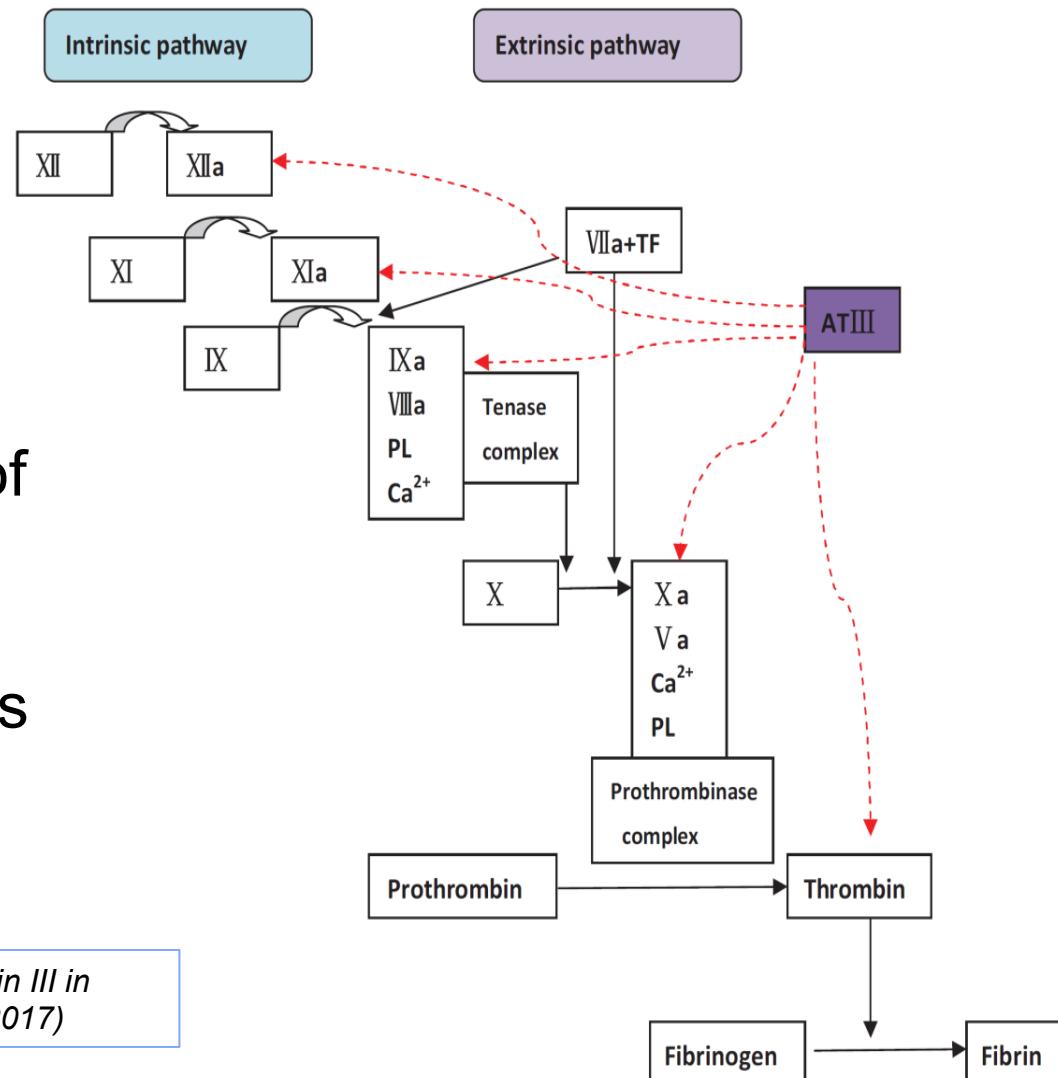


At weeks 34 and 41,
obese vs lean animals:

- **ENSRNOT00000003748**
48: higher in obese animals
- **ENSRNOT00000046512**/
ENSRNOT00000076043: higher in lean animals

Serpinc1 Has Impact on Kidney Inflammation and Fibrosis

- **Antithrombin III**, encoded by the gene *SerpinC1*, and **ATIII** deficiency is associated with **kidney injury**
- Isoform usage change of Serpinc1 can have an impact on kidney inflammation and fibrosis

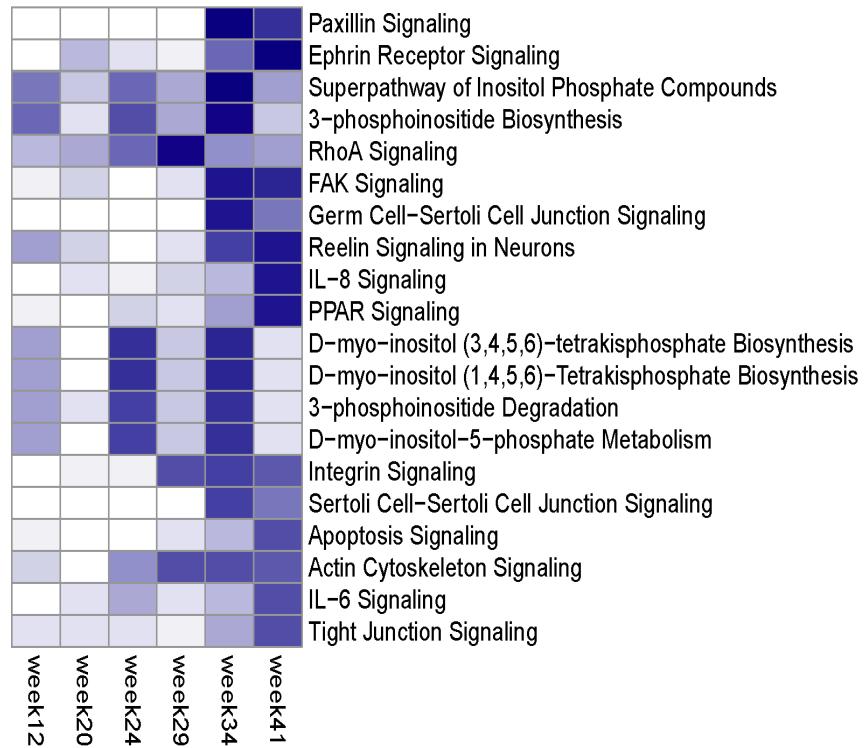


Lu, Z., Wang, F. & Liang, M. SerpinC1/Antithrombin III in kidney-related diseases. *Clin Sci*, 131, 823-831 (2017)

Differential Splicing Versus Expression: Different Pathways Enriched

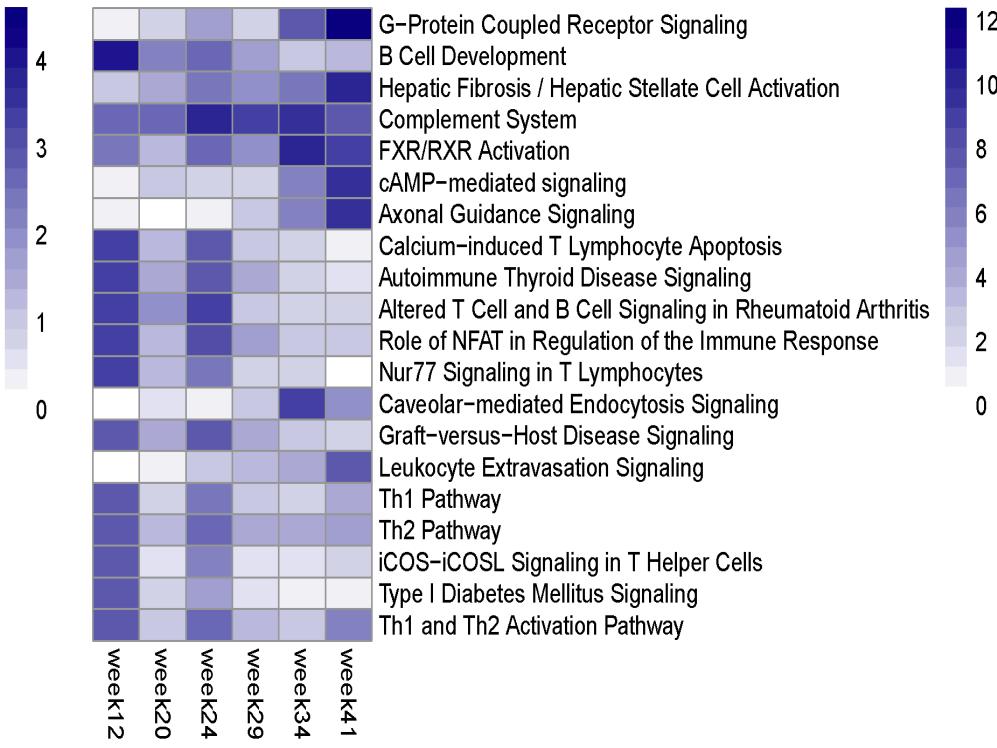
Differential splicing (DS)

A



Differential Expression (DE)

B



Top 20 canonical pathways were selected according to the smallest P value reported by IPA across all time points, and colored by their corresponding $-\log_{10}(P \text{ value})$.

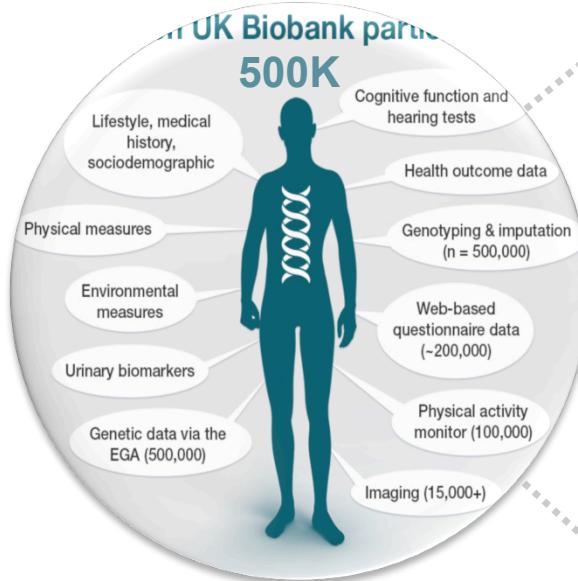
Understanding Diseases from Genetics in Big Scale

RGC to sequence exomes from 500,000 people by end of 2019; data will be paired with detailed, de-identified health information

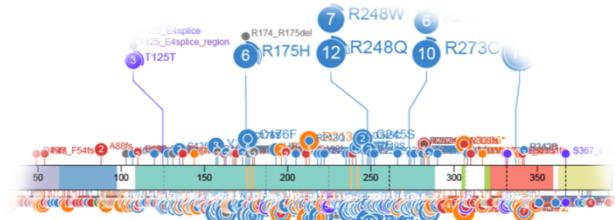
All data will be openly available to the global research community

Largest database of its kind may have profound impact on human health

Additional collaborators expected

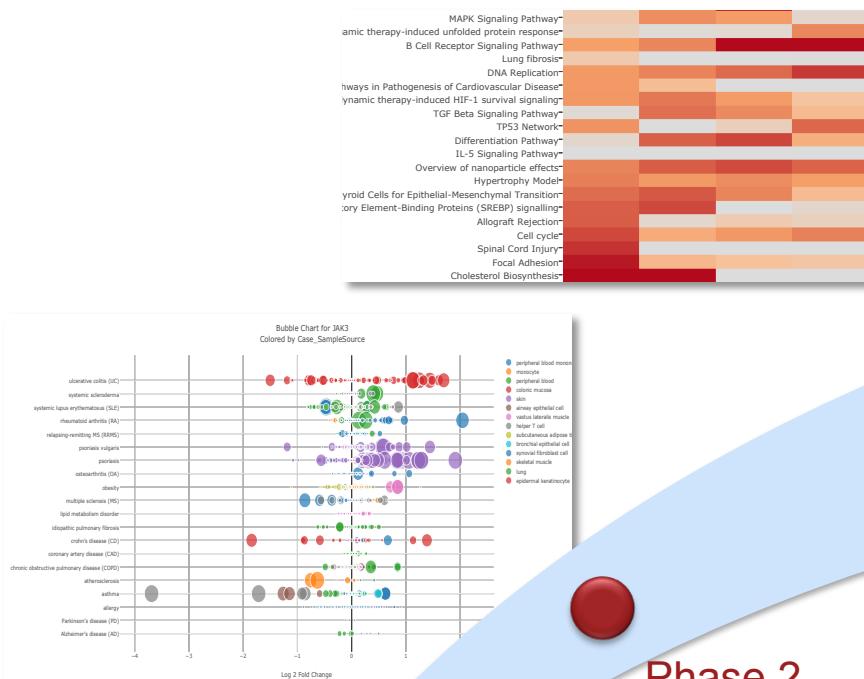


- Does people take any other medicines, such as anti-inflammation have delayed Alzheimer's?
 - Are any loss of function mutations protective from any diseases?



ProteinPaint

Roadmap to Understand Diseases Broadly and Deeply



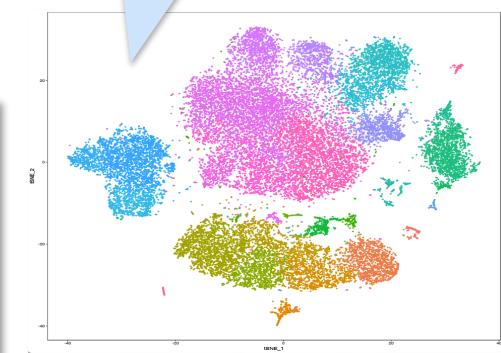
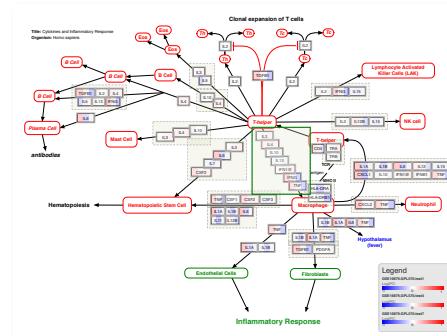
- Phase 1**
- Gene expression
 - Gene changes in diseases

- Phase 2**
- Gene set enrichment
 - Comparative pathway analysis

- Phase 3**
- Internal data sets
 - Controlled access

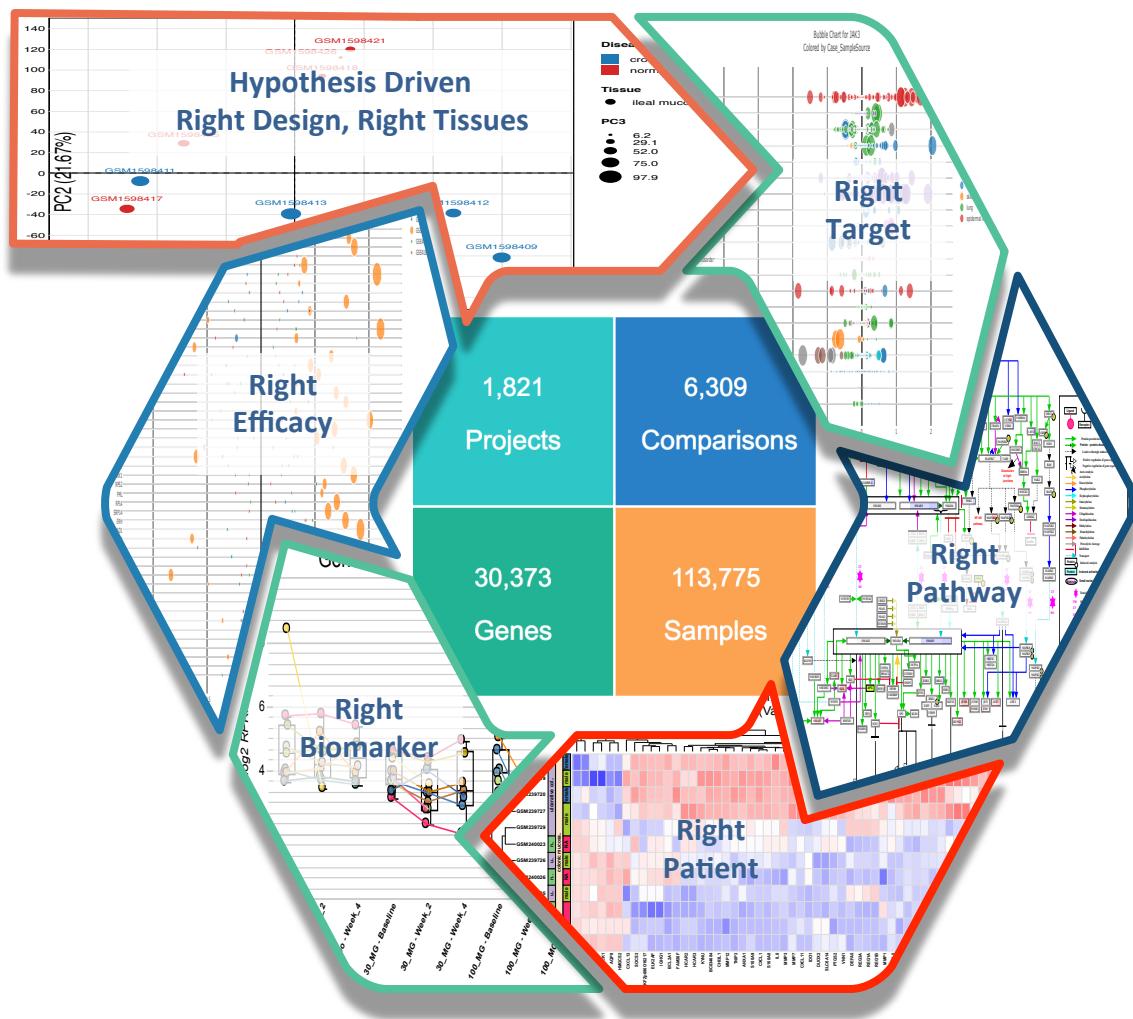
Phase 4

- Single cell
- Integrated omics



- Won 2017 WRD Achievement Award
- Reduces data acquisition cost by 90%
- Speeds up analysis from days to minutes

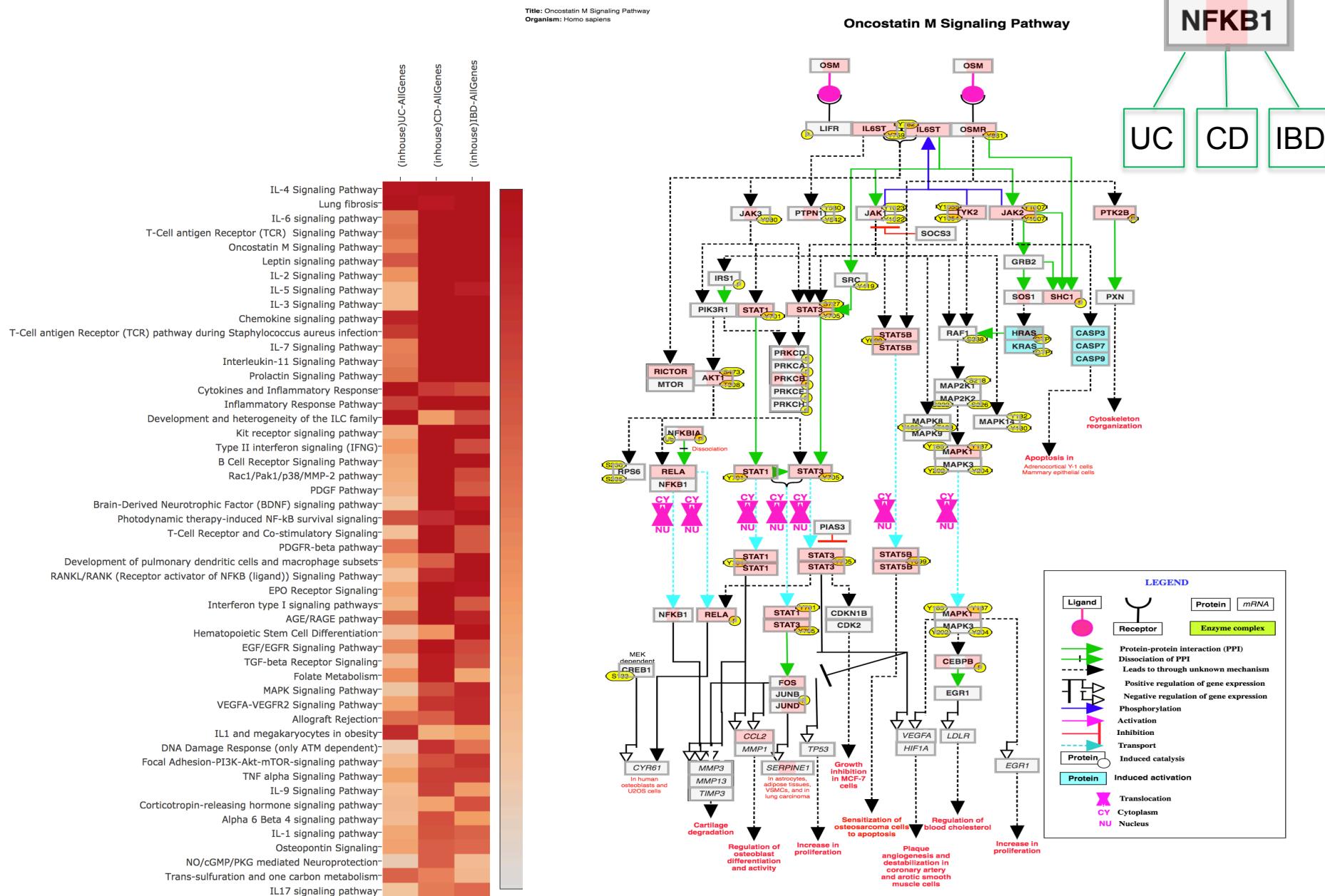
Jomics – Data Driven Decision Making



- Expression of a Gene
 - Expression of Multiple Genes
 - Bubble Plot of Differentially Expressed (DE) Genes
 - Heatmap of Gene Sets
 - Enrichment Analysis
 - Pathway View
 - Pathway Heatmap
 - Pathway Match Analysis
 - Comparative Pathway View
 - Meta Analysis
 - Correlation Analysis
 - PCA Analysis
 - Clustering Analysis
 - scRNA-seq View

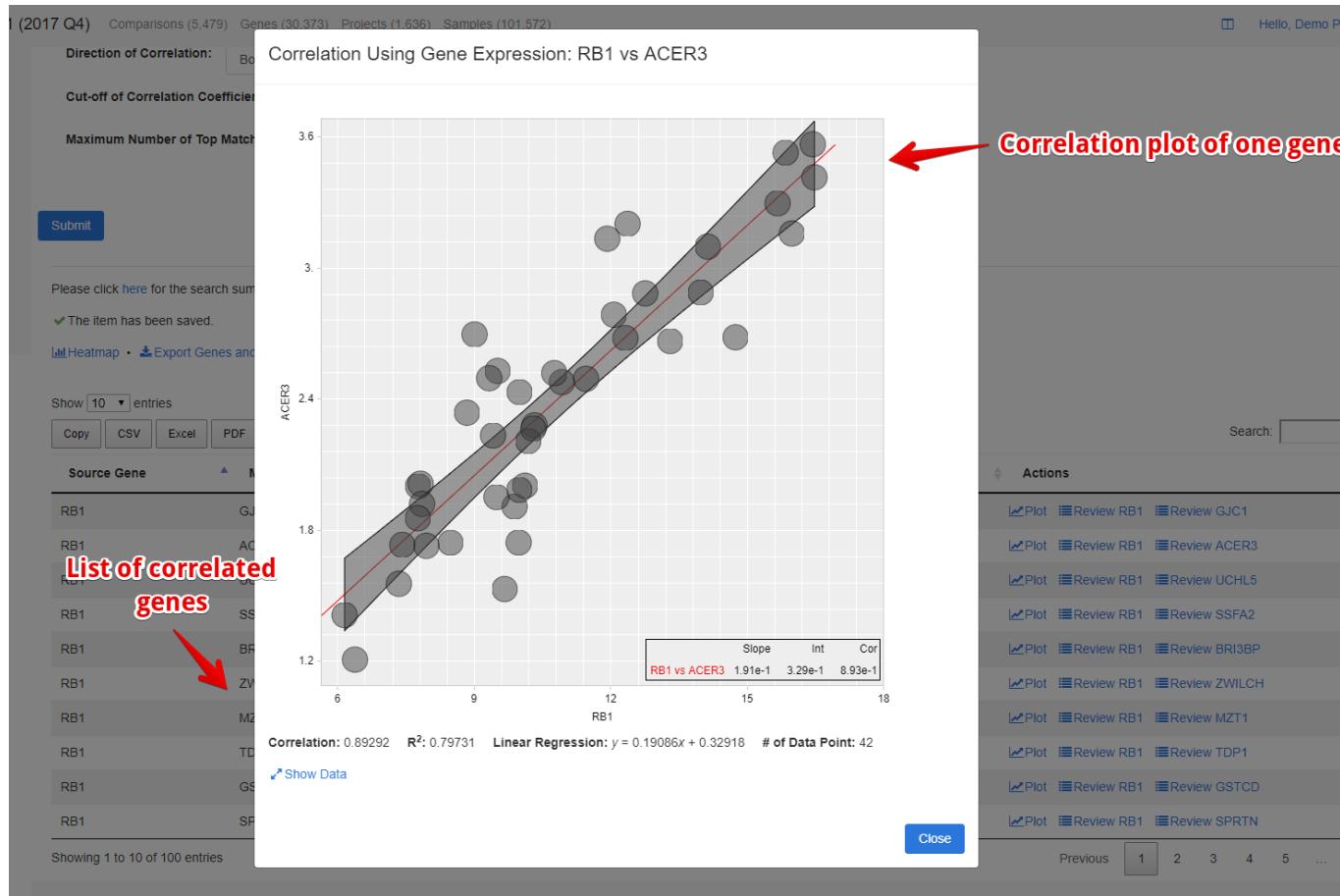
Data Source: OmicSoft (A Qiagen company) and Internal

Comparative Pathway Analysis



Genes Correlated with RB1

Correlation tools can identify other genes with similar patterns to the gene of interest.



QuickRNASeq - Reproducible Data Analysis Pipeline

<https://baohongz.github.io/ZSF1/>

ZSF1: Integrated and Interactive RNA-seq Analysis Report

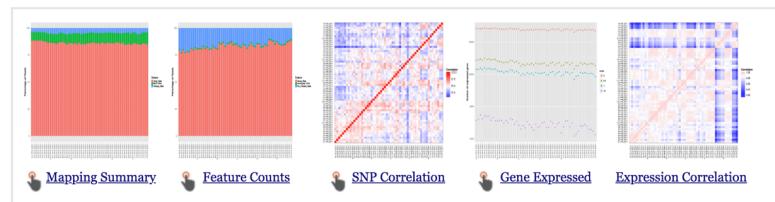
Overview

It provides the QC metrics and expression tables, including overall read mapping statistics, QC report for each sample, interactive gene expression visualization and analytical tools to gain insights about the data.

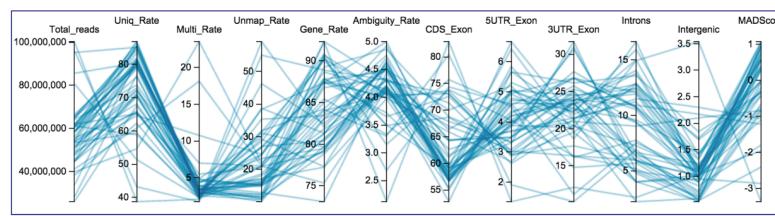
Methods

Glomerular enriched tissue was isolated by mechanical sieving of kidney cortex using an established protocol. Briefly, fresh cortex tissue (approximately one and a half kidneys per animal) was passed through a serial stack of mesh filters of 180, 106, and 75 μ pore size. An optimal yield and enrichment of glomeruli was obtained by gently forcing cortex tissue through the 180 μ mesh with the flat end of a 30 mL syringe plunger, followed by gravity flow collection on the lower mesh sizes with large volume PBS rinses. Materials trapped on the 106 μ and 75 μ mesh filters were collected in PBS by gentle flushing towards and aspiration from the edge of the mesh filter, combined, and placed back on a clean 75 μ mesh for additional large volume PBS rinses and final collection. The material was then pelleted, resuspended in Qiazol (Qiagen Cat# 79306), and stored at -80°C prior to RNA isolation. Total RNA was isolated for all samples after study termination on the same day using miRNeasy RNA isolation kits (Qiagen Cat# 217004) on a Qiacube instrument (Qiagen). For all samples, RNA quality and integrity was confirmed on an Agilent 2100 bioanalyzer. RNA sequencing was performed on oligo(dT) purified poly(A)+ mRNA. A standard TruSeq mRNA library was constructed using TruSeq RNA Sample Prep Kit v2 (Illumina, Cat.# RS-122-2001). The library was sequenced by Illumina HiSeq 2000 using a paired-end run (2 \times 100 bases). After sequencing, 40M 217 100 bp pair-end reads were generated for downstream analysis.

QC Metrics



Parallel Plot of QC

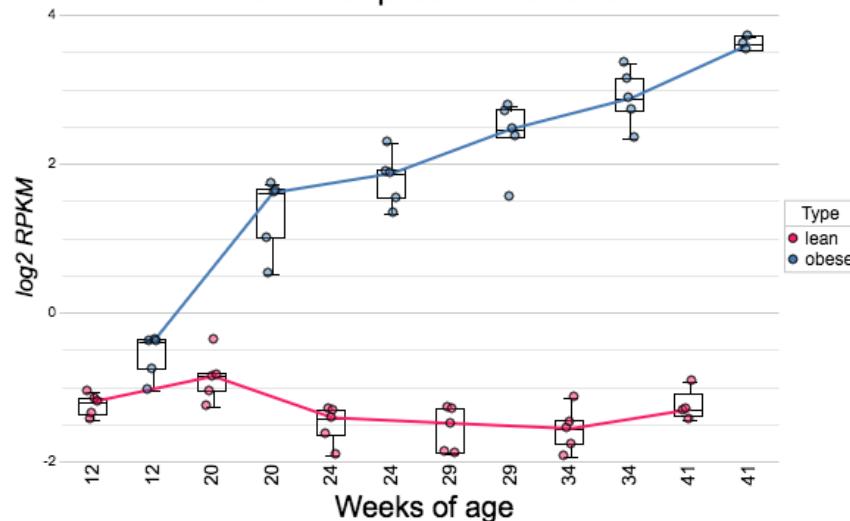


Click on the image above to access the interactive plot. [Raw Data](#)

Expression Tables

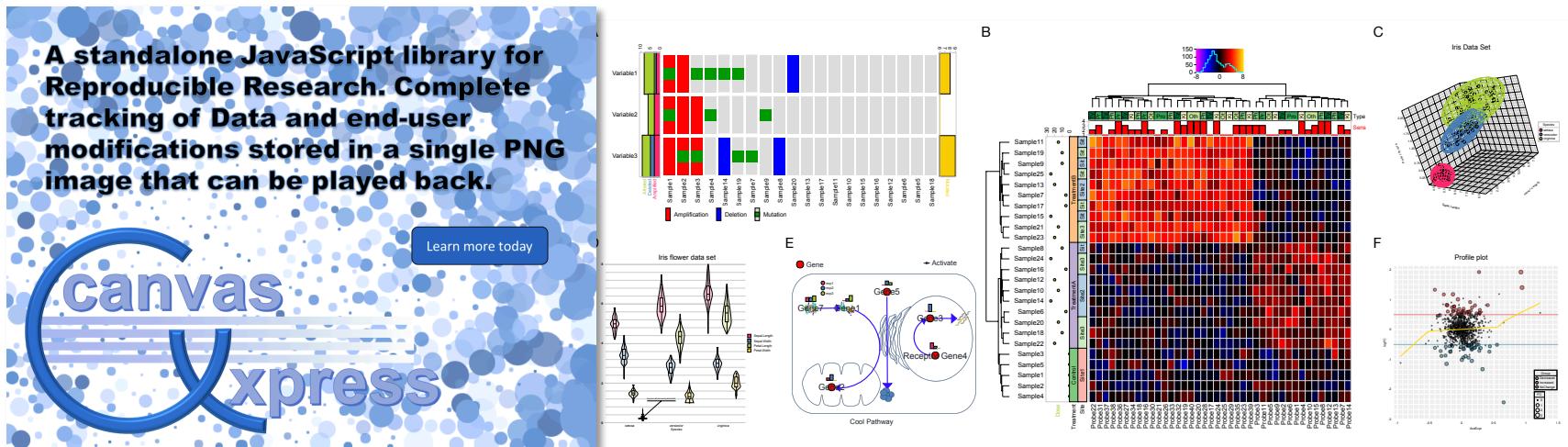
- [Raw Read Counts on Genes](#)
- [RPKM Values on Genes](#)

Gene expression of C4a



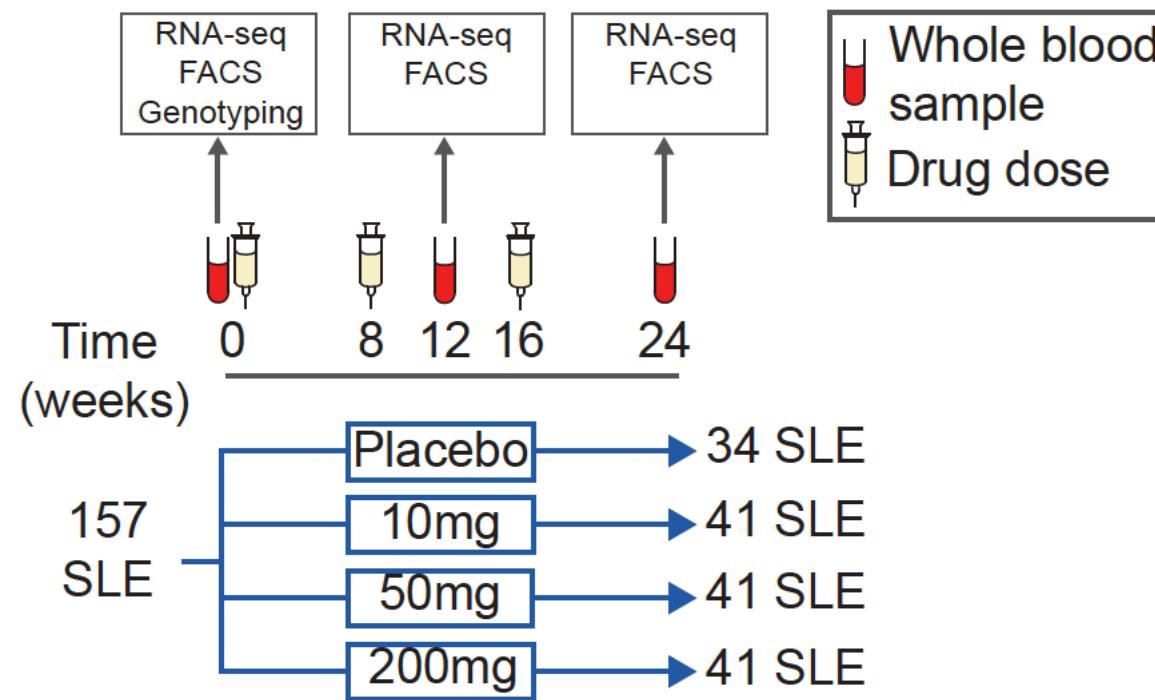
- Self-contained, database-less, server-less, network-less, easy-to-share interactive report
- QuickRNASeq is widely used in almost all Pfizer R&D lines
- Will be implemented as an enterprise solution by Genedata
- Utilized by academic collaborators from Hospital of Special Surgery, University of Rutgers, and Harvard University

Reproducible Big Data Visualization Toolkit



Molecular Profiling in Clinical Trials

- First Pfizer Phase 2 study in systemic lupus erythematosus (SLE)
- Broad specimen collection and molecular profiling to understand disease mechanism and identify biomarkers to predict response



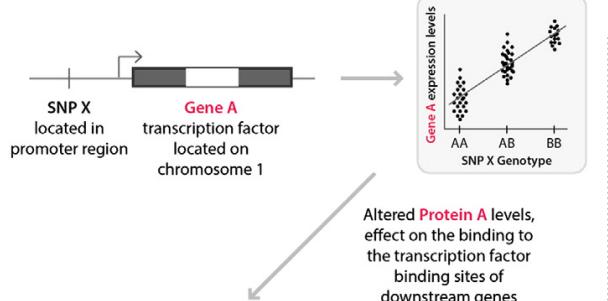
- Lupus is a chronic autoimmune disease.
- There is still no cure for lupus; there is still no known cause of lupus.
- It affects more women than men.
- There's an estimated five million people living with lupus worldwide.
- Systemic Lupus Erythematosus is the most common form of lupus and is considered more serious than the other forms. It affects many parts of a person's body including kidneys, lungs, blood, skin, and others.

Integrative Analysis Linking Genetics, Gene Expression and Transcription Factor Binding Motif

Linking Genetics with Gene Expression by eQTL (expression quantitative trait loci)

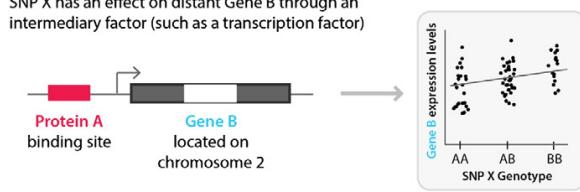
Cis-eQTL

SNP X has an effect on local Gene A

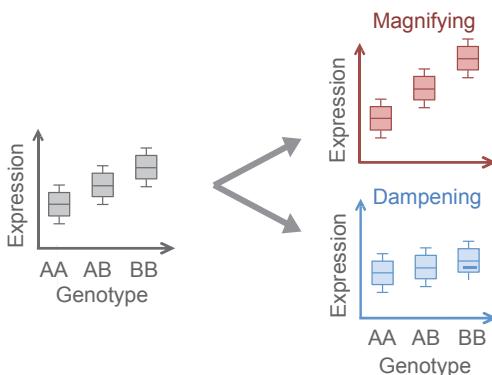


Trans-eQTL

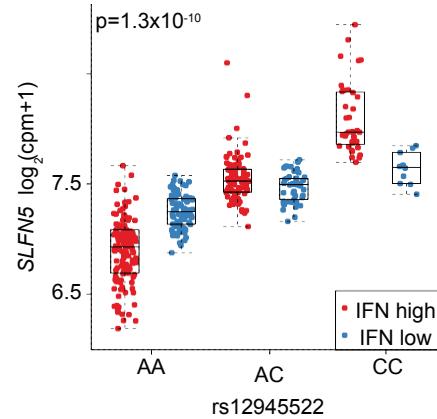
SNP X has an effect on distant Gene B through an intermediary factor (such as a transcription factor)



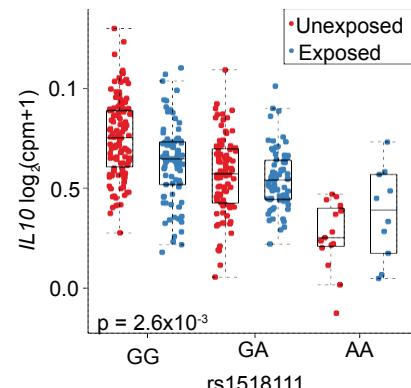
eQTL interactions can **magnify** or **dampen** the original eQTL effect



eQTL interactions with Interferon Status and drug expose

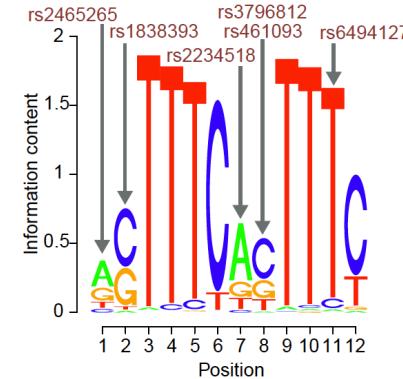


- 185 INF-eQTL interactions with $p < 0.01$
- The eQTL effect is dampened in IFN low samples in the example



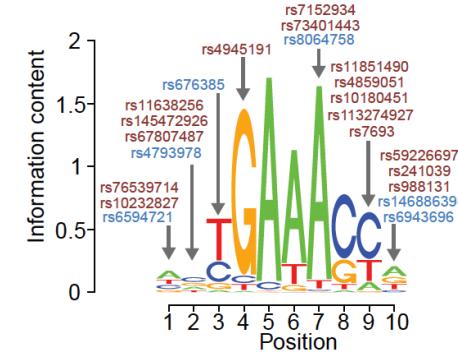
- 126 drug-eQTL interactions with $p < 0.01$
- The eQTL effect is dampened in drug exposed samples in the example

Transcription Factor Binging Enrichment



IRF1 binding site motif enriched for IFN-eQTL dampeners.

IRF1 is a transcription factor known to be important in the response to IFN.

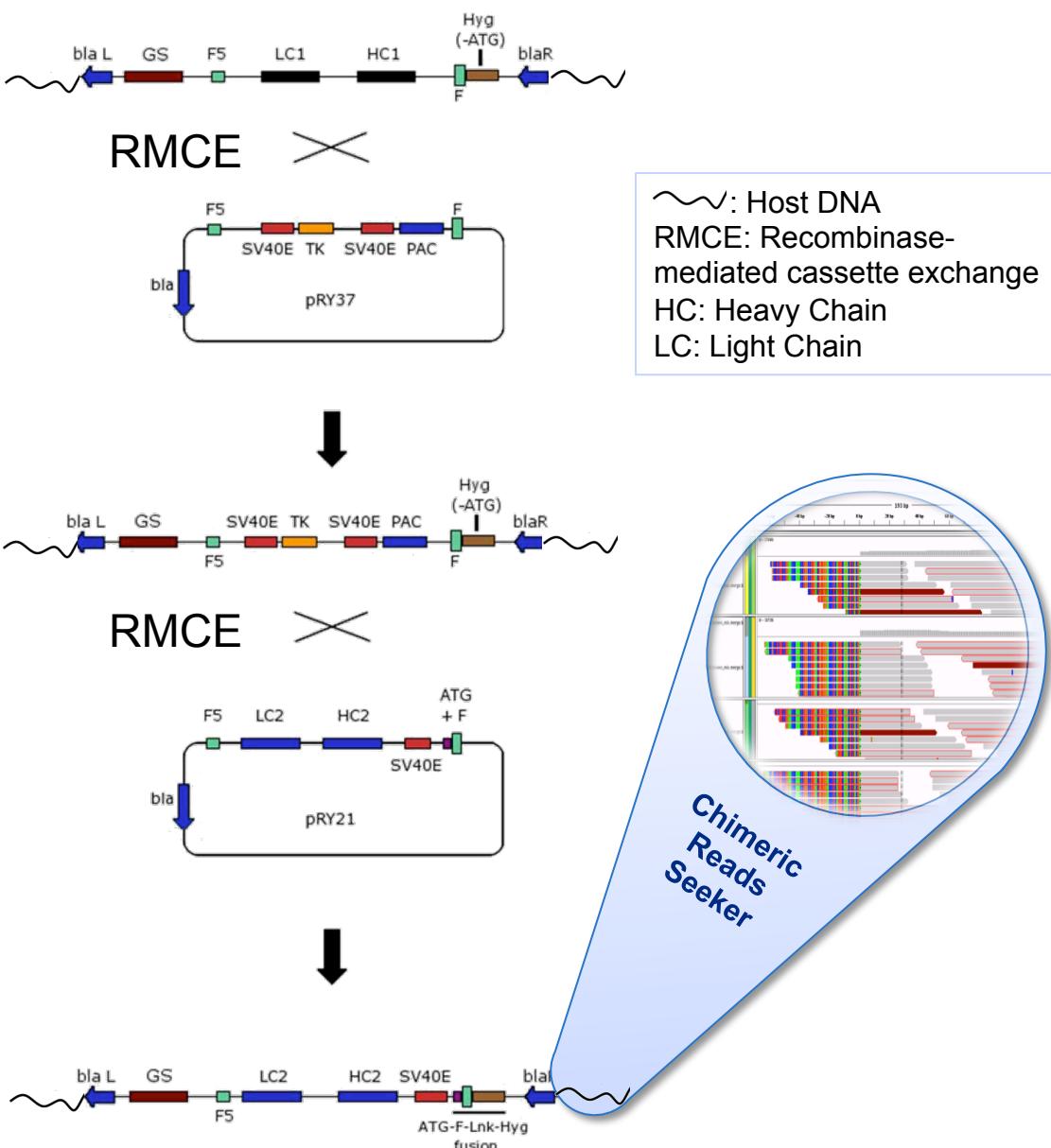


IRF4 binding site motif enriched for drug-eQTL magnifiers.

IRF4 can work downstream of IL-6 to regulate production of IL10 and other genes.

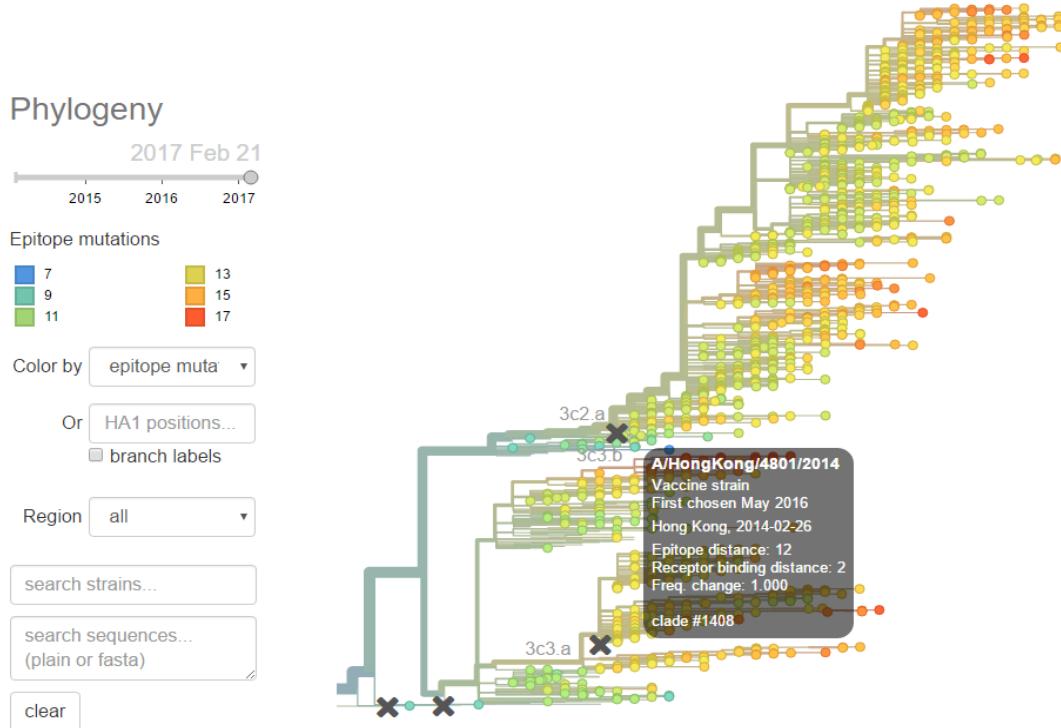
IRF4 is associated with rheumatoid arthritis⁴ (another autoimmune disease).

Sequencing in PharmaSci CHO Cell Line Engineering



- Requires resource-intensive and time-consuming efforts to screen hundreds to thousands of clones to obtain a commercially compatible production cell line.
- Minimizing the effort required when selecting high-expressing cell lines with predictable performance characteristics and consistent behavior would be of significant advantage in time and resource savings.
- Introduced next generation sequencing technology to PhamarSci CHO Cell Line engineering team to identify key integration sites of high-yield, stable therapeutic protein producers for intellectual property protection.
- Designed algorithm to search for chimeric reads, where part of the sequence is from the host genome and the other part is from the integrated vector.

FluTracker – Real Time Influenza Sequence Data Analysis



- Scientists in Vaccine RU used all the available tools in the FluTracker to predicted next vaccine viruses before the WHO recommendations!
- Nominated as a small team for 2017 WRD Achievement Award

Influenza vaccine project team was looking for tools to continuously monitor and analyze 1000's of viral sequences in support of predicting next vaccine viruses. When I approached Baohong for his input on developing a phylogenetic tree with regular automatic updates which involves little or no manual intervention, he enthusiastically suggested to work with Bioinforx . Baohong's outpouring positive attitude, attention to the details, seamless knowledge in articulating biological concepts to developers was very visible at every meetings. Though more work needs to be done, the phylogenetic tree Baohong supported building is user friendly, which is packed with ocean of information at scientist disposal.

From Sequence to Structure to Function

Fold and function predictions for *Mycoplasma genitalium* proteins

Leszek Rychlewski, Baohong Zhang and Adam Godzik

Background: Uncharacterized proteins from newly sequenced genomes provide perfect targets for fold and function prediction.

Results: For 38% of the entire genome of *Mycoplasma genitalium*, sequence similarity to a protein with a known structure can be recognized using a new sequence alignment algorithm. When comparing genomes of *M. genitalium* and *Escherichia coli*, > 80% of *M. genitalium* proteins have a significant sequence similarity to a protein in *E. coli* and there are > 40 examples that have not been recognized before. For all cases of proteins with significant profile similarities, there are strong analogies in their functions, if the functions of both proteins are known. The results presented here and other recent results strongly support the argument that such proteins are actually homologous. Assuming this homology allows one to make tentative functional assignments for > 50 previously uncharacterized proteins, including such intriguing cases as the putative β -lactam antibiotic resistance protein in *M. genitalium*.

Protein Science (1999), 8:1104–1115. Cambridge University Press. Printed in
Copyright © 1999 The Protein Society

The *Helicobacter Pylori* Genome: From Sequence Analysis to Structural and Functional Predictions

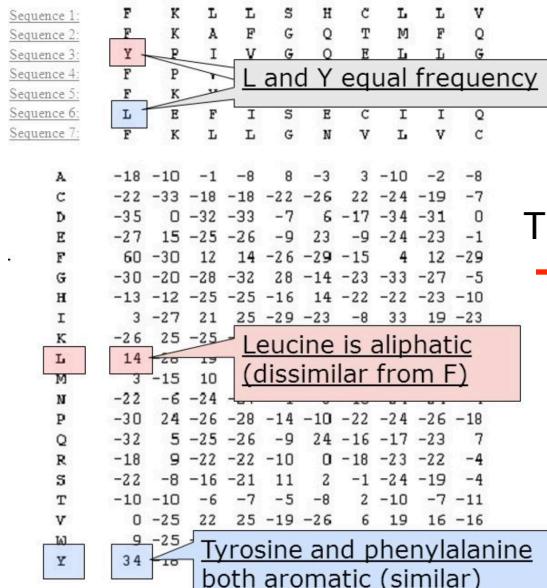
Krzysztof Pawlowski, Baohong Zhang, Leszek Rychlewski, and Adam Godzik*
The Burnham Institute, La Jolla, California

From fold predictions to function predictions:
Automation of functional site conservation analysis for functional genome predictions

Protein Science (1999), 8:614–624. Cambridge University Press. Printed in the USA.
Copyright © 1999 The Protein Society

BAOHONG ZHANG,¹ LESZEK RYCHLEWSKI,² KI
JACQUELYN S. FETROW, JEFFREY SKOLNICK,¹A

Functional insights from structural predictions:
Analysis of the *Escherichia coli* genome



Library of template structures

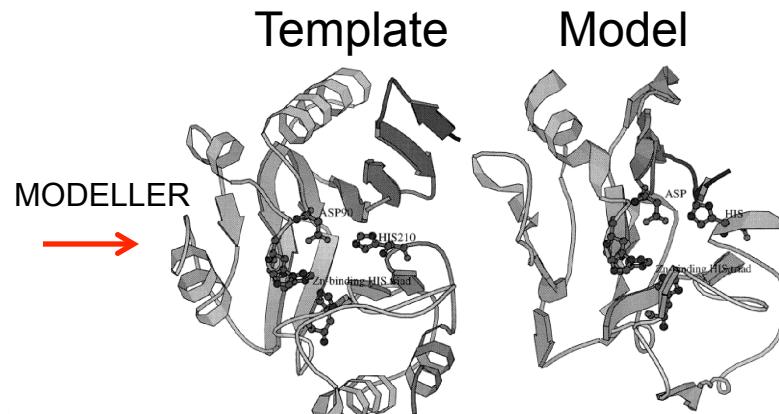
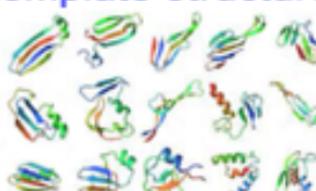
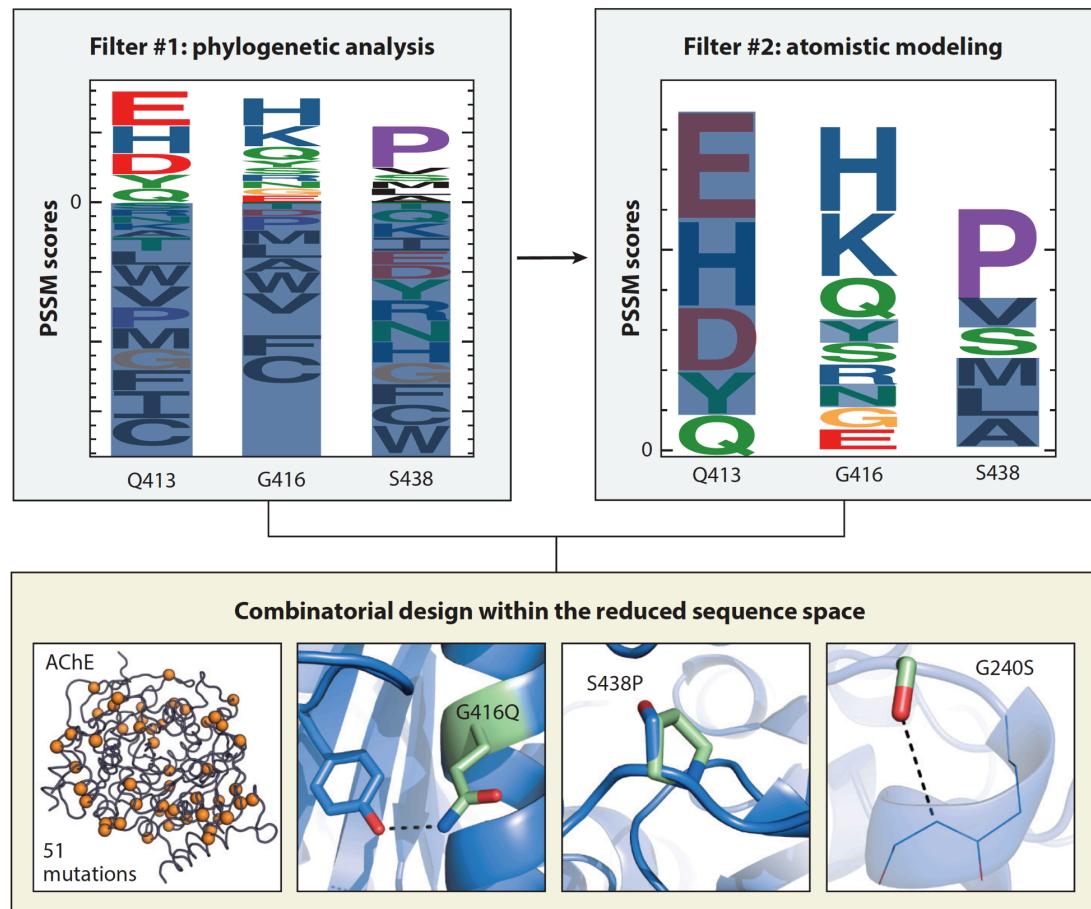
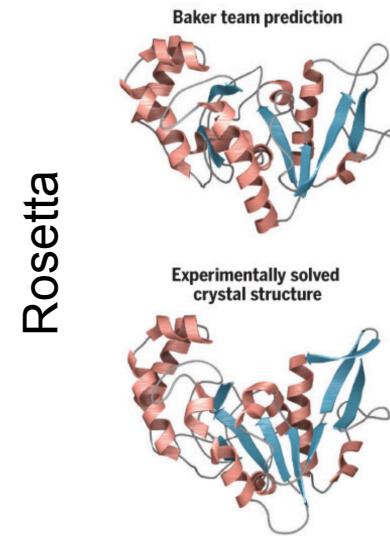


Fig. 3. Active site and catalytic residues in zinc- β -lactamase. (A) *Bacillus cereus* β -lactamase experimental structure, PDB code 1bmc. (B) A model of HP0813 based on the 1bmc template with BASIC alignment as created by MODELLER.

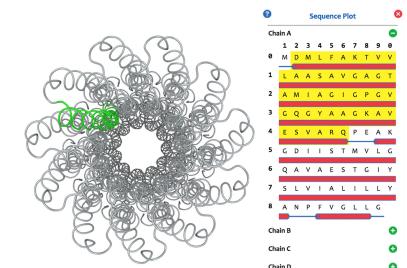
From Sequence to Structure to Function, ver. 2018



Goldenzweig et al. (2016) developed an automated algorithm for protein stabilization requiring minimal experimental testing; for instance, the five tested variants of human acetylcholinesterase showed > 100-fold higher soluble bacterial expression and higher melting temperatures than wildtype.

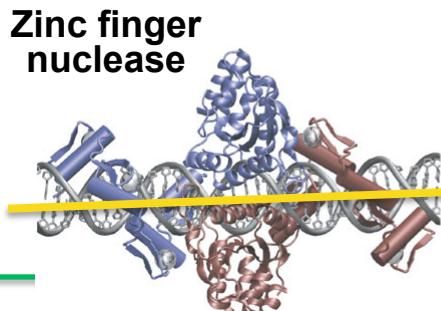


In a protein-folding competition, Baker's team stunned judges by almost matching the actual structure. v.
ALTOUNIAN/SCIENCE



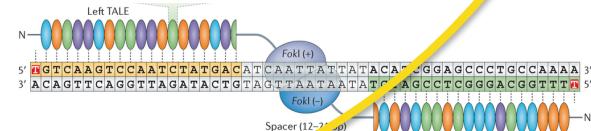
A constantly evolving technology: Gene(ome) editing toolbox is not static

Technology applications and utilization



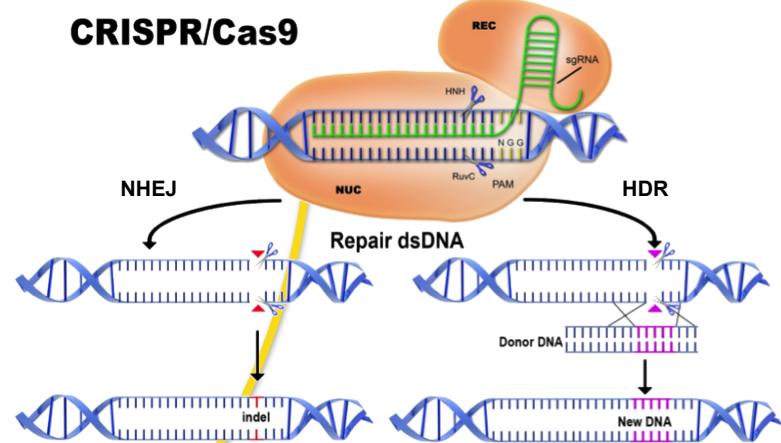
2003-2006

Many of the fundamentals of genome editing were developed with ZFN and TALEN but CRISPR has ushered in an **exponential** increase in utilization

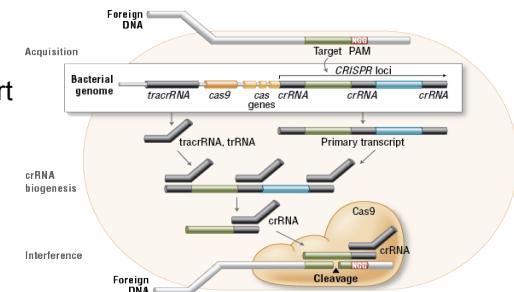


Transcription activator-like effector nuclelease

2009



2012-2014



Miller et al., *Nature Biotechnology* (2007)

Kim & Kim, *Nature Reviews Genetics* (2014)

Doudna & Charpentier, *Science* (2014)

J. Doench, *Slas* (2016)

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats
sgRNA: Single Guide RNA

PAM: Protospacer Adjacent Motif
NHEJ: Non-Homologous End Joining
HDR: Homology Directed Repair

CRISPR Improves The Entire Drug Development Pipeline

Target discovery and validation

- Functional genomic screens (LoF, GoF) to identify novel effectors of a disease phenotype
- Validation of GWAS hits – Generation of isogenic cell lines

Assay development and screening

- Knock-in (KI) of reporters into endogenous gene loci
- Utilization of genome-wide libraries that can Knock-out (KO) genes or modulate expression in high-throughput screens

Mechanism of action determination

- CRIPSR KO or gene expression modulation can be used in parallel with small molecules to identify targets. Can find genes that synergize with small molecule phenotypes
- Orthogonal validation of siRNA and shRNA phenotypes

In vivo analysis of efficacy, safety, & disease

- Accelerate KO and KI rodent model generation – knockout in 1-2 months and knockin in 2-4 months
- Ability to edit DNA easily to create transgenic of many species: cow, monkey, dog, sheep, pig, etc.

Biologics and bioprocess

- Modify cell lines for enhanced bioprocess efficiency (custom CHO, microbes, etc.)
- Integrate reporters for improved yield and scalability

In-house Interactive Guide RNA Designing Tool

View Submission Details

Mouse over for details. ■ High Quality Guide ■ Mid Quality Guide ■ Low Quality Guide All Guides ▾

Guide #19 (50) ► ■
Guide #5 (70) ► ■
Guide #16 (53) ► ■
Guide #17 (53) ► ■
Guide #21 (30) ► ■
Guide #10 (60) ► ■
Guide #2 (74) ► ■
Guide #1 (79) ► ■
Guide #3 (74) ► ■
Guide #18 (52) ► ■
Guide #8 (63) ► ■
Guide #10 (61) ► ■
Guide #13 (57) ► ■
Guide #6 (68) ► ■
Region_A
Reference
Region_B
Guide #4 (70) ► ■
Guide #15 (56) ► ■
Guide #11 (61) ► ■
Guide #7 (66) ► ■
Guide #9 (62) ► ■
Guide #20 (50) ► ■
Guide #14 (57) ► ■

Custom features

Guides ranked by score
Click a guide to view offtarget at right side

Guides on forward strand (color coded for quality)

Target sequence

Guides on reverse strand

All Guides Download All Guides & Offtargets

Click help for more

Info about score

Off targets for selected guide

Check here to view offtargets in exons

Aggregated Score: The higher the better. This score shows inverse likelihood of offtarget binding.

Click the guide number to view the offtargets.

Aggregated Score	Sequence
Guide #1	79 TTCAAGGGCTCACTGACCGT CGG
Guide #2	74 CTTCAGGGCTCACTGACCGT TGG
Guide #3	74 AGGGCTCACTGACCGTGGG GGG
Guide #4	70 ATGTCCGAATAAAGAGTC TGG
Guide #5	70 GGCACAGCACACACTGGG AGG
Guide #6	68 ATTTCAGTCGGCTTGAAG AGG
Guide #7	66 CACAGGGATTGCTCAGCTGG AGG
Guide #8	63 AGCATAGTTTAGACAATAG AGG
Guide #9	62 GAAATCRAAAGACAGGTGG AGG
Guide #10	61 GCATAGGTTTAGACAATAGA GGG
Guide #11	61 ACAGGGATTGATCCAGTCA GGG
Guide #12	60 GCAGACTGACCTCACTTCAG GGG
Guide #13	57 CATAGGTTAGACAATAGAG GGG
Guide #14	57 CAAAGTCAGGACTCTCAC CCG
Guide #15	56 ATGCCAACTACTAAATTC TGG
Guide #16	53 TGTCCCCAGCTCTTTATG CCG
Guide #17	53 TATACCCAAATTAGTAGT TGG
Guide #18	52 GTGACTGGAATCACAGCAT AGG
Guide #19	50 GGTGGGCACGCRCAACCTA GGG
Guide #20	50 CACTCATCCATCAAGC CGG
Guide #21	30 GAGGTGGTAACAGAAAGGTC AGG

Guide #21 Download

Quality Score: 30

info about score (Strand +)

Guide Sequence: GAGGTTGGTAACAGAAAGGTC AGG

On-target Locus: superscaffold23:17418331-17418353+ (Details)

Offtarget Sites (862)

The score for offtarget measures the likelihood of offtarget binding. Show Exon Only

No.	Sequence	Offtarget Score	Mismatches	Annotations	Locus
1.	GAGGTTGGTAACAGAAAGGTC AGG	20	1		C56984333:104-126:-
2.	GAGGTTGGTAACAGAAAGGTC AGG	20	1		scaffold316:87462-87484:-
3.	GAGGTTGGTAACAGAAAGGTC AGG	5.028	2		superscaffold3:27421193-27421215:-
4.	GAGGTTGGTAACAGAAAGGTC AGG	5.028	2		superscaffold40:14399861-14399883
5.	GAGGTTGGTAACAGAAAGGTC AGG	5.028	2		superscaffold7:13515676-3515698:+
6.	TAGGTTGGTAACAGAAAGGTC AGG	1.984	3		superscaffold16:19791589-19791611
7.	AAGGTTGGTAACAGAAAGGTC AGG	1.673	3		superscaffold14:933578-933600:+
8.	GAAGGTTGGTAACAGAAAGGTC AGG	1.535	3		superscaffold103:1573534-1573565:-
9.	GGGGTTGGTAACAGAAAGGTC AGG	1.53	3		scaffold118:97356-97378:-
10.	AGGTTGGTAACAGAAAGGTC TGG	1.491	4	Intron: AKP13_HUMAN	superscaffold73:431520-431524:-
11.	AAAGGTTGGTAACAGAAAGGTC AGG	1.471	4	Intron: Dzip11	superscaffold6:6289041-6289063:+
12.	CAGGGTTGGTAACAGAAAGGTC AGG	1.458	4	Exon: CH0000548.1; Upstream: CH0000548.1	C56524390:128-150-
13.	GAAGGTTGGTAACAGAAAGGTC TGG	1.388	4		scaffold148:101790-101812:-
14.	GACRGGGTATCAGAAAGGTC TGG	1.342	4		superscaffold7:27167760-27167782:-
15.	GGGGTTGGTAACAGAAAGGTC AGG	1.341	3	Downstream: SPAG1_HUMAN	superscaffold104:1619724-1619746:-
16.	GAAGGTTGGTAACAGAAAGGTC TGG	1.284	3		superscaffold54:6503404-6503426:+
17.	CATGAGGTTGGTAACAGAAAGGTC AGG	1.161	4		scaffold216:243682-243704:+
18.	CATGAGGTTGGTAACAGAAAGGTC TAG	1.161	4		superscaffold122:1661591-1661613:-
19.	GAAGTGGGGTGGTAACAGAAAGGTC AGG	1.16	3		superscaffold80:3737857-3737879:+
20.	AACTGGGGTGGTAACAGAAAGGTC TAG	1.152	4		superscaffold92:689008-689030:-
21.	GAGGTGGTAACAGAAAGGTC TAG	1.128	3	Intron: Ephx1	superscaffold18:16682657-16682679:-
22.	CACTCATCCATCAAGC CGG	1.083	2		superscaffold47:318337-318350:-

- Integrated two open-source tools, Cas-Finder and Cas-OFFinder
- Guides are scored based on penalties on off-targets
- Off-targets are mapped to genes
- Locus can link to genome browser
- Primer design on the fly
- Helped CHO cell line engineering group and Vaccine research unit to advance projects

Aach J, Mali P, Church GM. 2014. CasFinder bioRxiv doi: 10.1101/005074

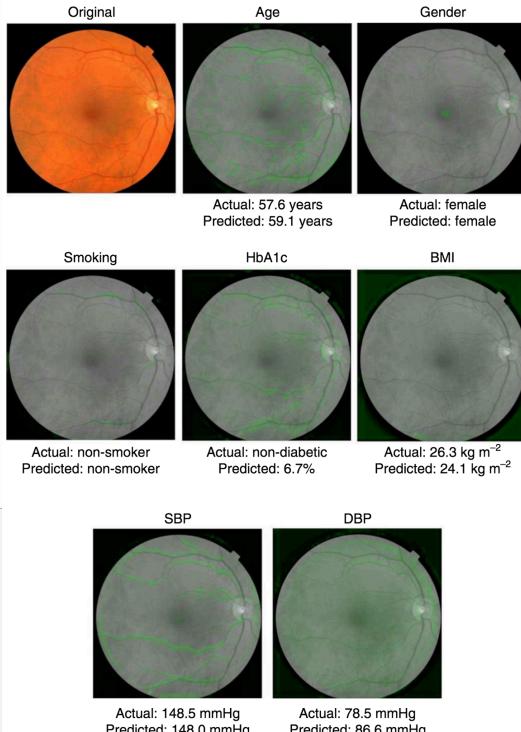
Bae S, Park J, Kim JS. Cas-OFFinder. Bioinformatics. 2014 May 15;30(10):1473-5.

What AI Excels at: Pattern Recognition in Large Data



Nature, Oct 19, 2017, 550:354-359

284K Patients, 1.7M images from UK Biobank and EyePACS



Breakthrough image based diagnostics powered by million scale images

Learning new, related categories requires less data.



Opportunity for sgRNA design
Cas modeling?

Google AI Assistant



"Do you have anything between 10 am and 12 pm?"

Real Person



"We do not have a 12 pm available.
The closest we have to that is a 1:15."

Google I/O, May 8, 2018

Images ✓

Board Games ✓

Natural Language ✓

Development of sgRNA On-target, Off-target Design Tools

Alignment Based

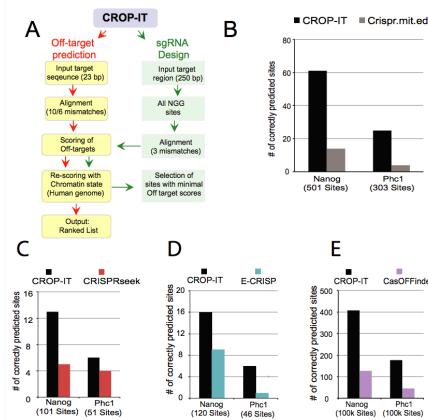
A. DNA Target site	guide sequence	PAM sequence
	CCACTGTGTGCACTTCATCCTGG	
B.		sequence (fasta)
gRNA Identification Key Parameters <i>PAM.size</i> <i>PAM</i> <i>gRNA.size</i> <i>gRNA.pattern</i> <i>findRNAsWithREcutOnly</i> <i>REpatternFile</i> <i>minREpatternSize</i> <i>overlap.gRNA.positions</i> <i>findPairedgRNAOnly</i> <i>min.gap</i> <i>max.gap</i>		Description PAM length PAM sequence guide length required or excluded sequences (regular expressions) require overlapping RE site file with all RE patterns minimum RE pattern length re set which guide positions must site require paired configuration distance between paired sites

sgRNAs are aligned from the given genome purely by locating the PAM

sgRNAsCas9, CRISPRdirect
CRISPRseek, Cas-OFFinder,
 CasFinder. CRISPRer

2013-

Hypothesis Driven

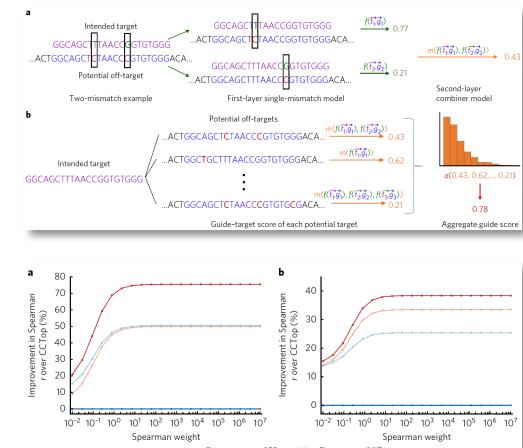


sgRNAs are aligned and scored with the contribution of specific factors (e.g., GC content, exon position, etc.) to sgRNA on-target efficacy

CRISPR-ERA, CHOPCHOP,
 Cas-Designer, CROP-IT

2015-

Machine Learning / Artificial Intelligence



sgRNAs are scored and predicted from a training model by considering different features affecting the efficacy

Elevation, Azimuth

2018+

36 sgRNA design tools and the list is growing.

We Are One

Acknowledgements

Precision Medicine

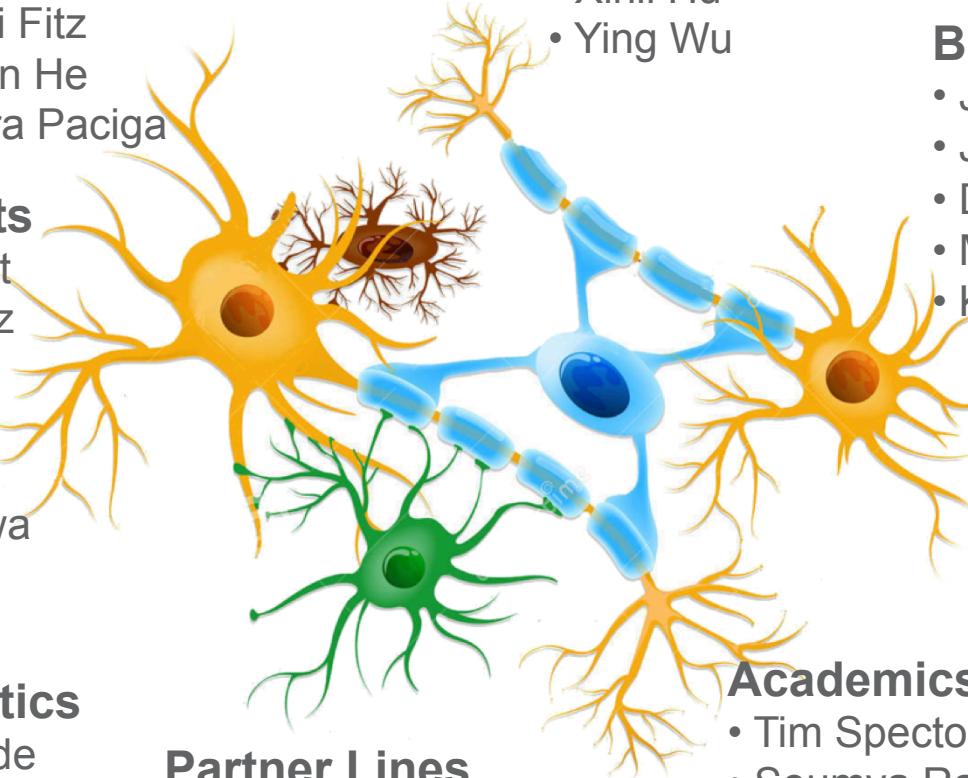
- Shanrong Zhao
- Ramya Gagini
- Lori Fitz
- Wen He
- Sara Paciga

Research Units

- Michael Vincent
- Robert Martinez
- Li-Ling Lin
- Ken Dower
- Bryce Johnson
- Christine Bulawa
- Will Gordon

Biostatistics

- Craig Hyde
- Huanyu Zhou



Partner Lines

- Lin Zhang
- Jessica Min-Debartolo
- Vishal Vaidya

Human Genetics

- Nan Bing
- Stefan McDonough
- Xinli Hu
- Ying Wu

Business IT

- Jay Bergeron (licensing)
- John Morris (storage)
- Dave Kannamer (server)
- Mike Miller (HPC)
- Kirk Watrous (cloud)

Ex-Pfizer

- Sally John
- Ciara Vangjeli
- Serena Scollen
- Chi Zhang

Academics

- Tim Spector (KCL)
- Soumya Raychaudhuri (Broad)
- Nir Hacohen (Harvard)
- Betsy J. Barnes (Rutgers)
- Marco Colonna (Washington U.)
- Accelerated Medicine Partnership

Challenges of Working as a Multidisciplinary Team

"The black sheep is beautiful!"
says the biologist

"Oh, no!" shouts the computer
scientist, "A special case!"

"Nah!" shouts the statistician,
"It is not statistical significant!"

"Aha," says the
engineer, "I see that
Scottish sheep are
black."

"Hmm," says the physicist,
"You mean that some Scottish sheep
are black."

"No," says the mathematician,
"All we know is that there is at least
one sheep in Scotland, and
that at least one side of that one sheep
is black!"





Blank Slide