# Design of individual gRNAs and Genome-wide Libraries of gRNAs

*Workshop 4*

Keystone Symposium, Precision Genome Engineering
Breckenridge, CO, January 8 – 12, 2017

John Doench
Broad Institute of MIT and Harvard

BROAD INSTITUTE

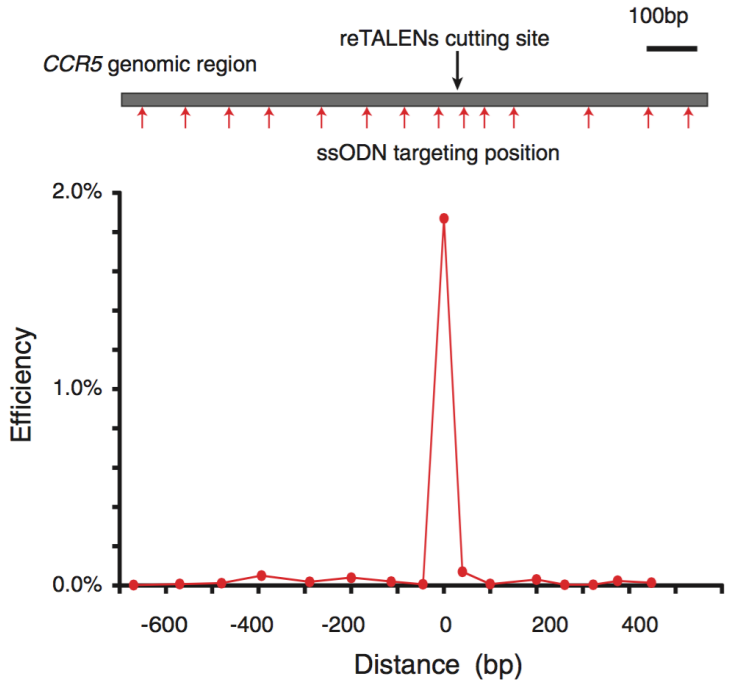
# I want the best tool for the job!
## *…but there are a lot of jobs out there!*

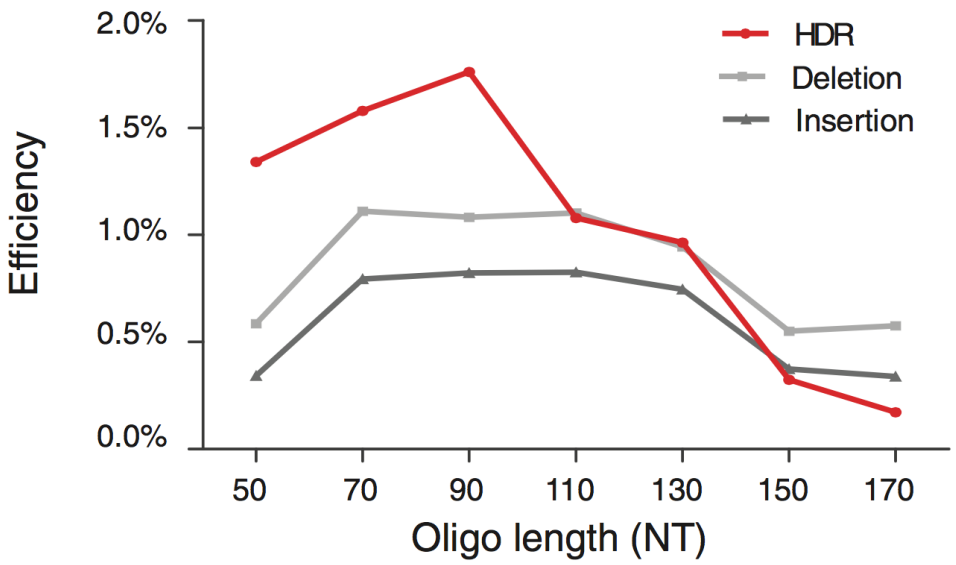
- Purpose
  - Gene editing or gene disruption
- Time
  - Long-term model or short-term investigation
- Scale
  - Study your favorite gene or genome-wide screen
- Delivery of Cas9
  - Lentivirus, AAV, plasmid, mRNA, protein
- Delivery of guide RNA
  - Lentivirus, AAV, plasmid, cr + trRNA, sgRNA

"The best" will vary based on what you're doing

# Gene editing

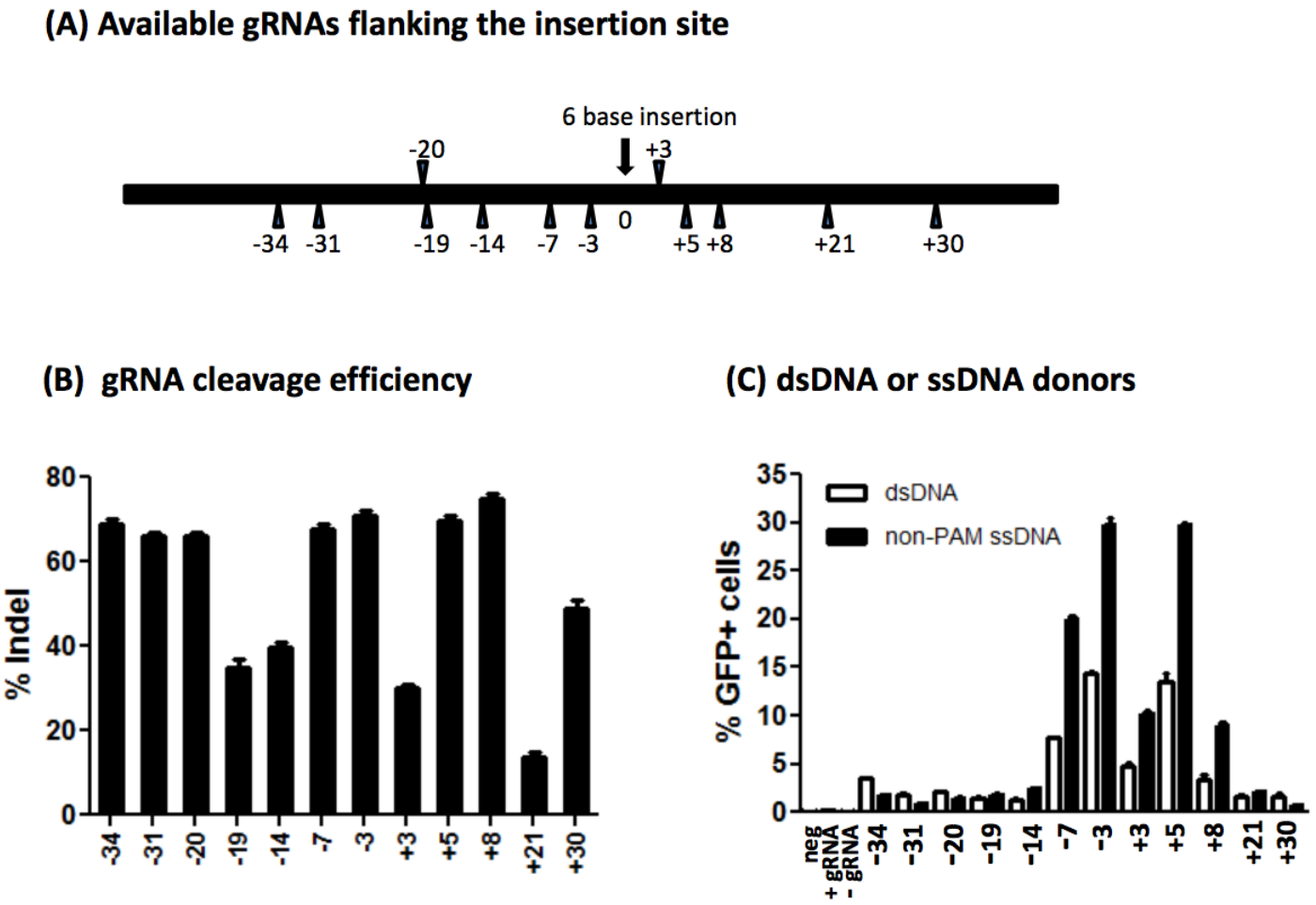
- Guide design is largely irrelevant, because you have few choices
  - Want the cut site to be as close as possible to the intended mutation




Yang, L. *et al.* Optimization of scarless human stem cell genome editing. *Nucleic Acids Research* **41,** 9049–9061 (2013).

# Gene editing

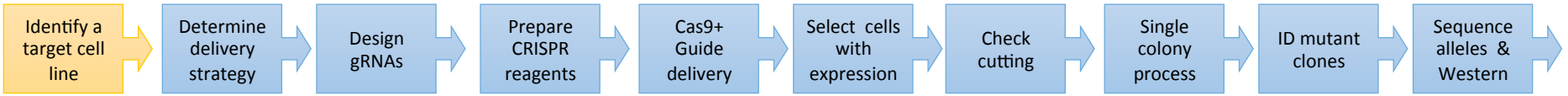



Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *Journal of Biotechnology* **241,** 136–146 (2017).

# What are the choices for gene editing?

- Large edit that can carry a selectable marker (i.e. knock-in) or single-base change?
- How well do your cells single cell clone?
  - If poorly, can you get away without single cell cloning with a smart assay?
- Modified RNA components?
- Modified Cas9 enzymes?
- NHEJ-inhibitors?
- Delivery delivery delivery

# CRISPR workflow for cell line engineering: Selecting a cell line




- HEK293, HCT116, and K562 cell lines are the most commonly used and represent the simpler cell lines to edit
- Choose a cell line that **best captures your biology** or question of interest and then build a CRISPR delivery solution around it
- **Characterize your cell lines!**
  - Copy number, best method of transfection, capable of single cell expansion, SNPs in gene of interest
- Unless a primary cells such as fibroblast can undergo 20-30 doublings without senescing, they are usually *not* amenable
  - Consider iPSc or immortalization
  - ES/iPSc editing requires special workflows to maintain pluripotency throughout the editing process

Make sure you fingerprint/ID and mycoplasma test your cell lines before starting a genome editing workflow – **garbage in, garbage out**

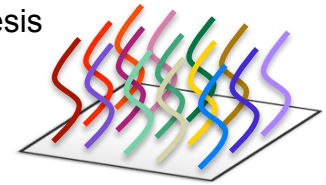
*Slide courtesy*
*Sam Hasson, Pfizer*

# For gene editing

- Design is more about *experimental* design rather than choosing the right guide RNA, because choices are so limiting. So, workflow:
  1. Optimize delivery to your system using pre-validated reagents (e.g. Cas9 that you know works, guide that you know works, etc.)
     - Also use these experiments to monitor cell viability
  2. Develop an assay to assess successful gene editing on the pool of cells (e.g. PCR & MiSeq)
  3. Test several combinations of guides and repair templates and used pooled assay from step 2
  4. Move forward with combos that work best, preferably with two different guides identified in step 3

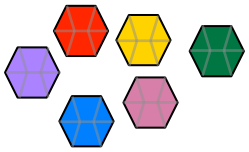
# Genome-wide Pooled Screening

1. Oligonucleotide Synthesis

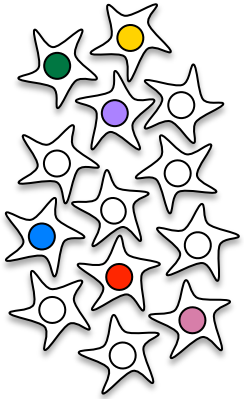


2. Plasmid Pool

3. Virus Pool



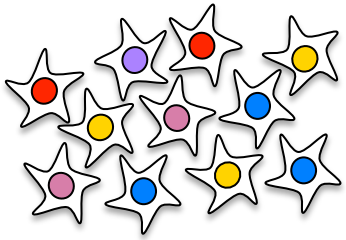
4. Pooled Infection



Pooled Screening

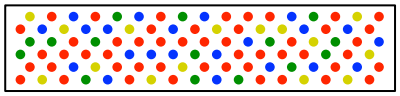
5. Positive Selection

5. Negative Selection

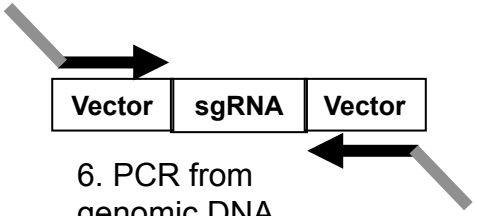


| sgRNA | Init. | Pos. | Neg. |
|---|---|---|---|
| | 16 | 0 | 21 |
| | 19 | 100 | 19 |
| | 17 | 0 | 0 |
| | 18 | 0 | 26 |
| | 14 | 0 | 18 |
| | 16 | 0 | 16 |

8. Matrix of sgRNA Abundance

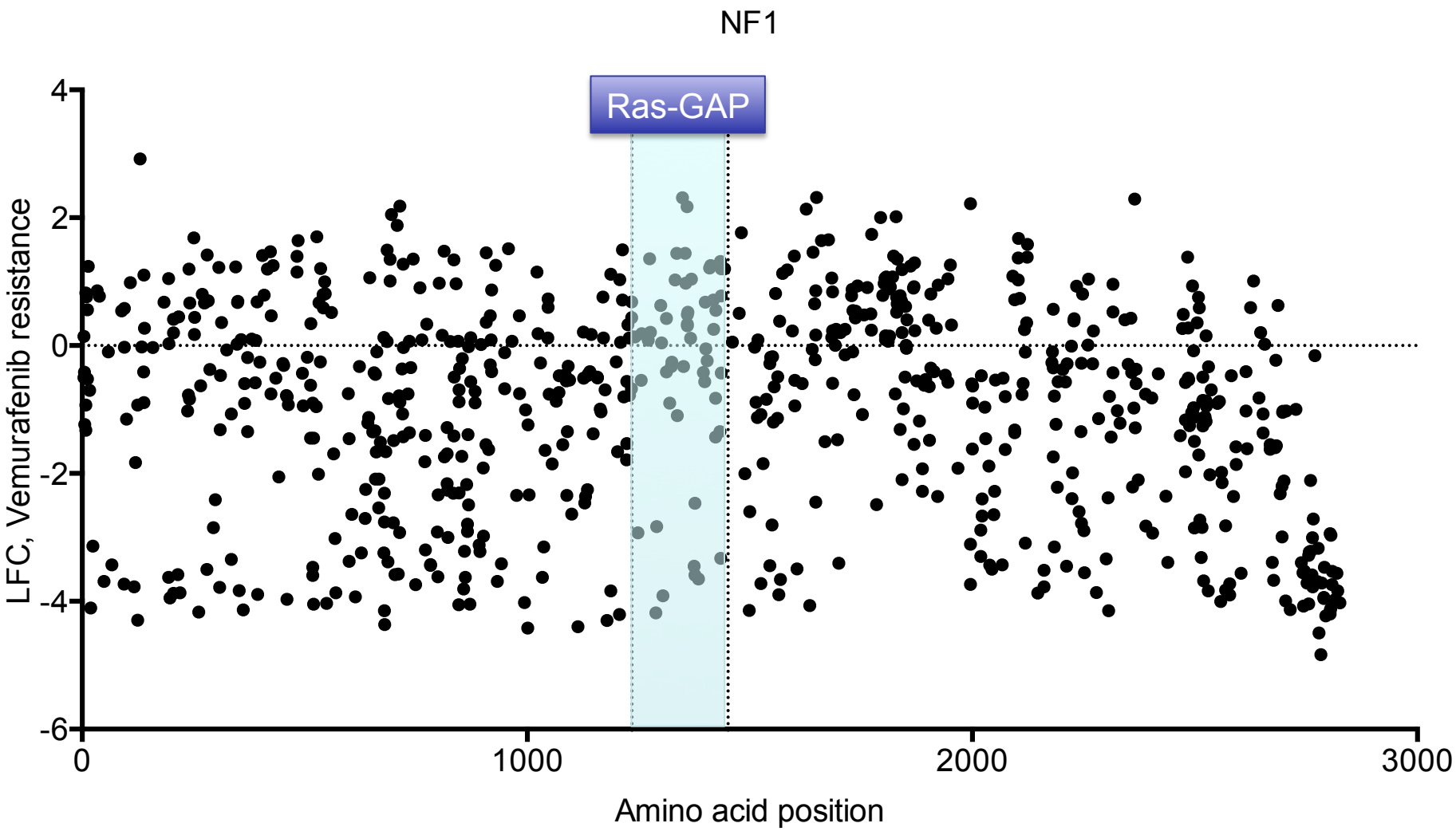


7. NextGen Sequencing

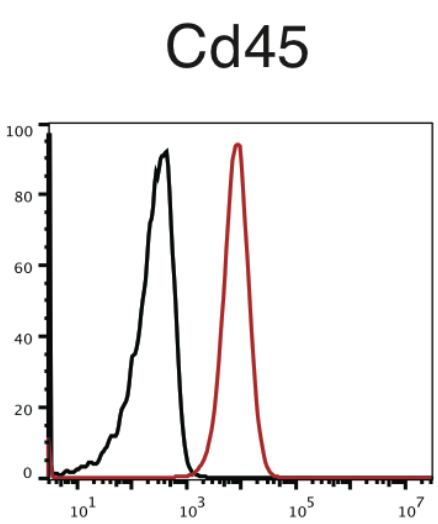


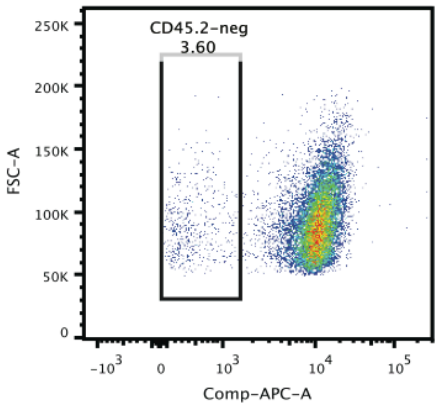

6. PCR from genomic DNA

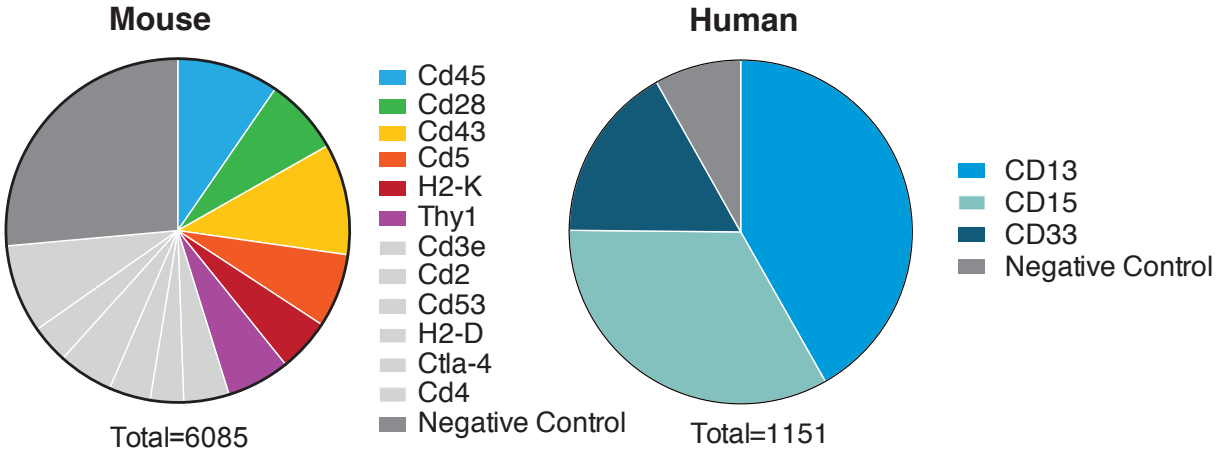
# Variability in sgRNA efficacy



# FACS based-readout: clear isolation of negative populations



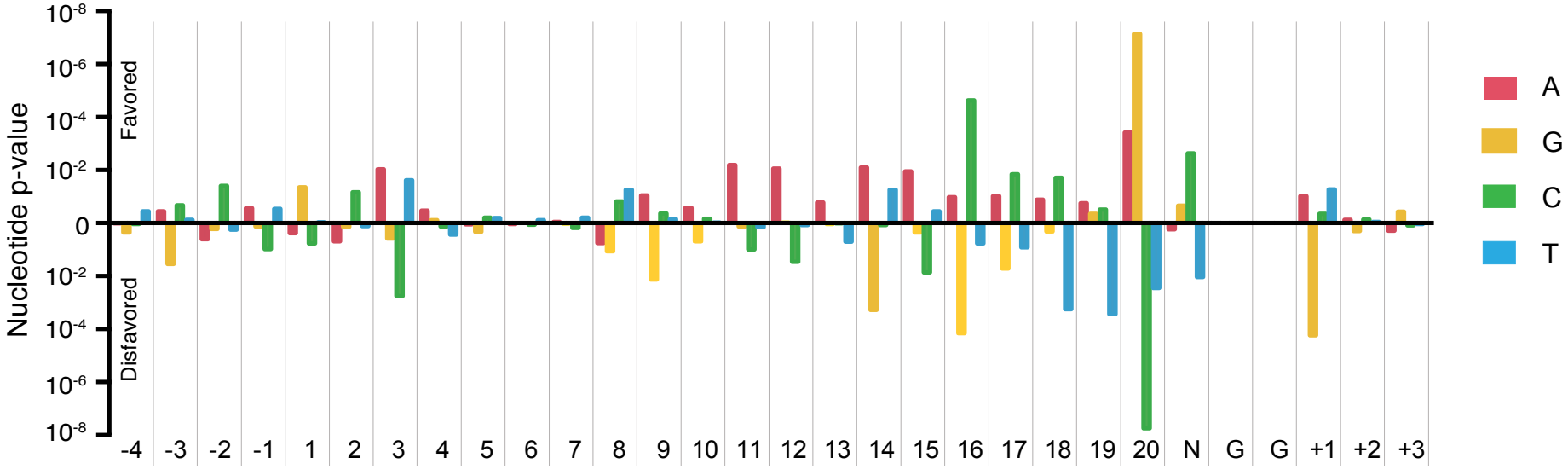




- Create all possible sgRNAs against panel of cell surface markers
- Isolate cells that are negative for these markers
- Sequence to determine the frequency of each sgRNA
- **Enriched** in the marker-negative population are **active** sgRNAs
- Use of positive selection assays to determine rules means we are *not* selecting for promiscuity, in contrast to rules based on cell death


Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat Biotechnol* **32,** 1262–1267 (2014).

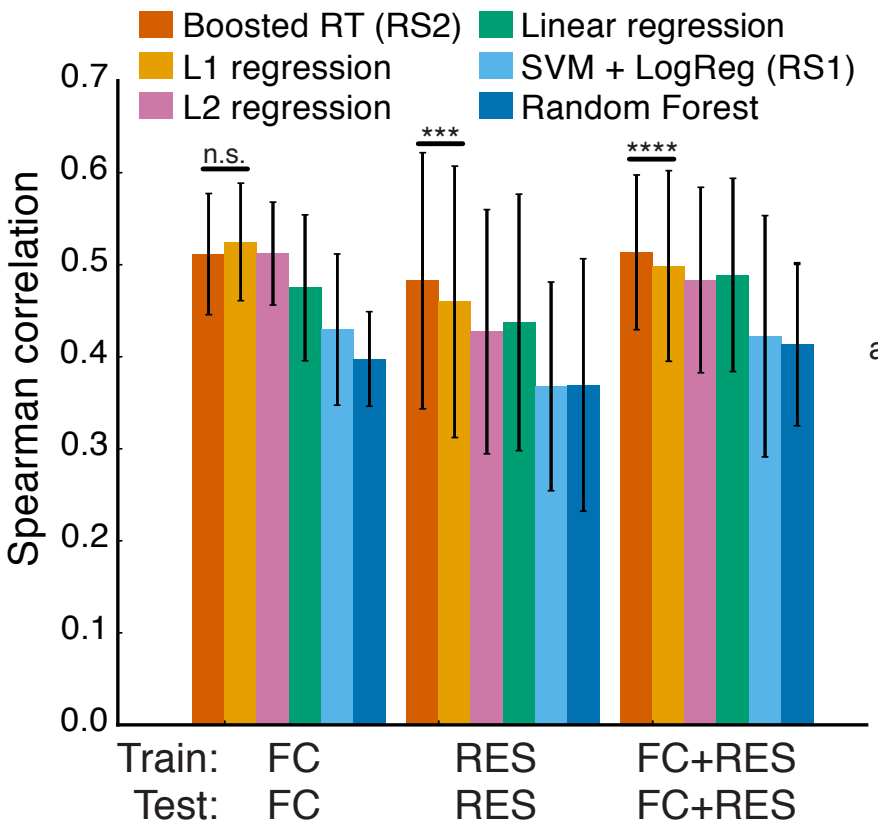
# Nucleotide frequencies indicate preferences for highly-active sgRNAs

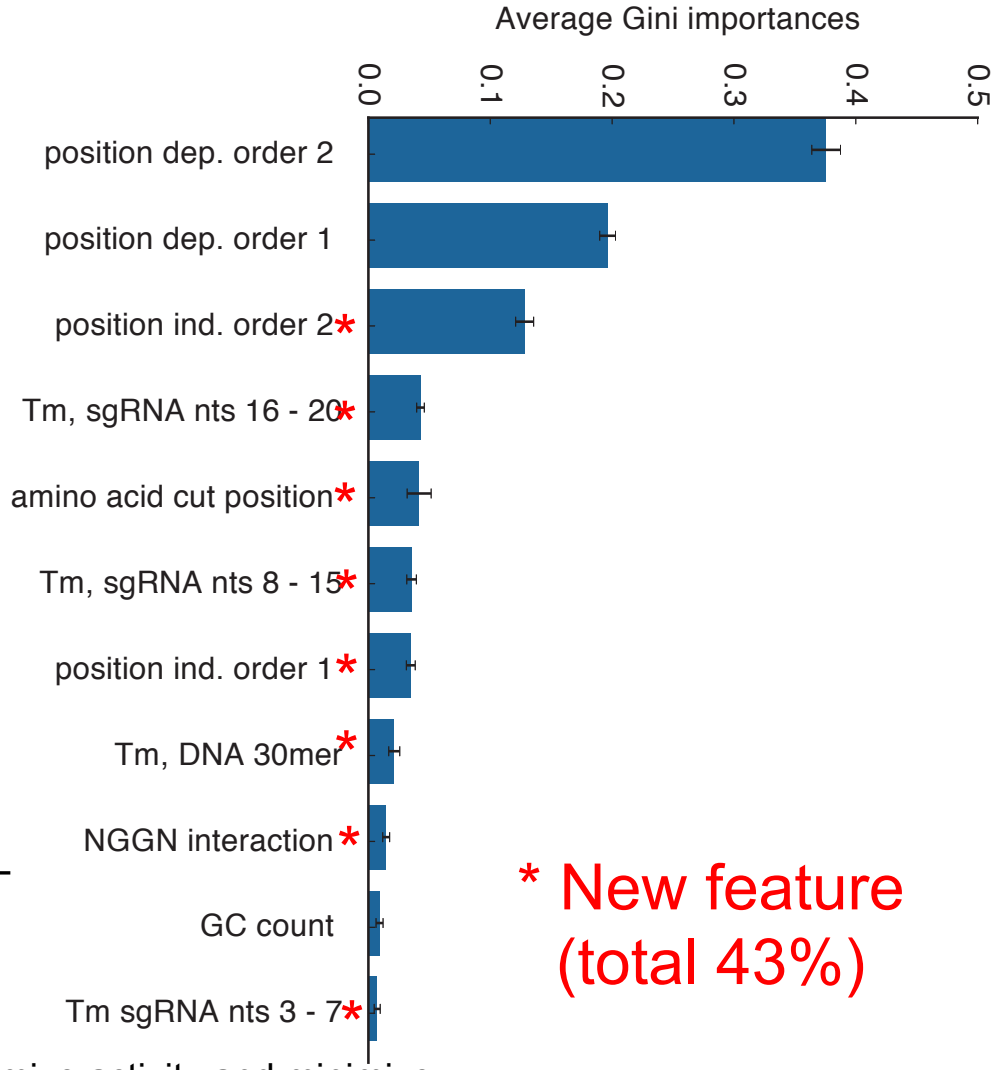



- For each gene, compare sequence content for 20% most-active sgRNAs to the remainder
- Linear weights of nucleotide features to build a classifier to predict sgRNA activity
- http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design

# From Rule Set 1 to Rule Set 2: More data, better computation

*with Jennifer Listgarten, Nicolo Fusi, Microsoft Research*

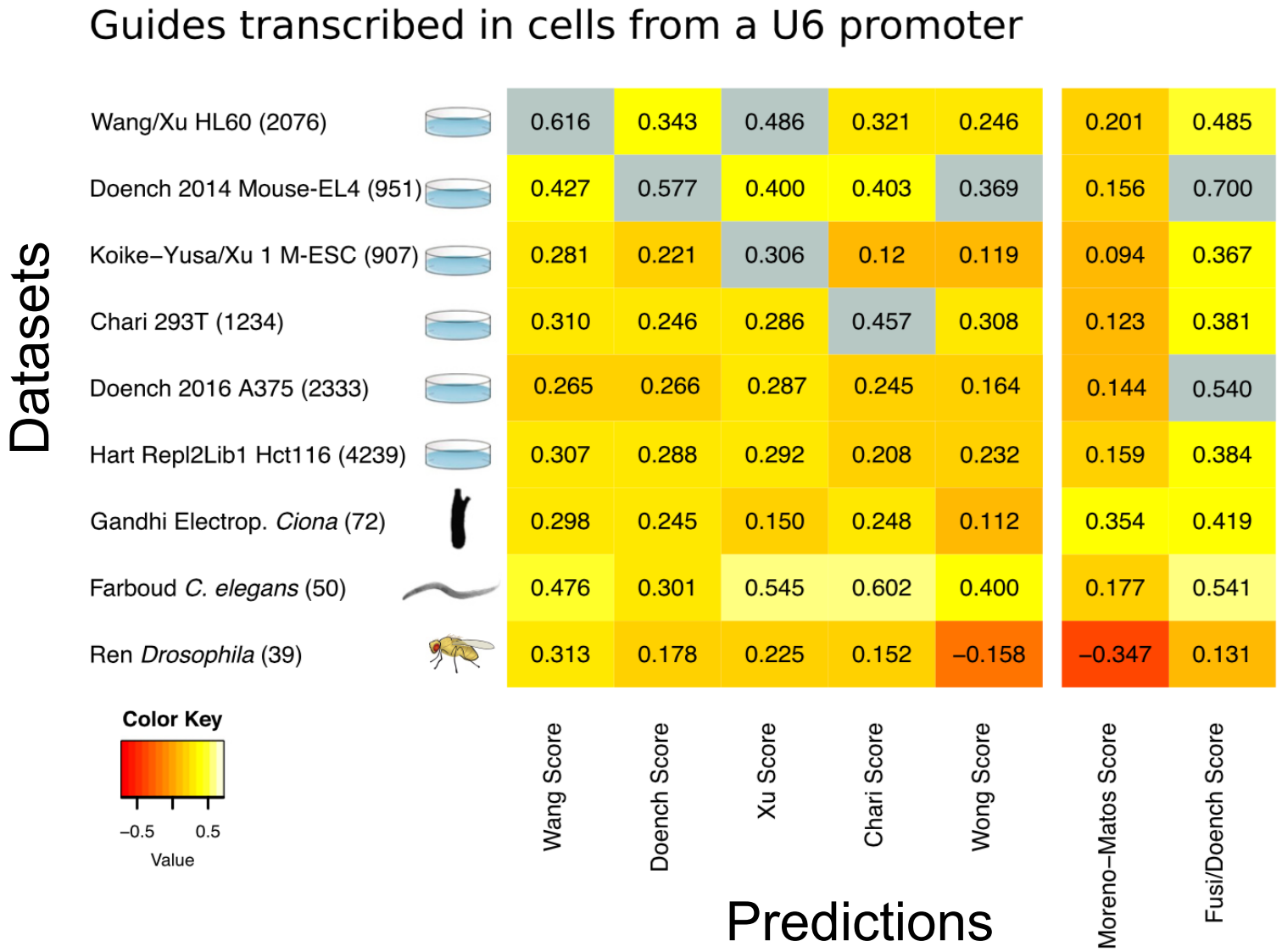



* New feature (total 43%)

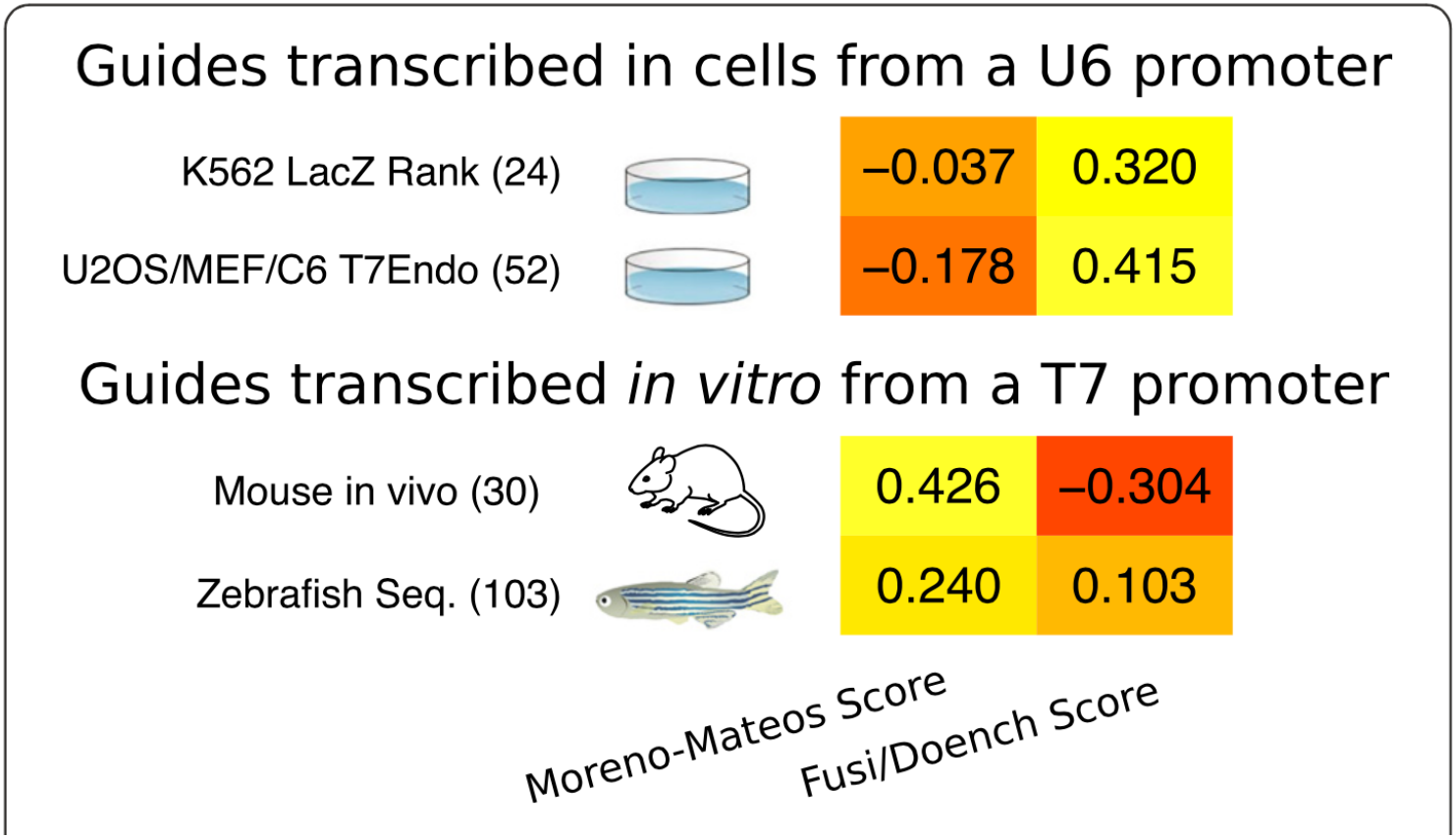
Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34,** 184–191 (2016).

# Unbiased comparison of prediction algorithms




Haeussler et al., *Genome Biology,* 2016

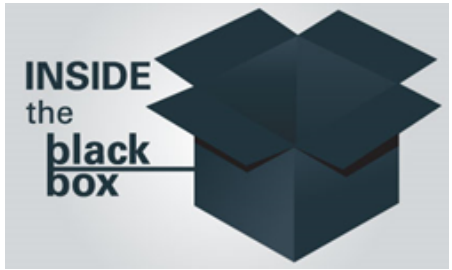
# How the sgRNA is made / delivered may affect prediction algorithm accuracy




Haeussler et al., *Genome Biology,* 2016

# A selection of design sites

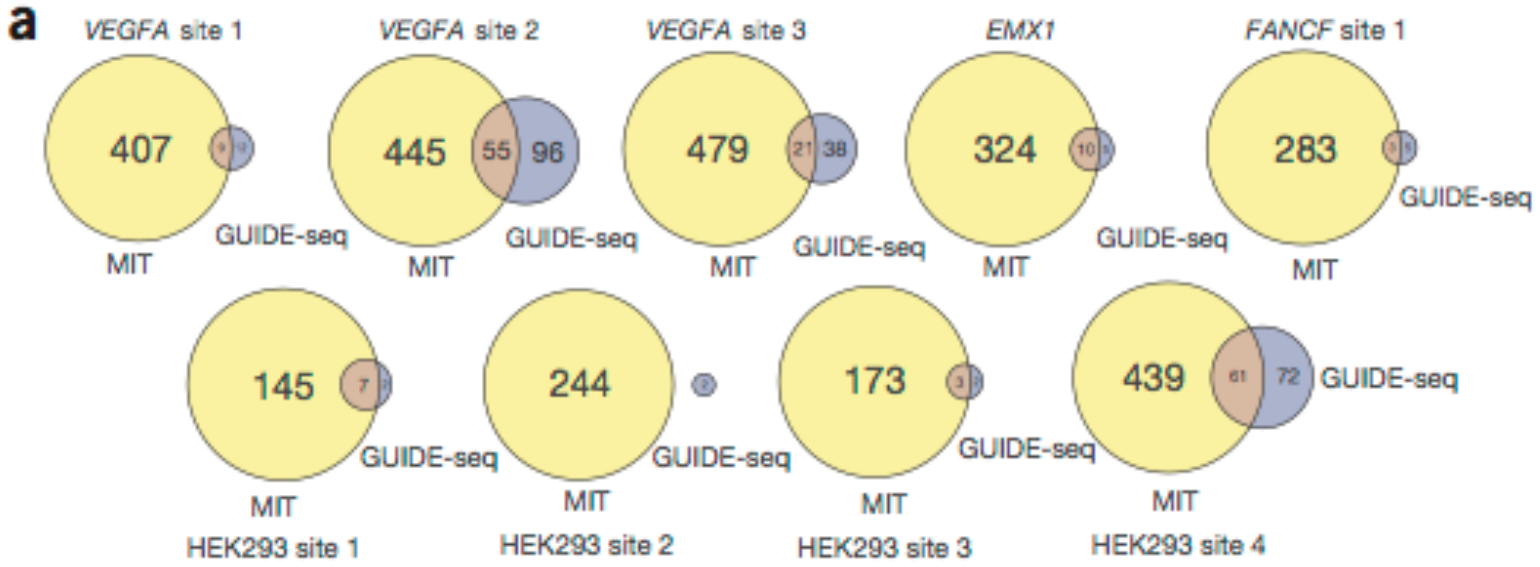
- Broad-GPP (for designing sgRNAs)
  - http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design
- CRISPOR (implements multiple scoring schemes)
  - http://crispor.tefor.net
- GenomeCRISPR (to search existing data)
  - http://genomecrispr.dkfz.de/#!/
- Companies
  - Desktop Genetics, Benchling, GE (Dharmacon), many others




*Don't use ANY tool unless you know what its limitations are and what purpose it was built for*

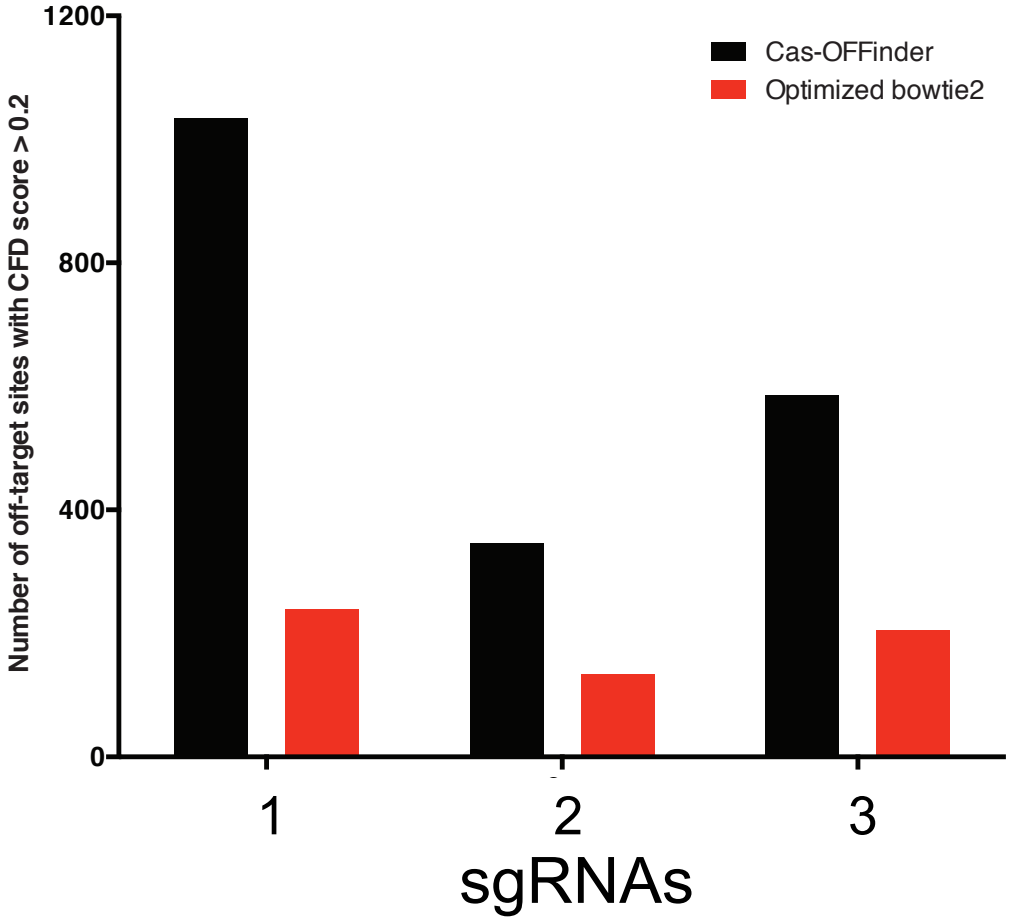
# GUIDE-seq: unbiased detection of off-target sites

"Having established the efficacy of GUIDE-seq, we next performed direct comparisons of our method with two popular computational programs for predicting off-target mutation sites: the MIT CRISPR Design Tool and the E-CRISP software. Both of these programs identify potential off-target sites based on 'rules' about mismatch number and position. In direct comparisons, we discovered that **neither program identified the vast majority of off-target sites** found by GUIDE-seq for the nine RGNs."




Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33,** 187–197 (2014).

# Often-used search algorithms miss potential off-target sites




Don't use design tools & websites that haven't been updated in several years, this field moves fast!

Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34,** 184–191 (2016).

# Genome-wide libraries

*Available via Addgene as plasmid AND lentivirus*

Human library
*Avana*


Mouse library
*Asiago*

From Brunello to Zinfindel – strategy of pooled oligo synthesis, pooled clone production, and pooled screening allows for *rapid*, *inexpensive* cycles of innovation


Human library
*Brunello*

Mouse library
*Brie*

# Comprehensive resource for pooled CRISPR libraries: Addgene

## Find Pooled Libraries of Interest

Click on different properties to create a custom filtered list of the CRISPR pooled libraries in Addgene's collection.

**Library Type**
- ○ Knockout
- ○ Activation
- ○ Inhibition

**Species**
- ○ Human
- ○ Mouse
- ○ Fly
- ○ *T. gondii*

**Library Size**
- ○ Genome-wide
- ○ Subpool

**Viral Prep Available**
- ○ Yes
- ○ No

Clear Filters

| Name | ID | Library Type | Species | PI | Lentiviral Generation | gRNAs per gene | Total gRNAs |
|---|---|---|---|---|---|---|---|
| Activity-optimized genome-wide library | 1000000067 | Knockout | Human | Sabatini and Lander | 3rd | 10 | 178,896 |
| Broad GPP genome-wide Brunello | 73179 (1 plasmid)<br>73178 (2 plasmid) | Knockout | Human | Doench and Root | 3rd | 4 | 76,441 |
| Broad GPP genome-wide Brie | 73632 (1 plasmid)<br>73633 (2 plasmid) | Knockout | Mouse | Doench and Root | 3rd | 4 | 78,637 |
| Broad GPP kinome Brunello | 75314, 75315 (1 plasmid)<br>75312, 75313 (2 plasmid) | Knockout | Human | Doench and Root | 3rd | 4 | 3,052 |
| Broad GPP kinome Brie | 75317 (1 plasmid)<br>75316 (2 plasmid) | Knockout | Mouse | Doench and Root | 3rd | 4 | 2,852 |
| CRISPRa | **Discontinued** | Activation | Human | Weissman | 3rd | 10 | 198,810 |

# Are we doing any better? sgRNA sets to assess library performance

*Article*

TRANSPARENT PROCESS

OPEN ACCESS

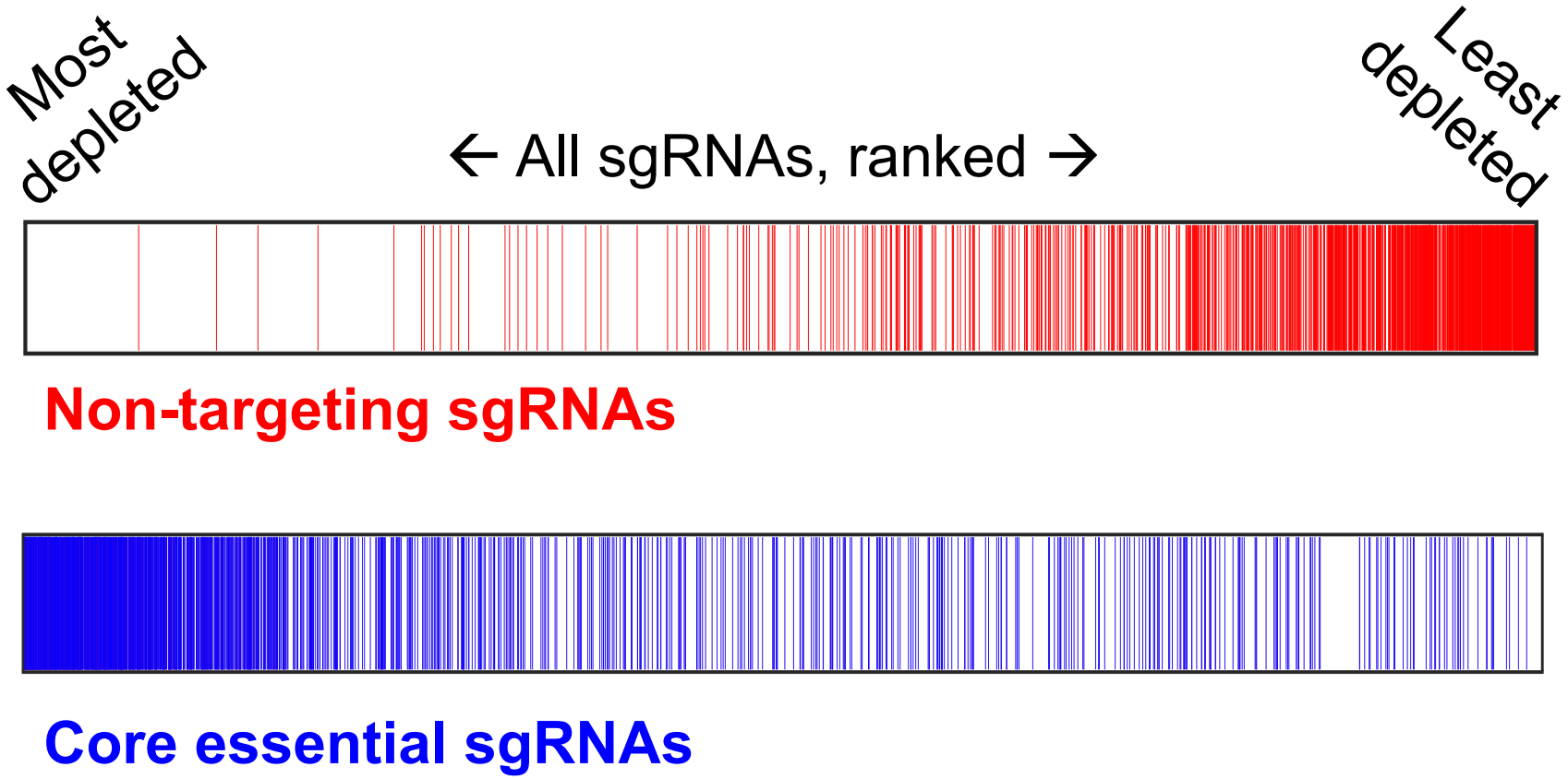
molecular systems biology

## Measuring error rates in genomic perturbation screens: gold standards for human functional genomics

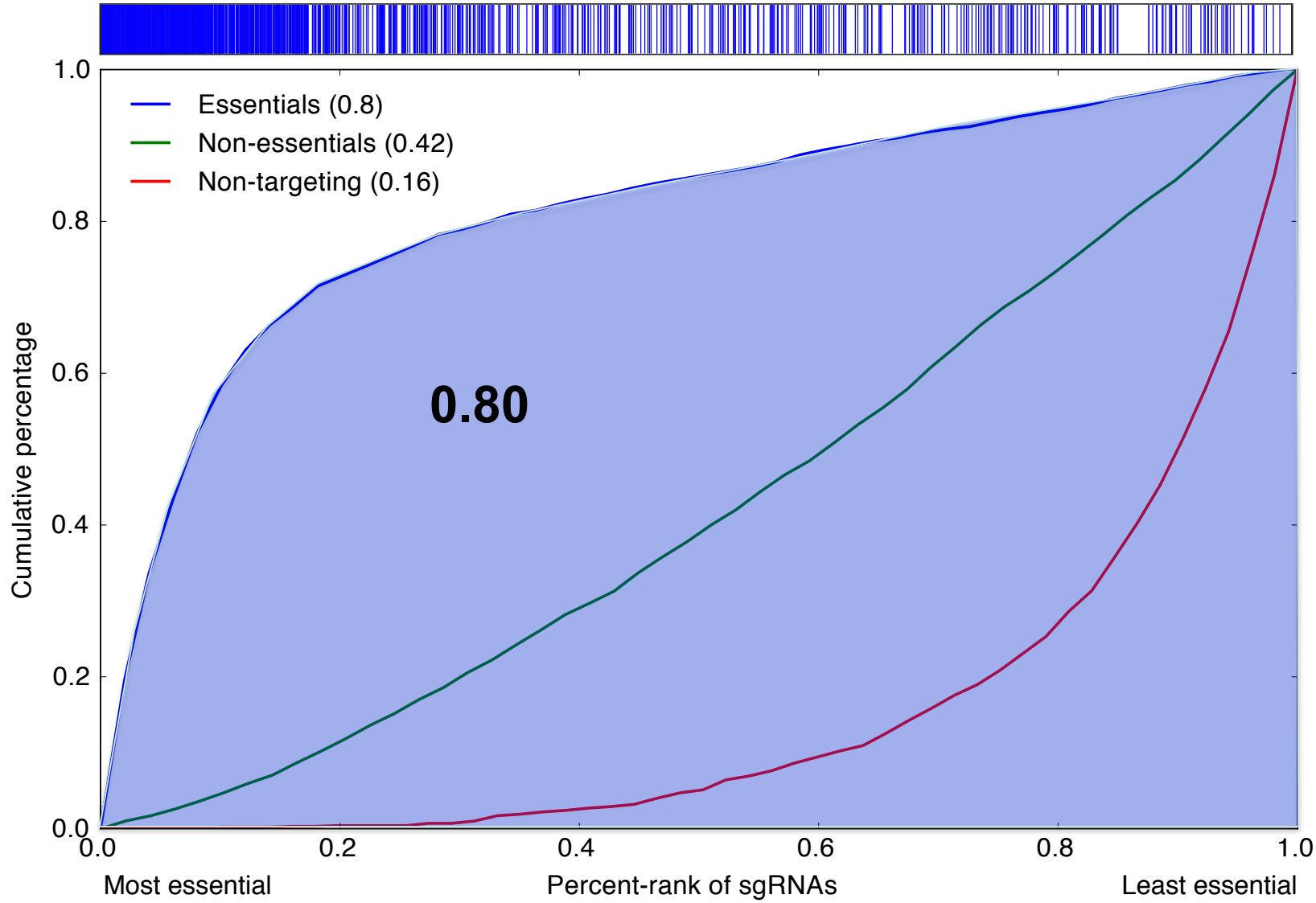
Traver Hart[1], Kevin R Brown[1], Fabrice Sircoulomb[2], Robert Rottapel[2,3,4] & Jason Moffat[1,5,*]

291 core essential genes
927 non-essential genes
1000 non-targeting control sgRNAs

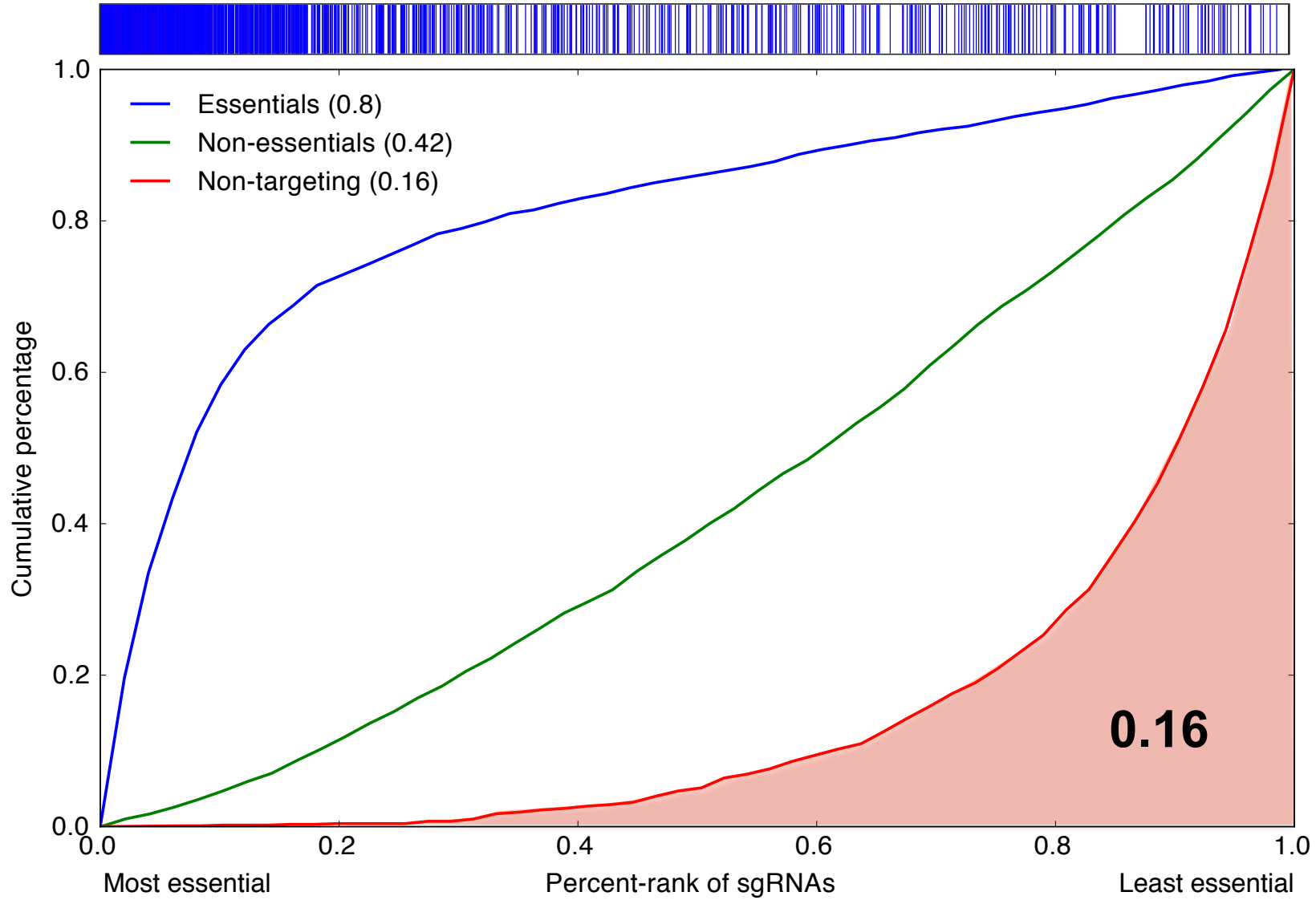
# Are we doing any better? sgRNA sets to assess library performance



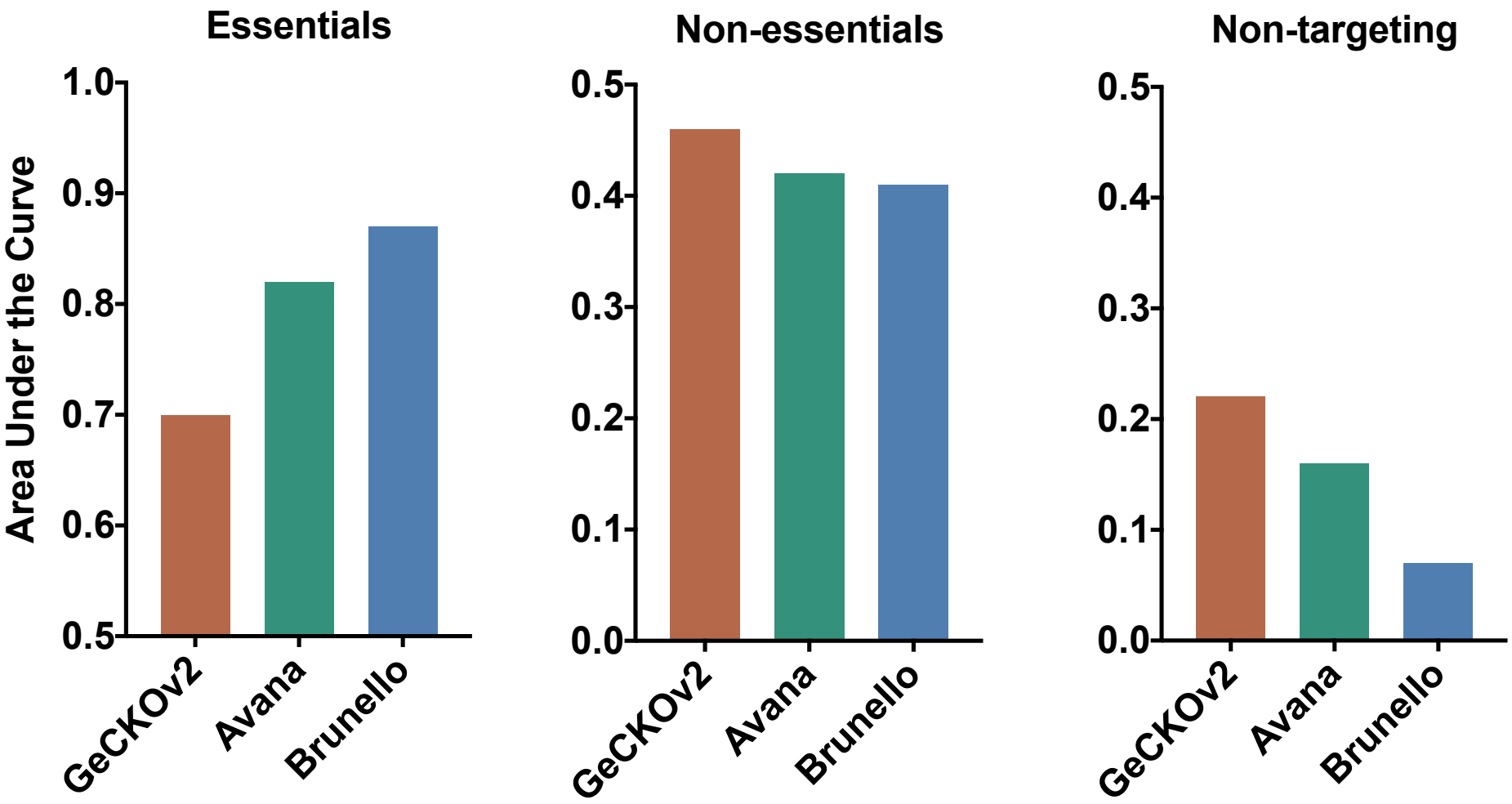
# Quantitate performance via Area Under Curve (AUC)



# Quantitate performance via Area Under Curve (AUC)



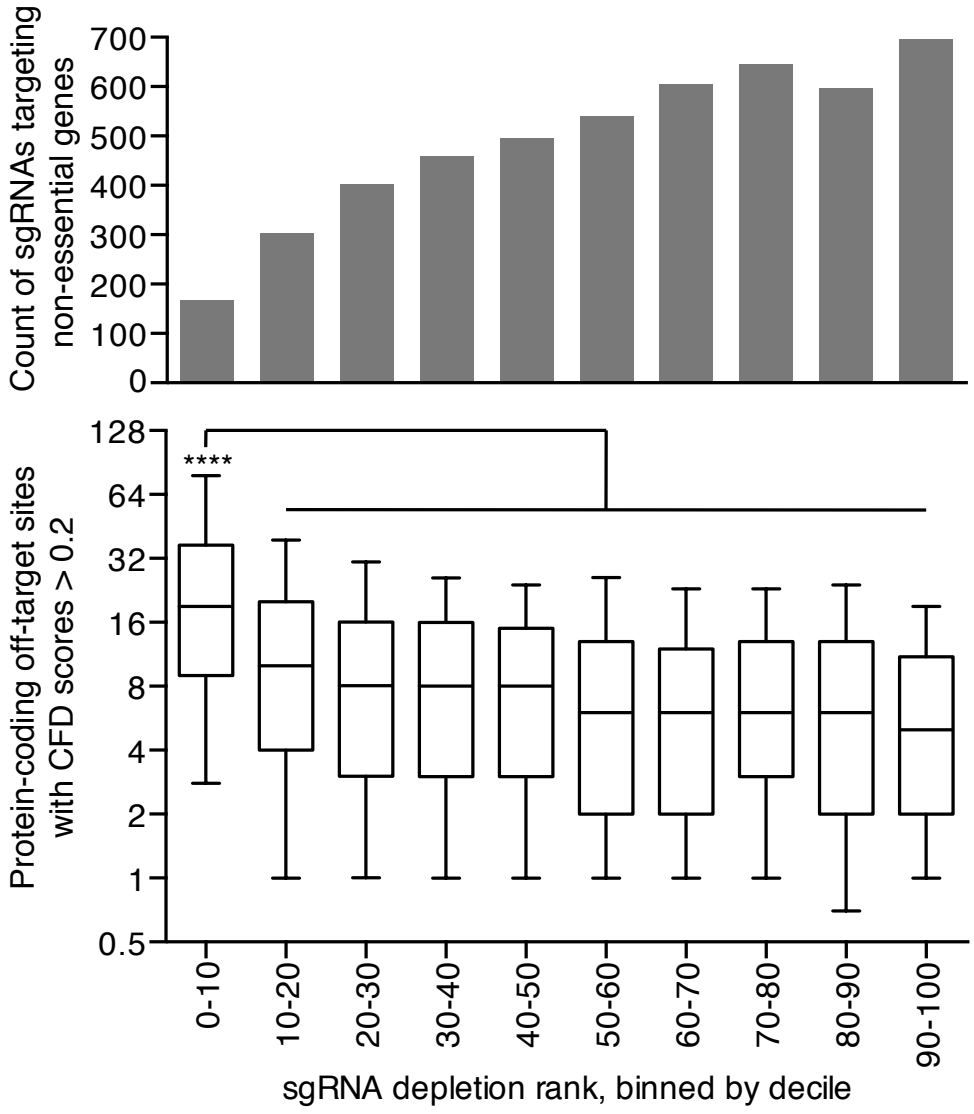
# Improvement with each library




Doench, unpublished

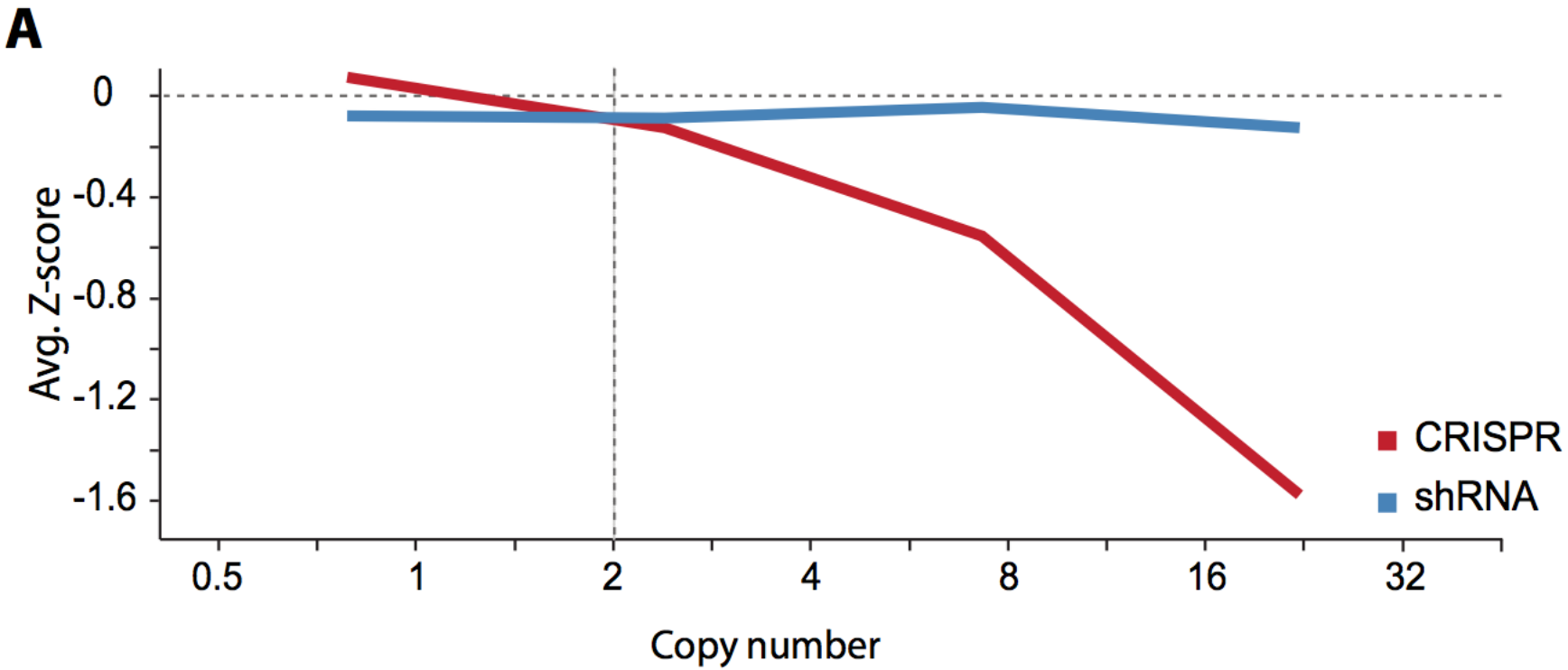
# Promiscuous sgRNAs kill cells

## *Analysis of sgRNAs targeting non-essential genes*



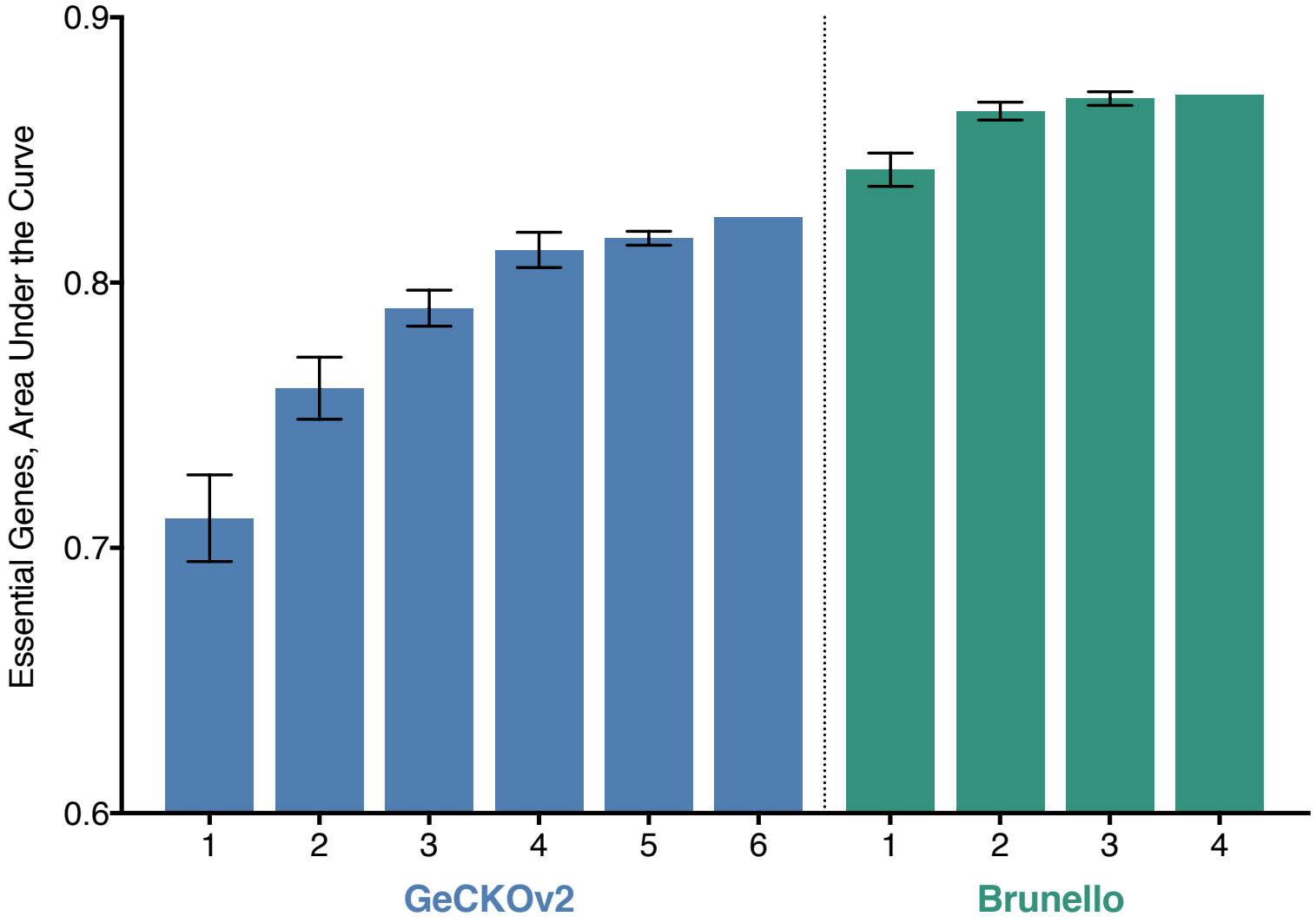

Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34,** 184–191 (2016).

# Promiscuous sgRNAs kill cells…
# … and copy number amplified regions




Munoz et al., *Cancer Discovery*, 2016
Aguirre et al., *Cancer Discovery,* 2016

# Subsampling to compare answers at the **gene** level: how many guides needed?




Doench, unpublished

# Smaller, more active libraries

- Why smaller libraries? No doubt that more sgRNAs gives more data, better statistical significance. But it comes at a cost:
  - More cells: puts some models out of reach (e.g. primary cells, *in vivo* applications, flow cytometry)
  - More dollars: more cell culture, more sequencing, etc.
- Primary screen: identify genes that *might* be hits by casting a wide net for follow-up (low false negative rate)
- Secondary screen: more sgRNAs per gene result in a low false positive rate moving forward

# What analysis *cannot* do for you

8. Matrix of sgRNA Abundance

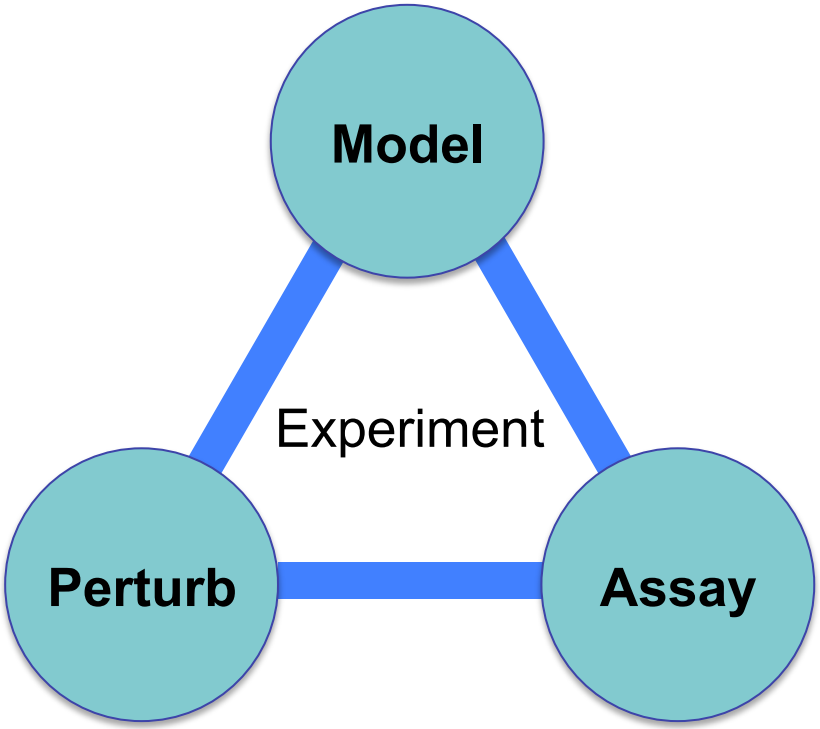
| sgRNA | Init. | Pos. | Neg. |
|---|---|---|---|
| | 16 | 0 | 21 |
| | 19 | 100 | 19 |
| | 17 | 0 | 0 |
| | 18 | 0 | 26 |
| | 14 | 0 | 18 |
| | 16 | 0 | 16 |

- Take technically poor data and make it technically sound
  - If your replicates don't line up well, "analysis" won't change that
- Tell you what assays to do next – you should have thought about this *before* the screen
  - Only rarely does the identity of genes in a hit list lead to an obvious hypothesis
- The #1 rule of screens with CRISPR: ***require multiple perturbations of independent sequence to give the same phenotype across all assays***

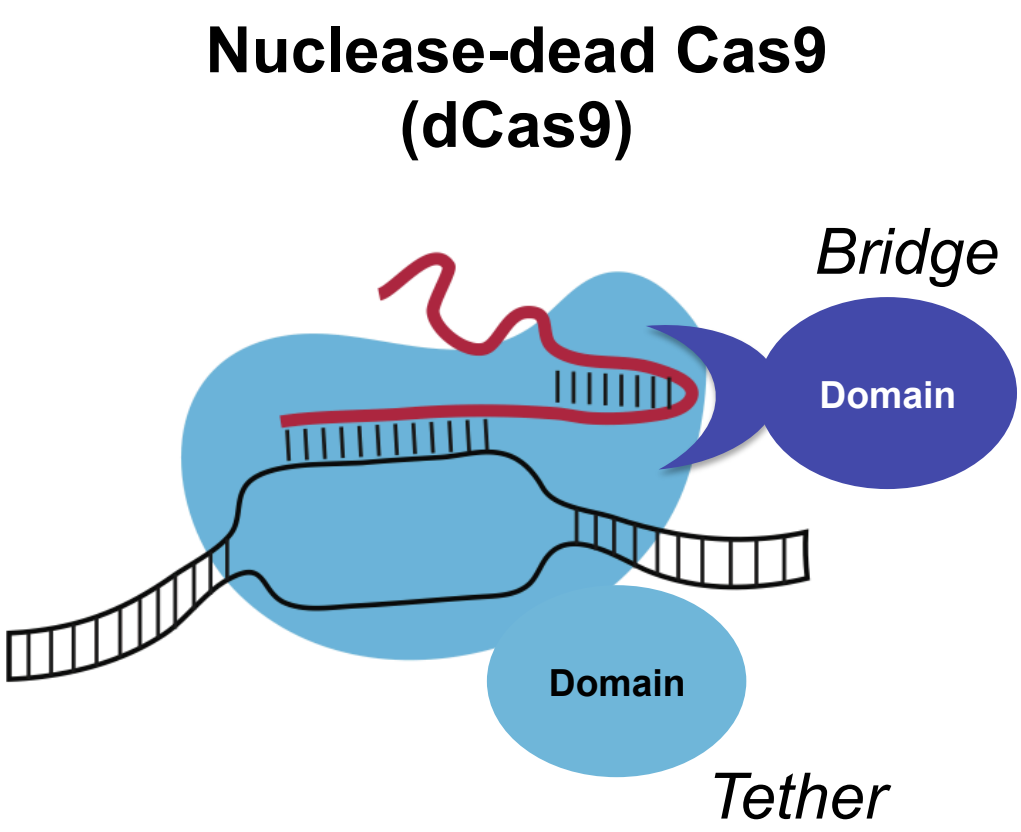
# Functional Genomics

***Pooled screens are great, you should do one!***

- CRISPR (knockout)
- ORF (overexpression)
- RNAi (knockdown)
- CRISPRa (overexpression)
- CRISPRi (knockdown)



# Diversity of Cas9 activities




**Programmable DNA binding protein:**

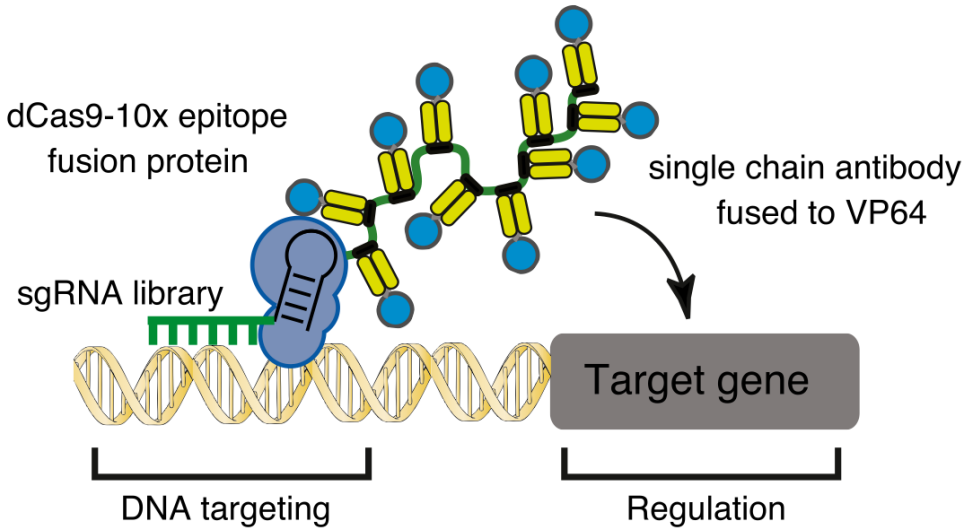
1. Tether protein domains to Cas9 itself
2. Bridge via an RNA stem loop engineered into tracrRNA along with fusion protein

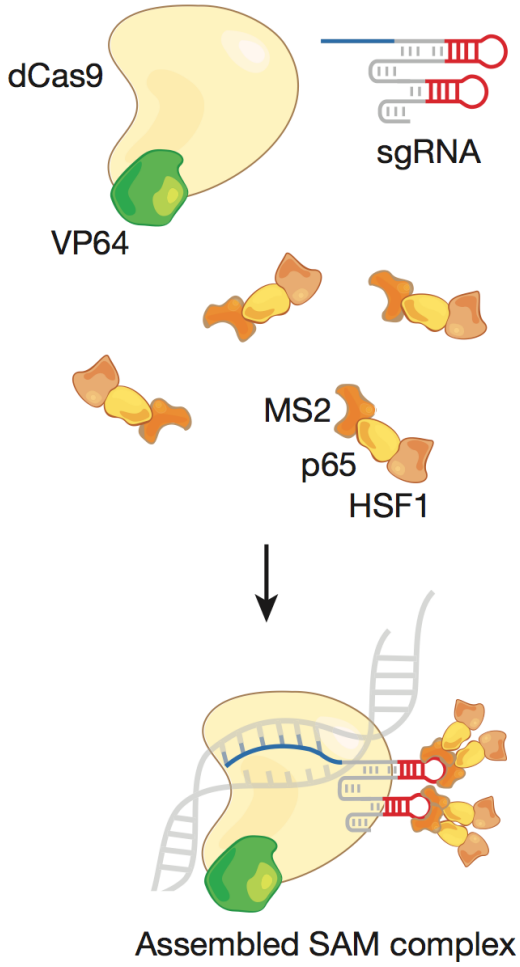
Activities

- Transcriptional activation (CRISPRa)
- Transcriptional repression (CRISPRi)
- Visualization
- Mutagenesis

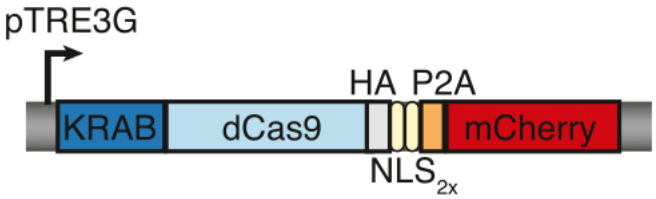
# dCas9 for transcriptional modulation

CRISPRa screening



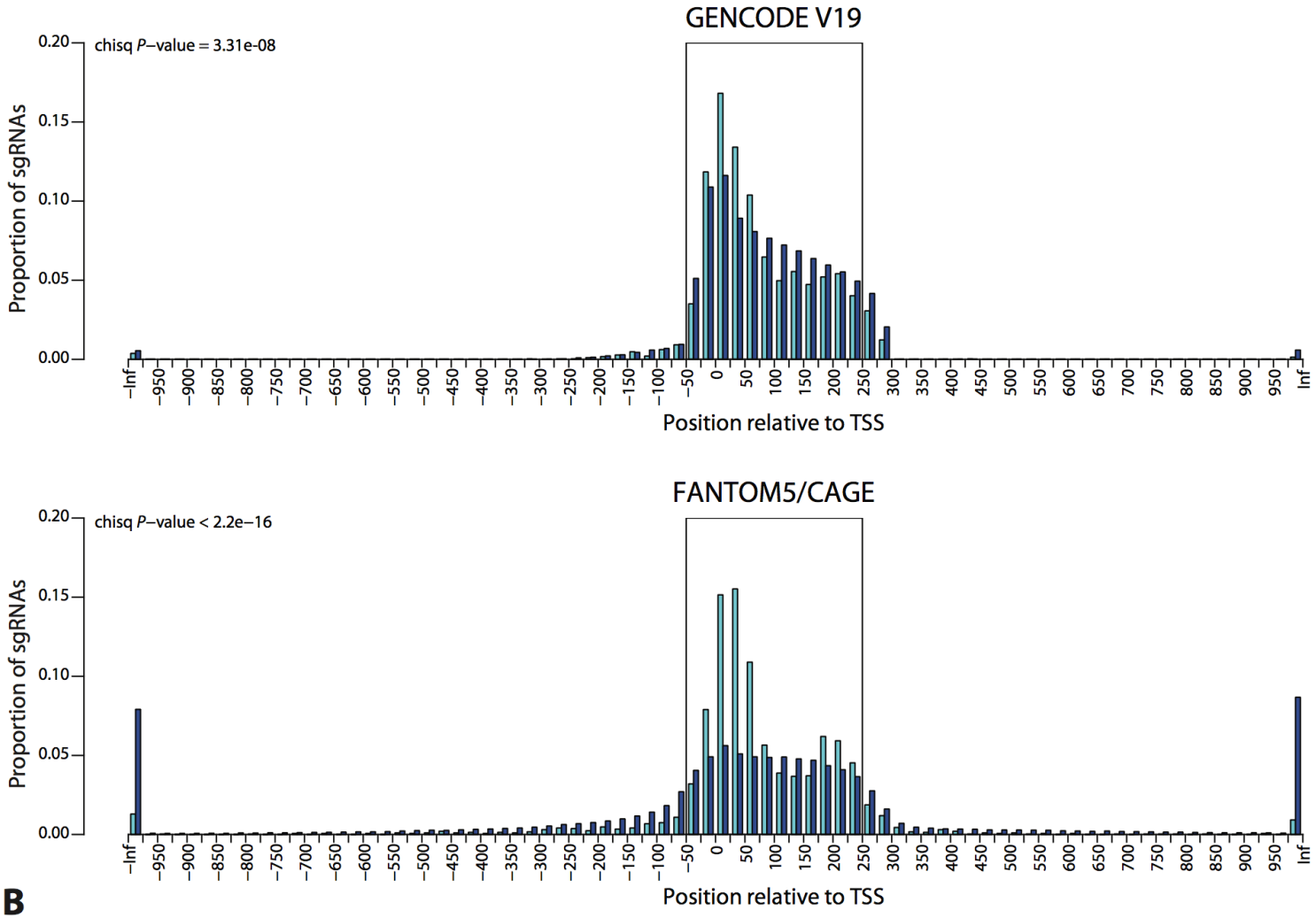

CRISPRi




Gilbert et al., Cell, 2014

Konermann et al., Nature, 2014

# CRISPRi (and CRIPSRa) designs require good transcription start site (TSS) annotation

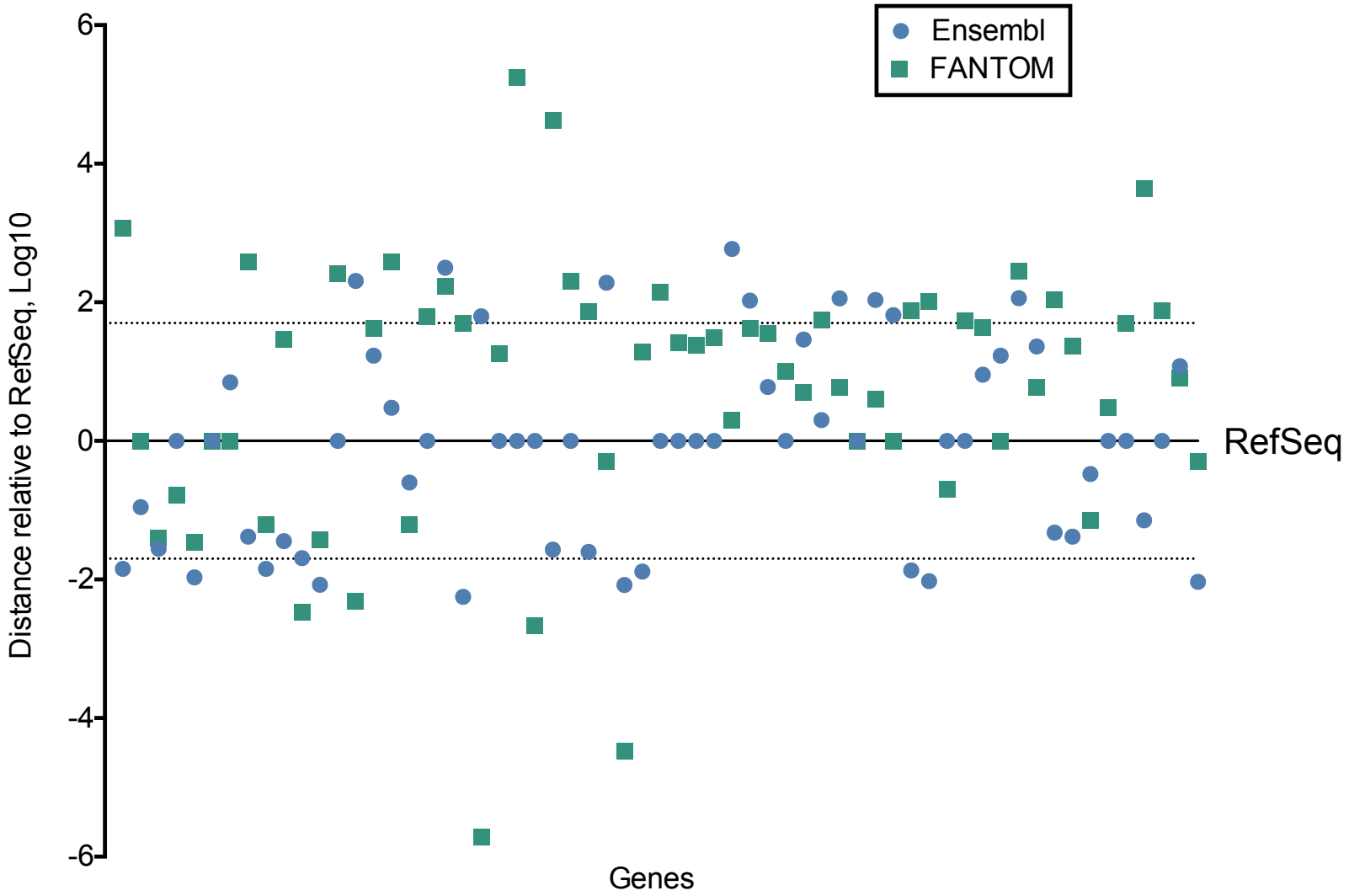



| | inside (-50...+250) |
|---|---|
| "non-hit" | 89% |
| "hit" | 92% |

| | inside (-50...+250) |
|---|---|
| "non-hit" | 56% |
| "hit" | 86% |

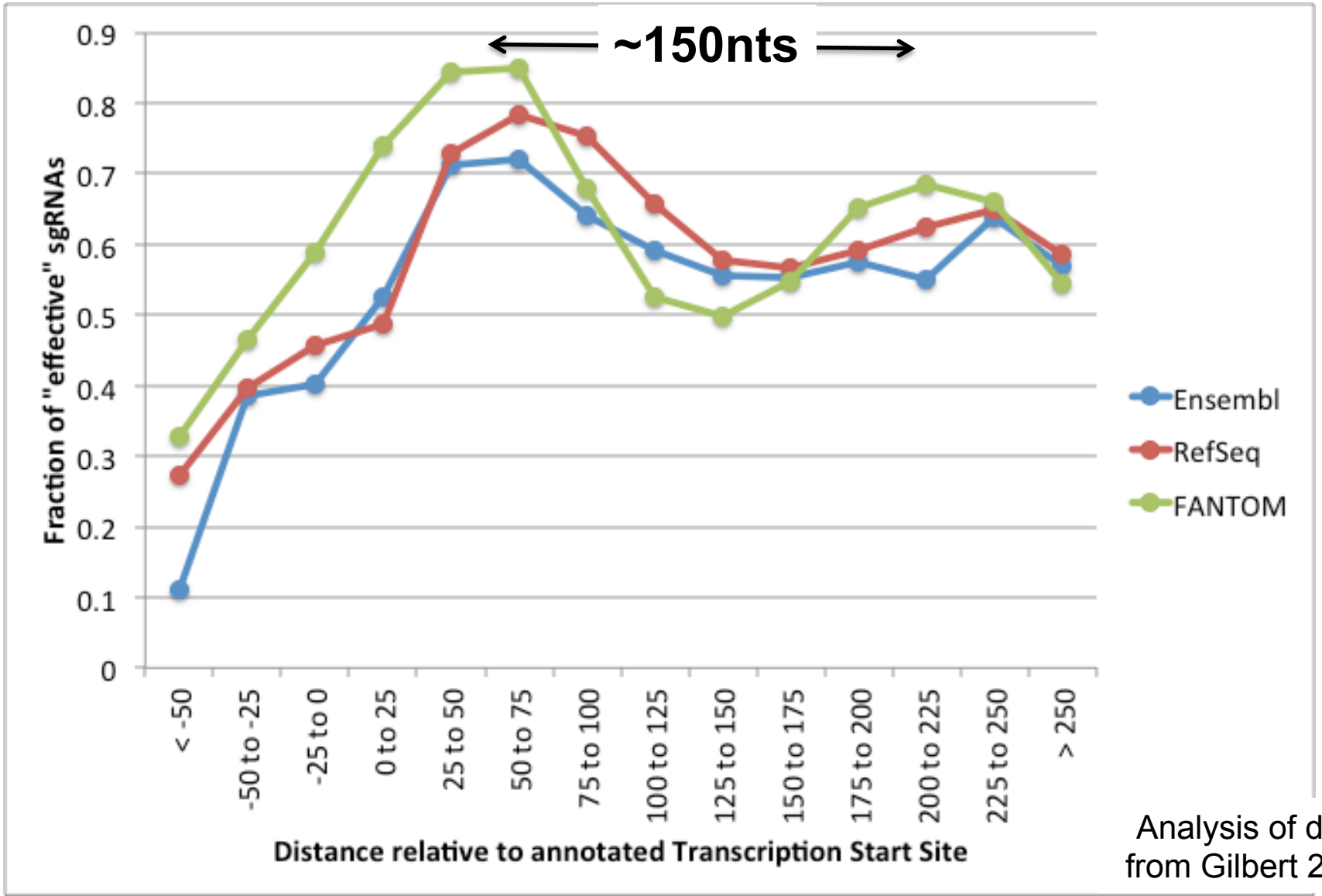
Radzisheuskaya, A., Shlyueva, D., Müller, I. & Helin, K. Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. *Nucleic Acids Research* **44,** e141–e141 (2016).

# Transcription Start Site (TSS) annotations differ across databases




Doench, unpublished

# CRISPRi location preference




Analysis of data from Gilbert 2014

# CRISPR: what it is and what it isn't

- For genetic screens (i.e. functional genomics) CRISPR is a game-changer.
  - Much greater on-target efficacy
  - Many fewer off-target effects
- For gene *editing,* CRISPR makes it possible, but that is not the same thing as going from hard to easy.
  - Homology-directed repair low efficiency
  - Delivery challenges
  - Single cell cloning
- Like any tool or technology, need to understand its strengths and weaknesses when using it.

# The Hype

"the capability to quickly and efficiently alter any gene sequence [will] have profound impacts on biological research and to yield potential therapeutic strategies for genetic diseases." – *Nat. Rev.*

**2013**
**TALENs**

**2003**
**RNAi**

"within a few years [this technology] should yield a rough idea of what each of our genes does." – *Fortune*

"a battle of biblical proportions over the patents for a crucial enzyme in molecular biology." – *The Scientist*

**1996**
**Taq polymerase**

# *The Economist*, 2015

Gartner, an American consultancy, has a simple yet elegant way of describing the life of a promising new technology. First, it is talked up to a **peak of inflated expectations**. Then it falls into a **trough of disillusionment**. After that, if it survives, it begins climbing the **slope of enlightenment**. Finally, it reaches the **plateau of productivity**. In the world of biotechnology CRISPR/Cas9 is still ascending towards peak expectations.