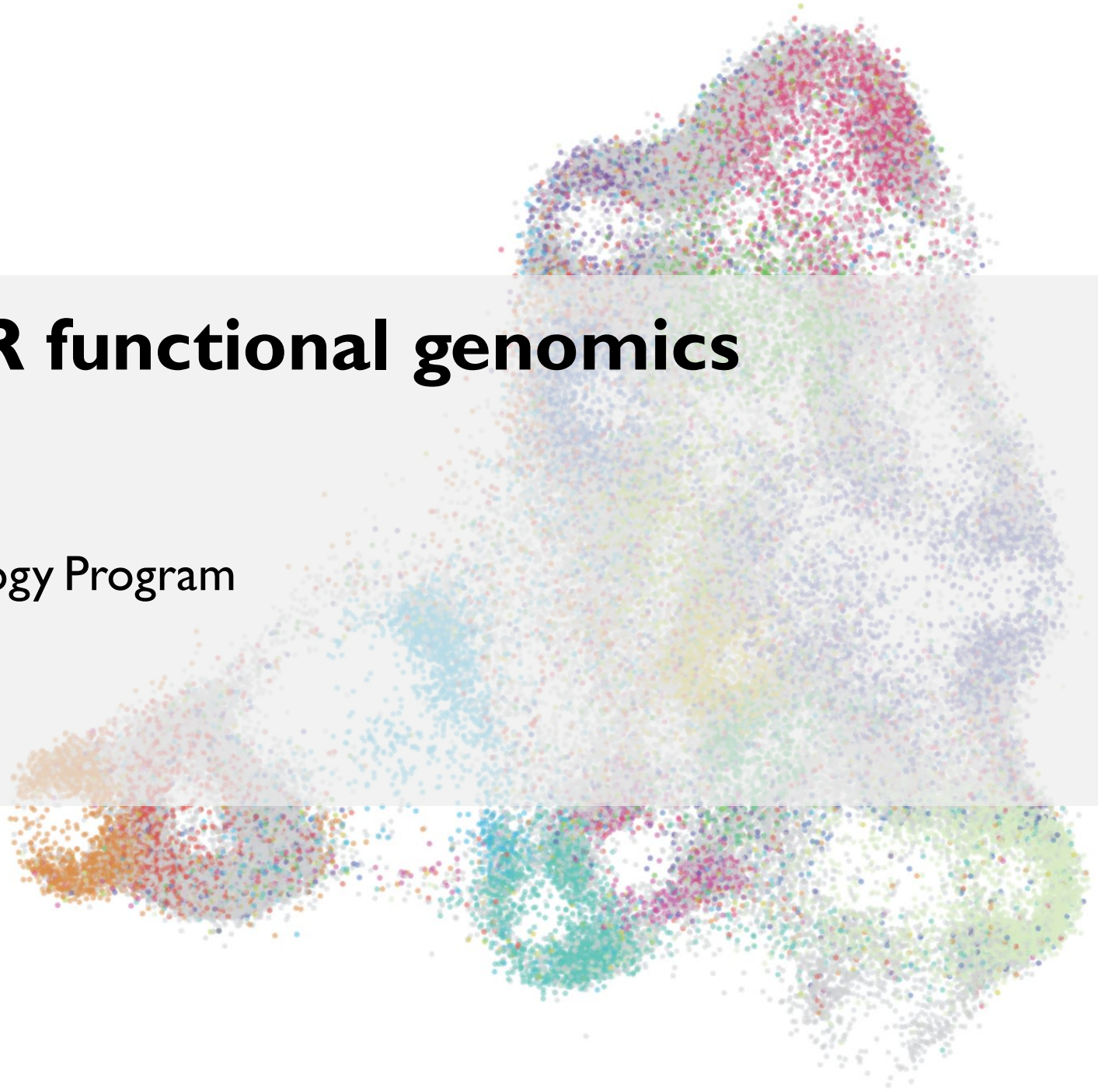
# Frontiers in CRISPR functional genomics

Tom Norman

Computational & Systems Biology Program
Sloan Kettering Institute

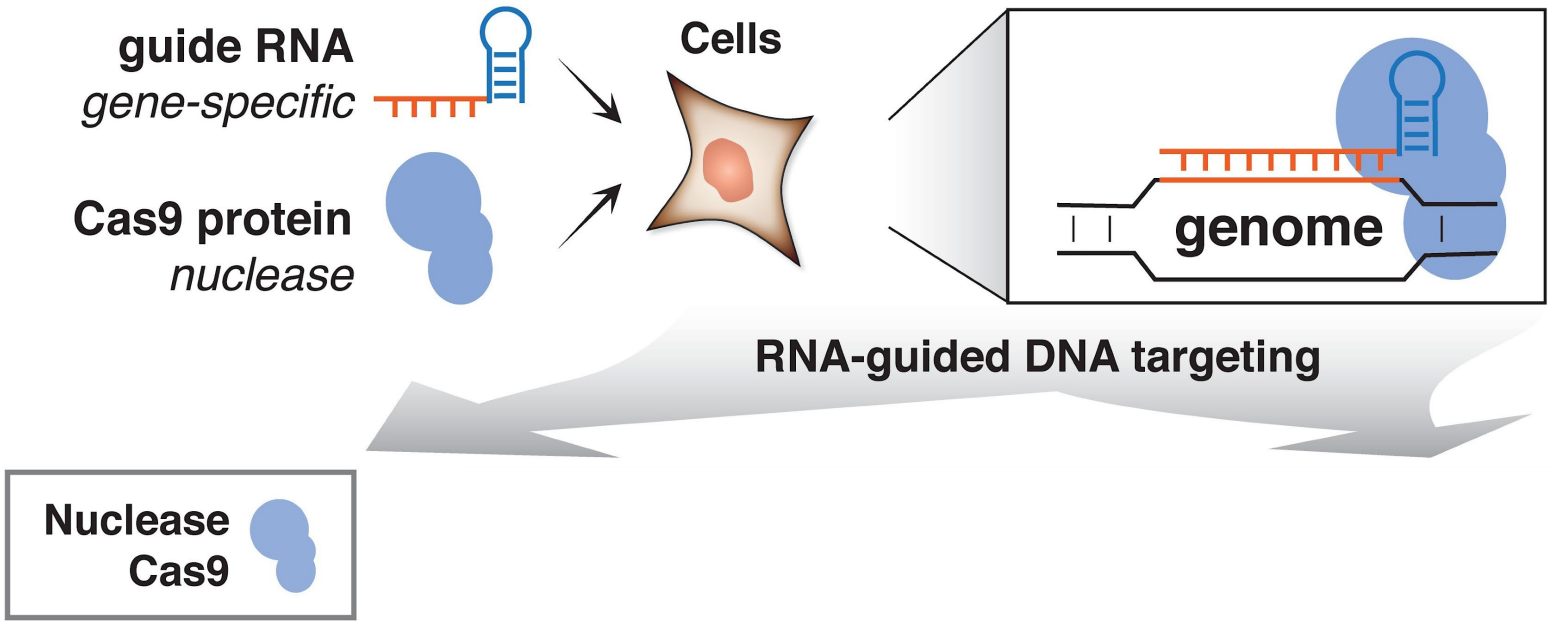
normantm@mskcc.org



# Outline

1. Functional genomics background
   - Current CRISPR tools
   - The four elements of a CRISPR screen
2. High-content readouts
   - The DepMap
   - Development of Perturb-seq
   - Using genome-scale Perturb-seq to assign gene function
   - Using imaging CRISPR screens to assign gene function
   - *In vivo* screens and adding lineage information
3. High complexity perturbations
   - Genetic interaction mapping
   - Highly multiplexed studies of enhancers

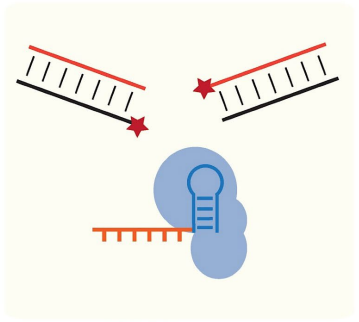
# Cas enzymes are programmable genome perturbation tools







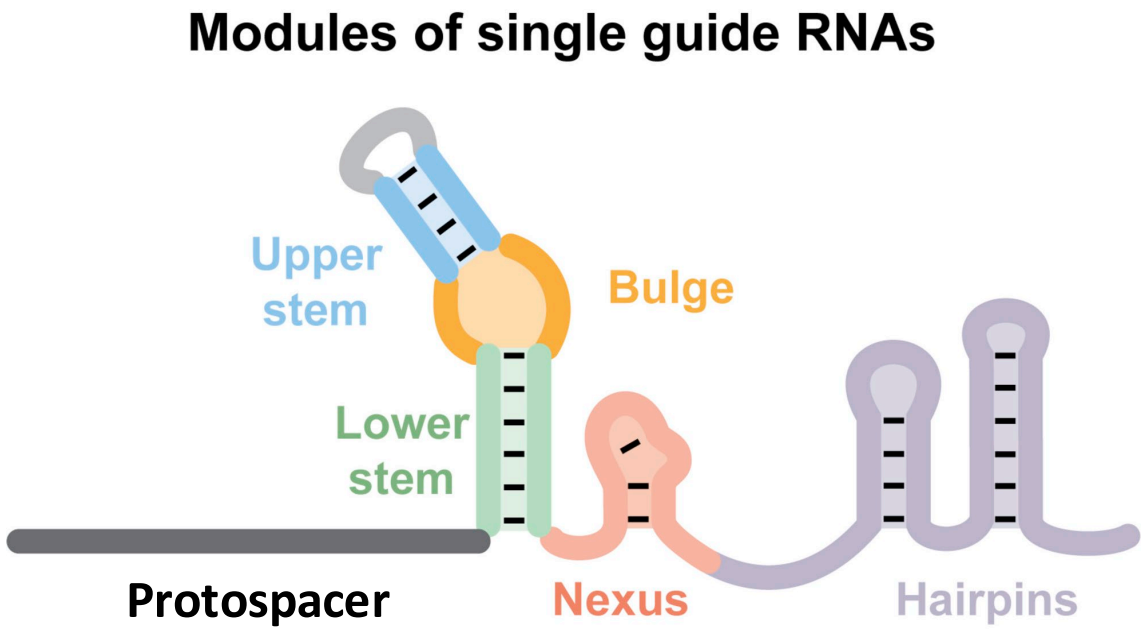
**Genome editing (CRISPRn)**



***DNA cutter***

# Structure of an sgRNA

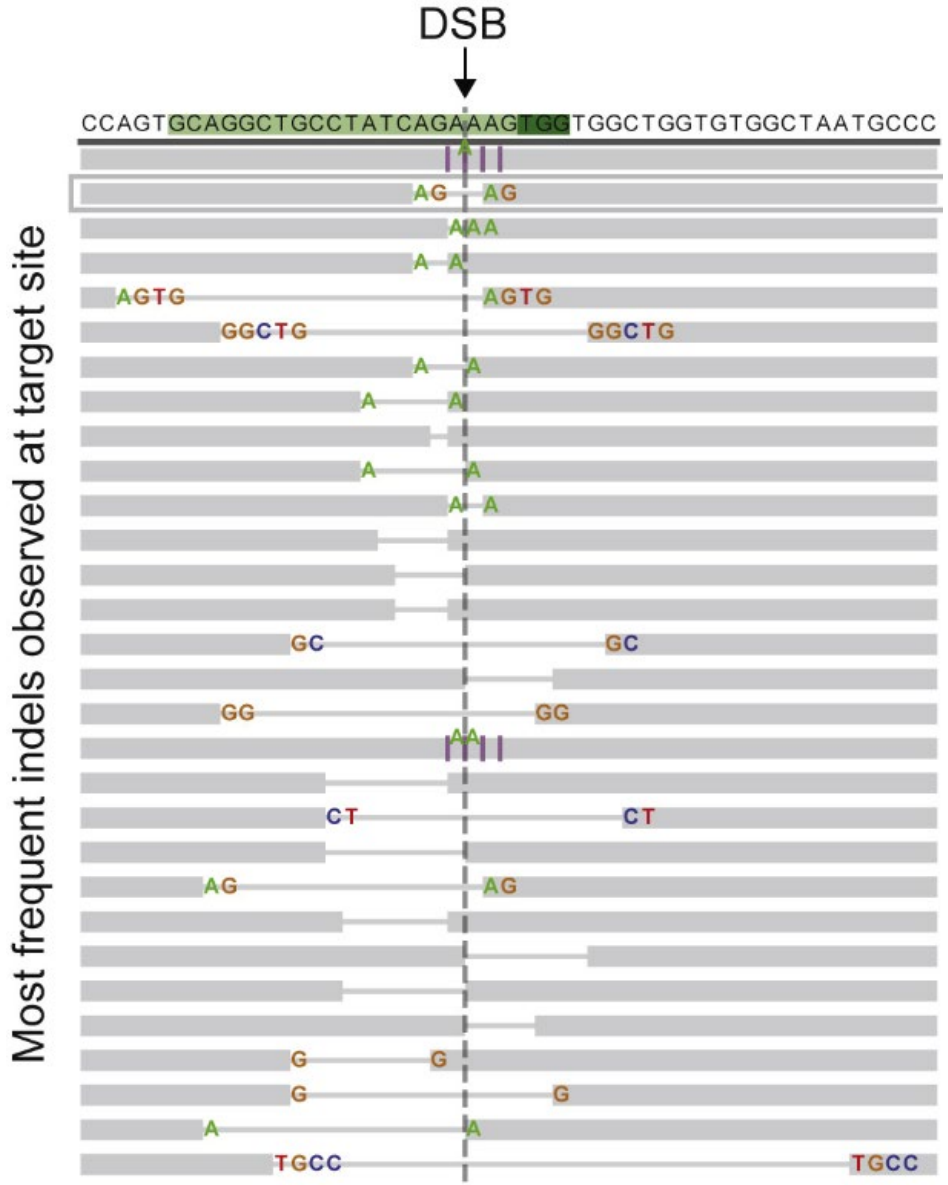
- sgRNAs are highly structured short RNAs expressed from Pol III promoters
- Target specificity set by the **protospacer** sequence, which is ~20 nt long
- The stem loop is responsible for binding to Cas9
- The first hairpin tolerates large insertions and is frequently modified:
  - in the SAM CRISPRa system, insert binding sites for phage RNA binding proteins
  - in direct capture Perturb-seq, insert a primer binding site to enable custom library prep




https://teichlab.github.io/scg_lib_structs/methods_html/10xChromium3fb.html

# CRISPR cutting

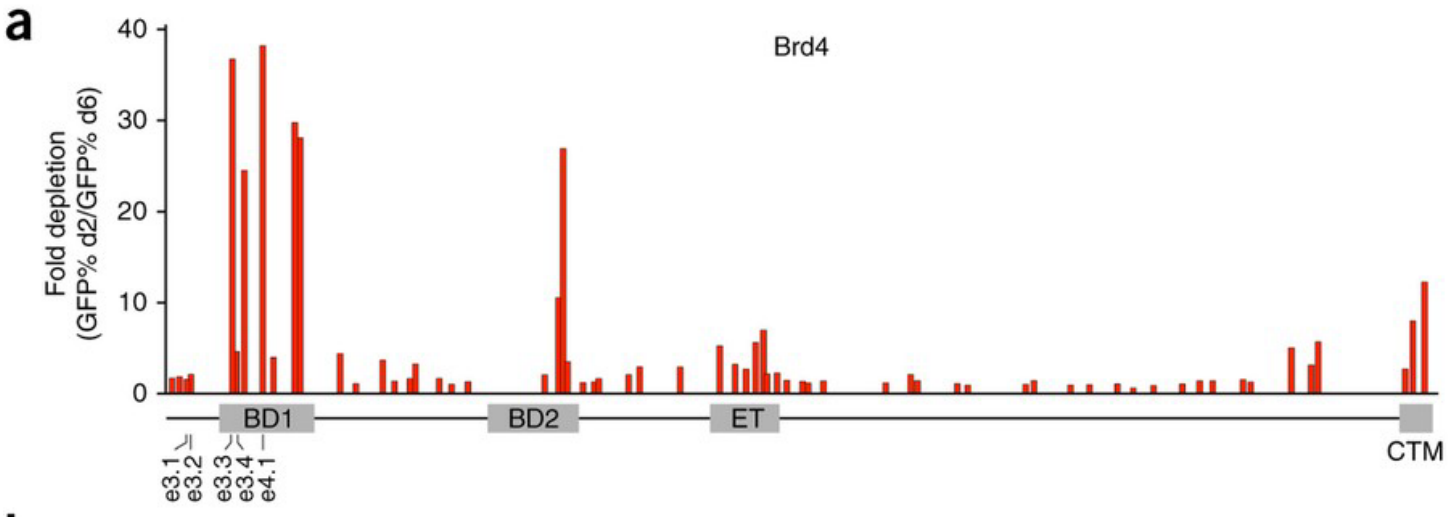
- Typically introduce a single double strand break away from the 3’ end of the targeted gene
- How this is resolved depends on relative activity of DNA repair pathways (NHEJ vs. MMEJ) in cell type, plus sequence context
- Most common indels are single base pair insertion (10-25%) and 1-2 bp deletions (20-25%), then a tail of increasingly diverse edits up to ~30 bp
- Some cell types seem to behave quite differently. E.g. ES cells generate large rearrangements quite frequently.
- **Not all breaks will result in frame shifts—i.e. not all edited cells will have loss-of-function of the targeted gene! (Some cells may also have heterozygous LOF.)**




https://doi.org/10.1016/j.cell.2021.10.002

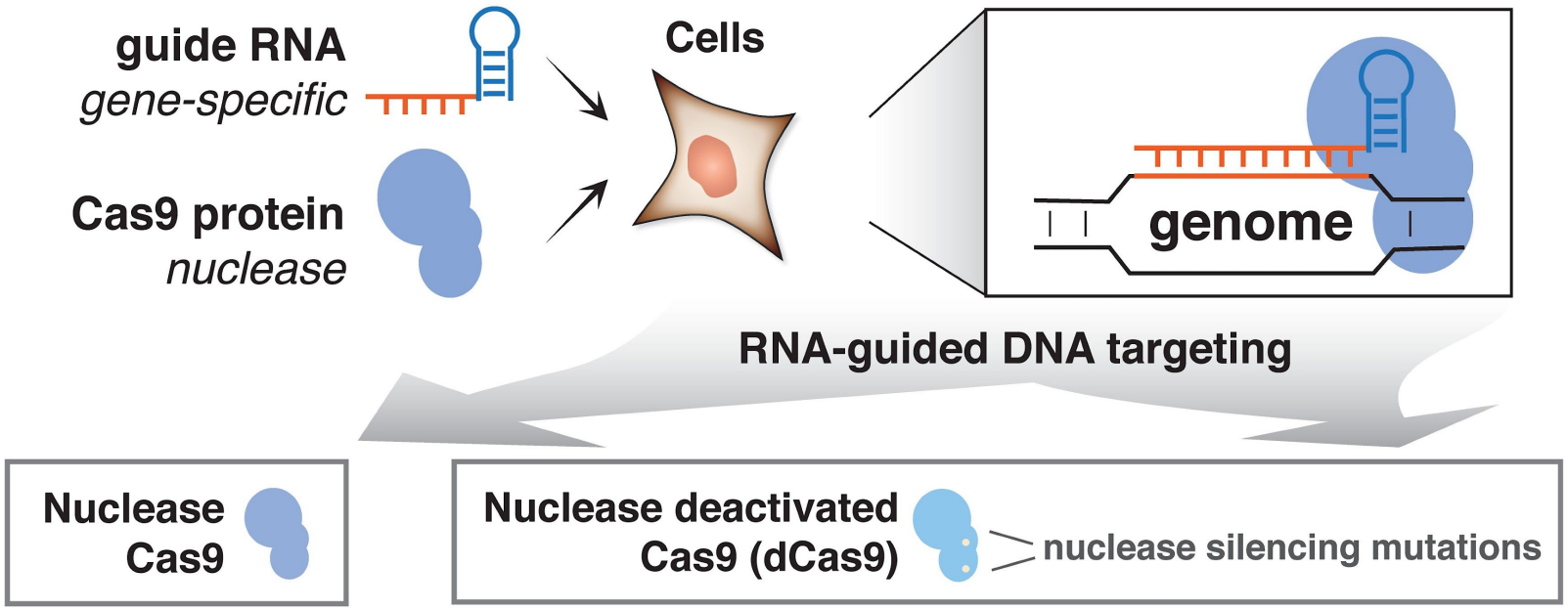
# CRISPR cutting

- One approach to improve activity is to target known functional domains or conserved regions. Even in-frame indels may then result in loss-of-function.
- Also useful in the opposite direction for mapping functionally important regions, which may not be well-characterized for many genes
- Note the variability in efficacy here for 64 different guides targeting *Brd4*

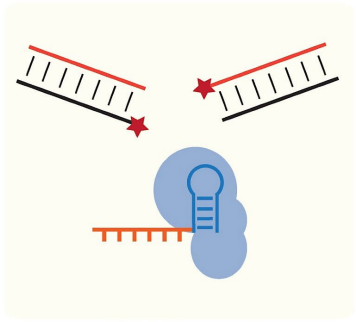



https://doi.org/10.1038/nbt.3235

# Cas enzymes are programmable genome perturbation tools

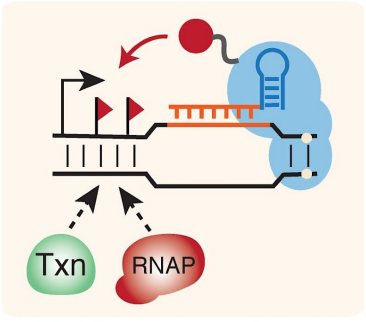



**Genome editing (CRISPRn)**



***DNA cutter***

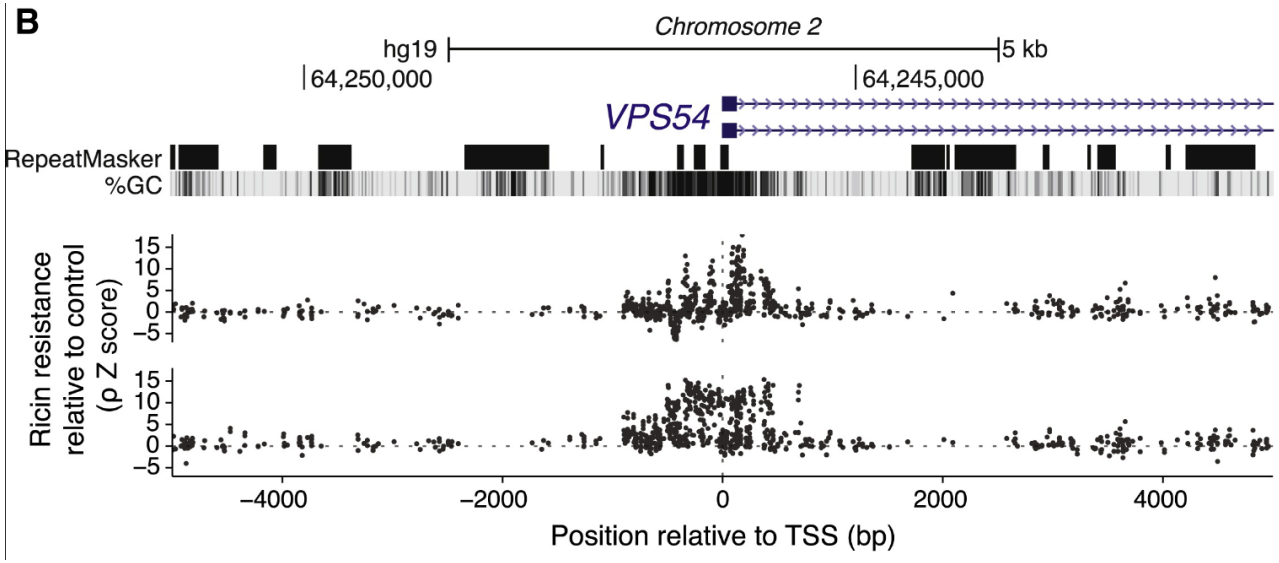
**Epigenome editing (CRISPRi/a/on/off)**




***Deliver regulators***

# Gene knockdown using CRISPR inhibition (CRISPRi)

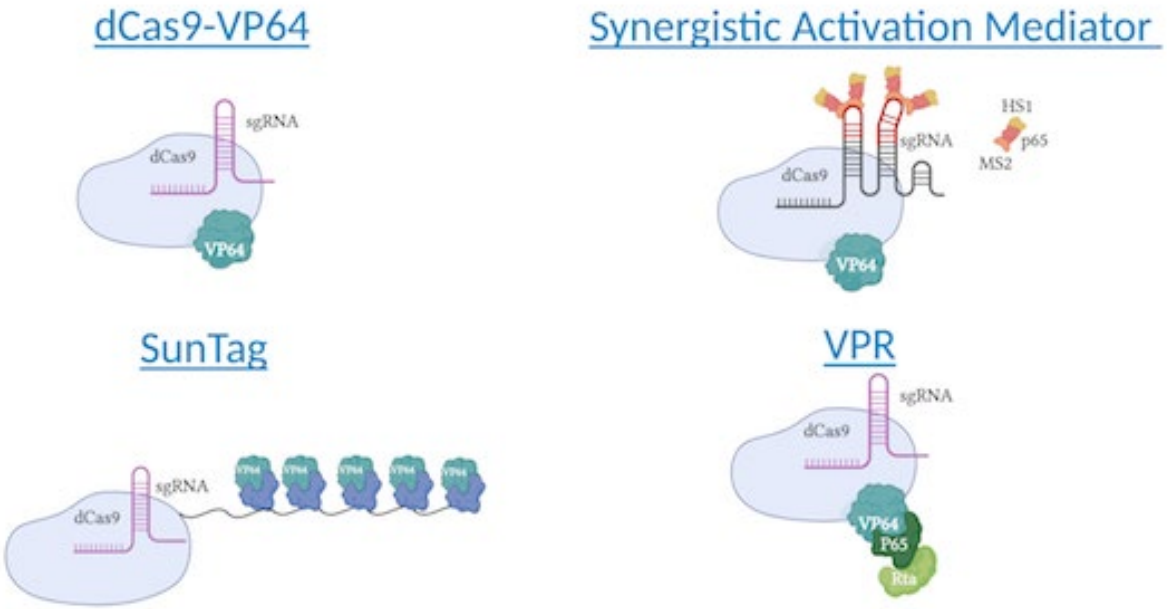
- Usually realized using a fusion of dCas9 with a KRAB domain from a zinc finger transcription factor e.g. ZIM3-dCas9
- Most effective when targeted just after the transcription start site of a gene
- Repression occurs in two ways:
  - Physical occlusion of RNA polymerase binding/elongation
  - KRAB domain triggers histone 3 lysine 9 trimethylation (H3K9me3)
- May act “digitally” – silencing is sticky once installed




https://doi.org/10.1016/j.cell.2014.09.029

# Gene activation using CRISPR activation (CRISPRa)

- Less understood and still a work-in-progress
- Overall goal is to associate various activating domains (usually virus-derived) with dCas9, but many strategies for tethering:
  - Direct fusions e.g. dCas9-VPR
  - SAM system: use phage RNA-binding proteins to recruit domains to sgRNA
  - Nanobodies to recruit multiple copies of activating domains to arrays of binding sites
- Activating genes is hard: still likely many false negatives. Works better at manipulating expression of genes that are already on rather than turning on dormant genes.




https://blog.addgene.org/crispr-activators-dcas9-vp64-sam-suntag-vpr

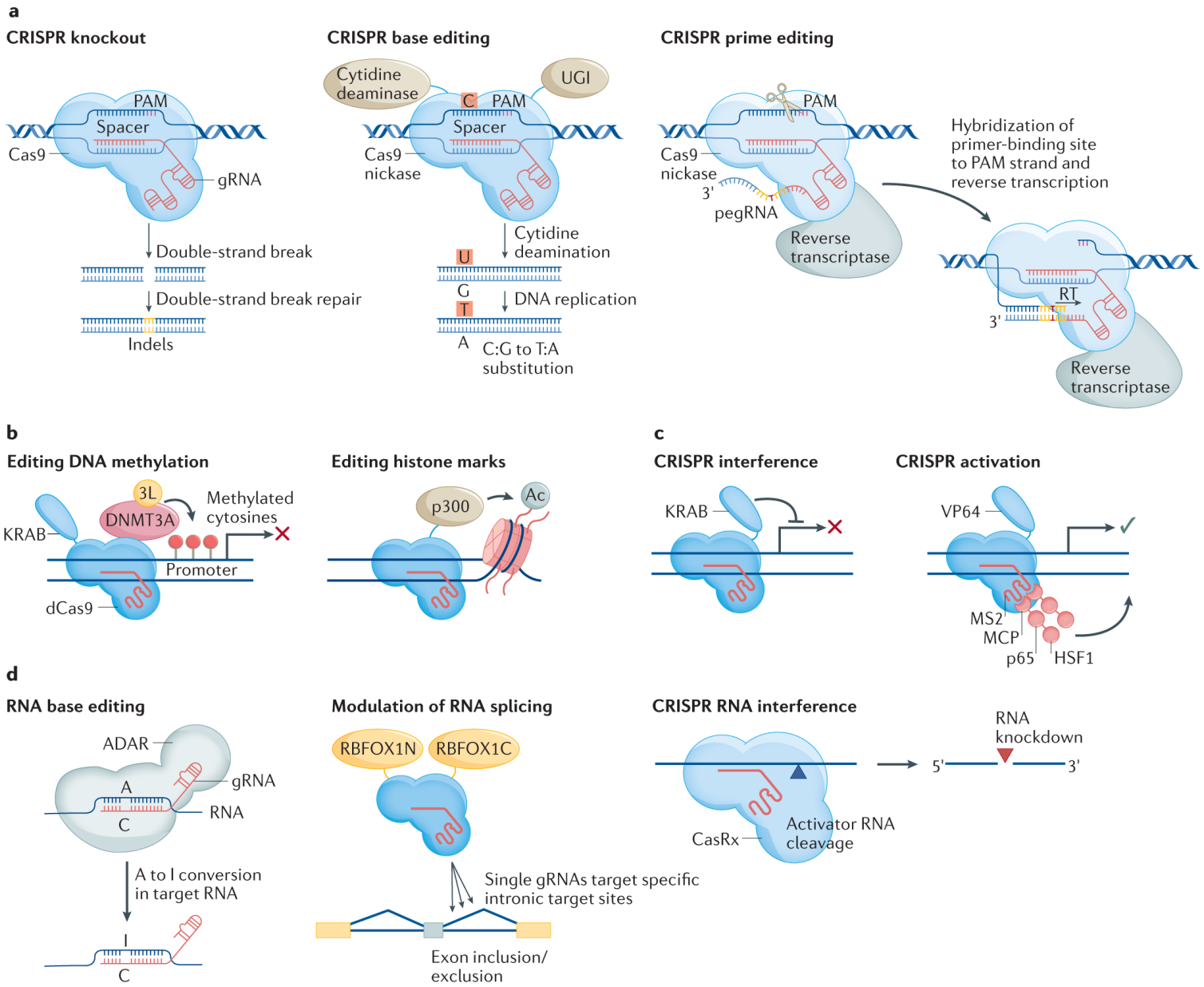
# Cas enzymes are programmable genome perturbation tools

## Established

- Cas9 (or Cas12a) – cut the genome
- dCas9 – activate (dCas9-VPR) or inhibit (dCas9-KRAB) gene expression with CRISPRi/a

## Emerging

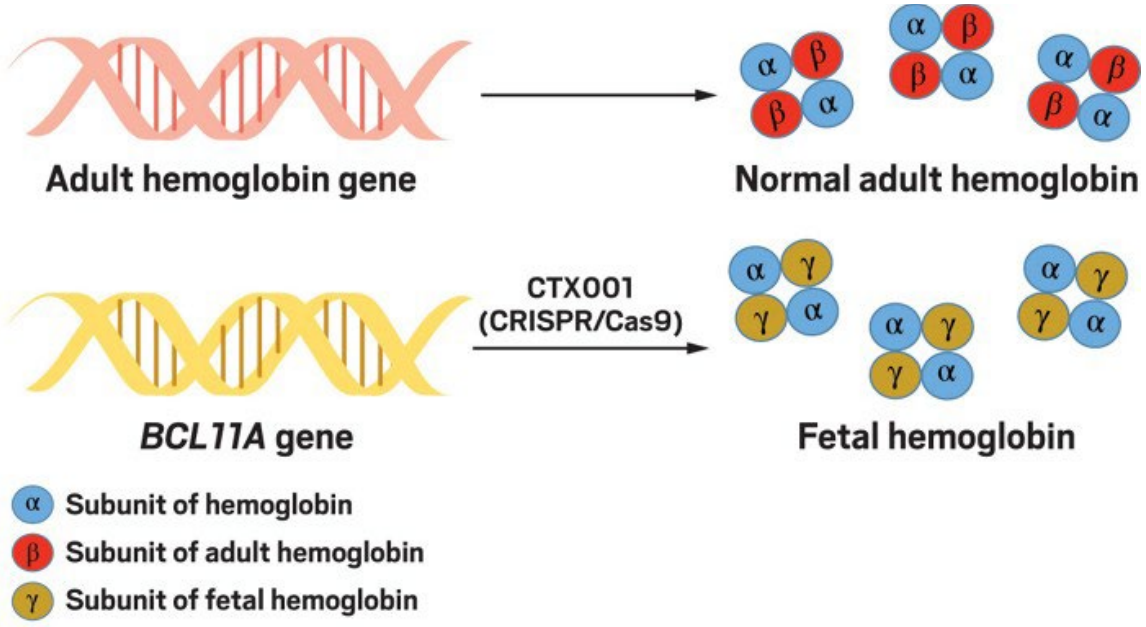
- Modify splicing or translation of RNA using Cas13d (potentially with splicing factors attached)
- Prime editing to insert defined indels (Cas9 nickase mutant + reverse transcriptase)
- Edit epigenetic marks with CRISPRon/off




https://doi.org/10.1038/s43586-021-00093-4

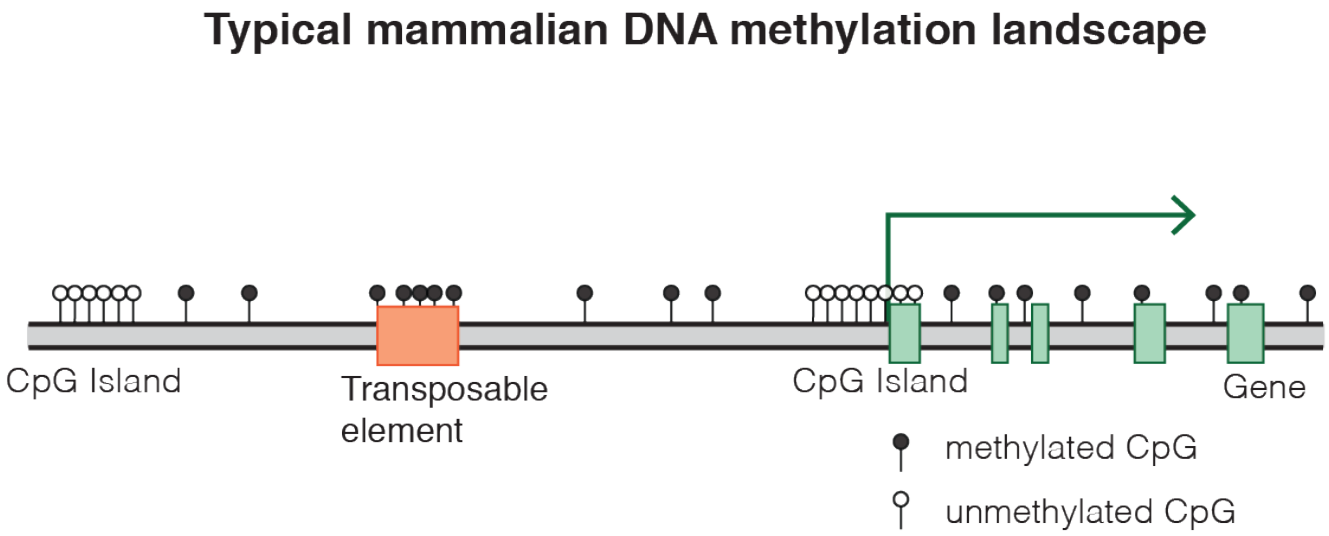
# Therapeutic appeal of “hit and run” epigenome editing

- It’s currently difficult to control the outcome of editing DNA using cutting, base editing, or prime editing. Truncated or mutant proteins can be immunogenic.
- Technologies like CRISPRi require constant expression of the effector, making them ill-suited as therapeutics
- **Could we instead permanently modify the epigenome by *transient* delivery of epigenome editors?**
  - E.g. Silence *CCR5* to confer resistance to HIV
  - E.g. Silence *BCL11A* to reactivate fetal hemoglobin expression to help treat sickle cell disease or β-thalassemia



# Permanent silencing using CRISPRoff

- Most human gene promoters contain CpG islands (region with high frequency of CG dinucleotides)
- >75% of CpG islands are methylated: this is the default state, and unmethylated CpGs generally indicate specific regulation
- Evolutionarily ancient form of gene regulation: CpG methylation correlates negatively with gene expression almost always in eukaryotes. Likely a form of genome defense against elements like transposons.

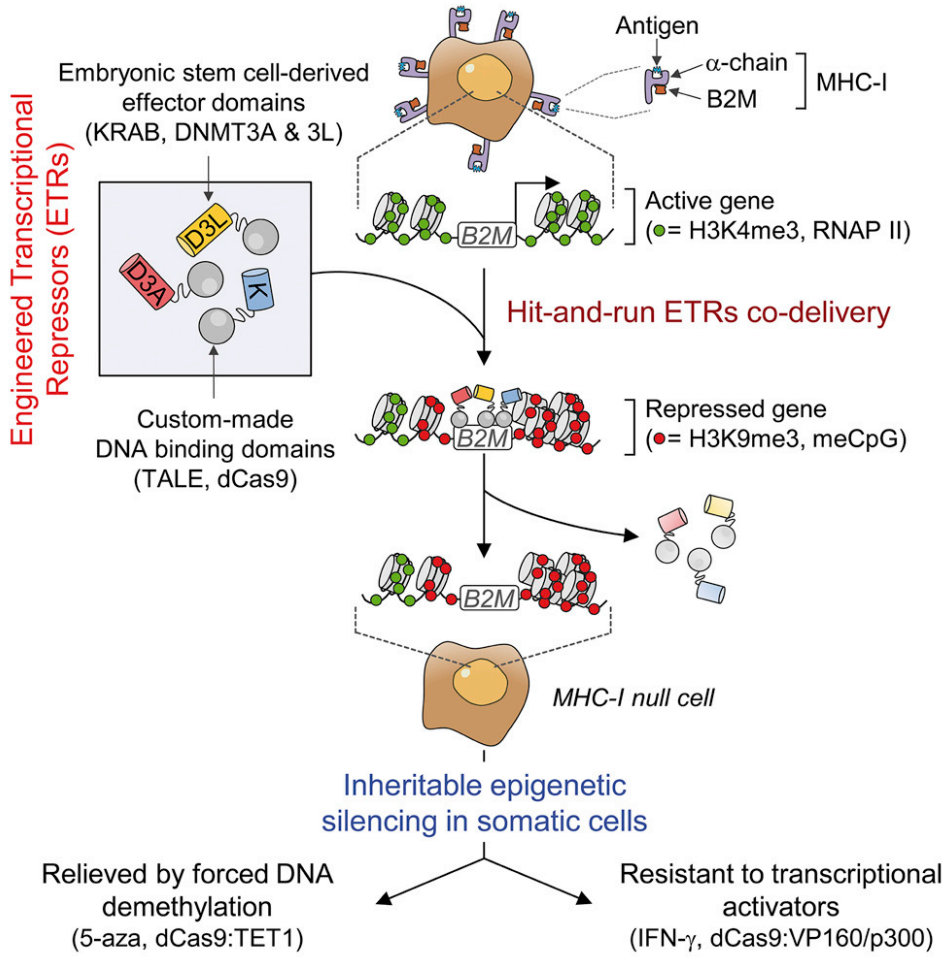


# Permanent silencing using CRISPRoff

- **KRAB zinc fingers:** are the initiators of the silencing cascade; bind to a site and recruiting epigenetic modifiers that set up a repressive chromatin environment

- **DNA methyltransferases (DNMTs):** eventually recruited and install CpG methylations
- DNMT3A and DNMT3B are the *de novo* methyltransferases that set up DNA methylation patterns early in development

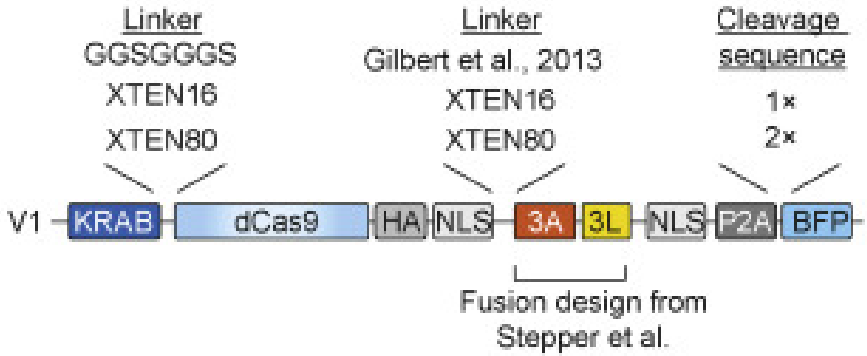
- This effect can "spread" over time upstream and downstream of the target

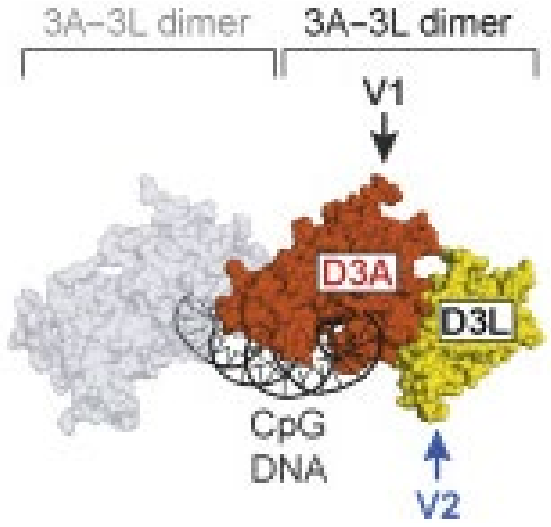



https://doi.org/10.1016/j.cell.2016.09.006

# Permanent silencing using CRISPRoff

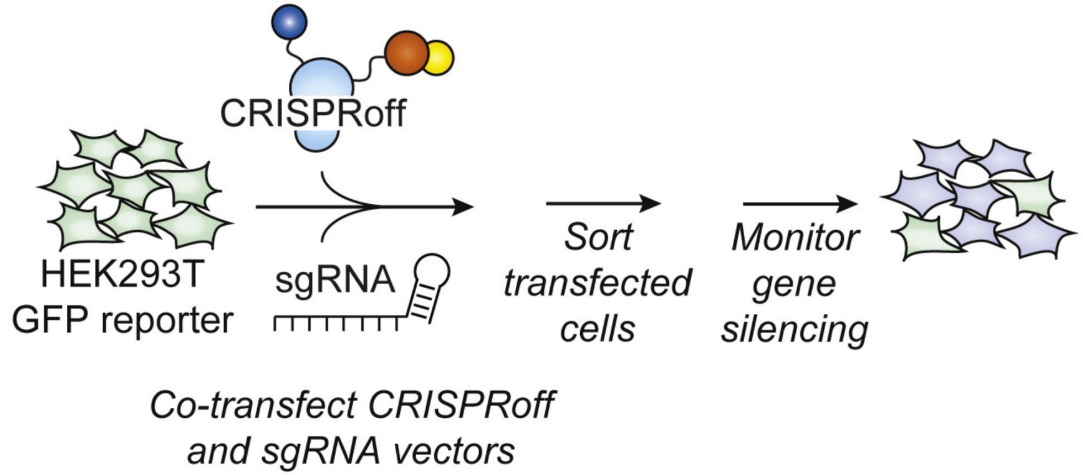
- In CRISPRoff, a four-part fusion of dCas9 to KRAB, DNMT3A, and DNM3TL (stimulates A) enables semi-permanent gene silencing through DNA methylation
- Design informed by structure of Cas9

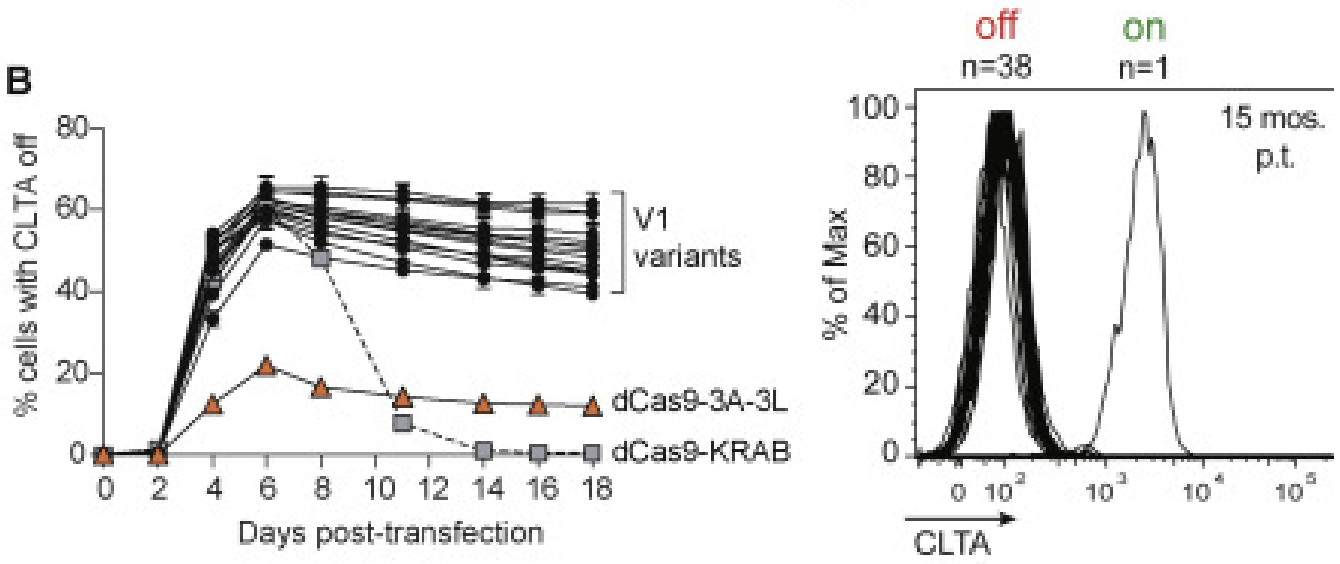


# Permanent silencing using CRISPRoff

- Delivered in “hit and run” format—transiently transfect effector + cocktail of sgRNAs
- (Note: now many studies indicating that using multiple sgRNAs can improve performance in various contexts)
- Can isolate clones that remain silenced for 15 months at some loci



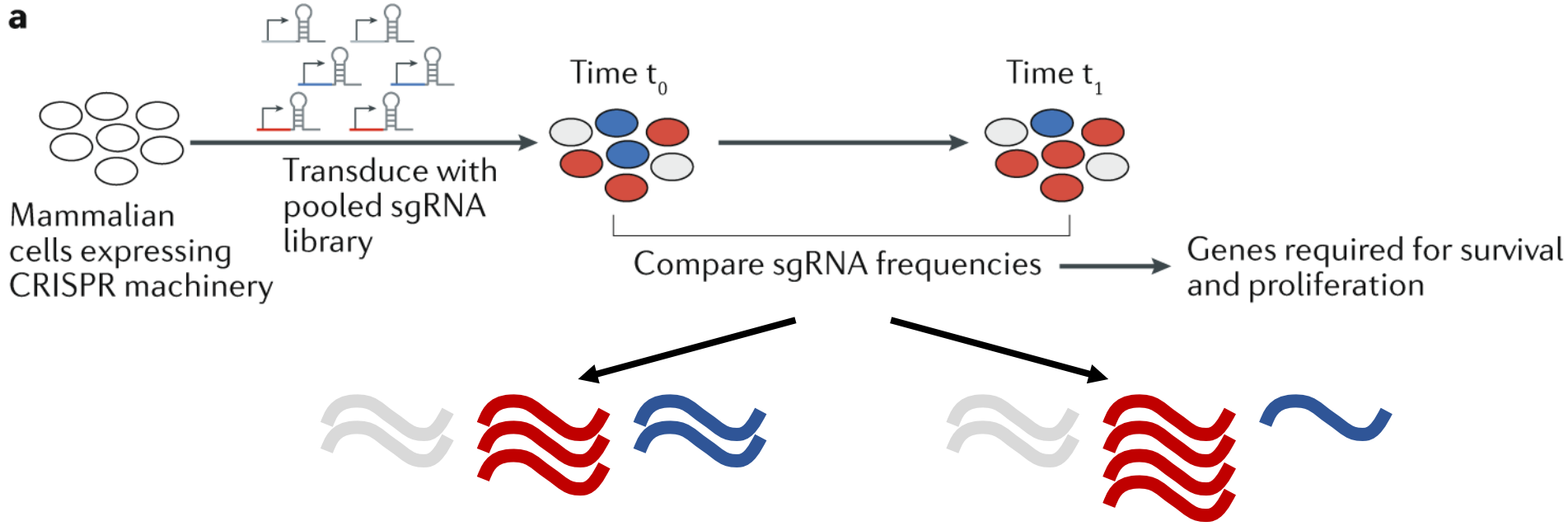
# Outline

1. Functional genomics background
   - Current CRISPR tools
   - **The four elements of a CRISPR screen**
2. High-content readouts
   - The DepMap
   - Development of Perturb-seq
   - Using genome-scale Perturb-seq to assign gene function
   - Using imaging CRISPR screens to assign gene function
   - *In vivo* screens and adding lineage information
3. High complexity perturbations
   - Genetic interaction mapping
   - Highly multiplexed studies of enhancers

# The four elements of a CRISPR screen

- **Model:** Which cells is the screen performed in?
- **Perturbations:** Which genes are perturbed and how?
- **Challenge:** What is the selective pressure?
- **Readout:** How is the phenotype of each perturbation quantified?

# The prototypical CRISPR screen: fitness/dropout

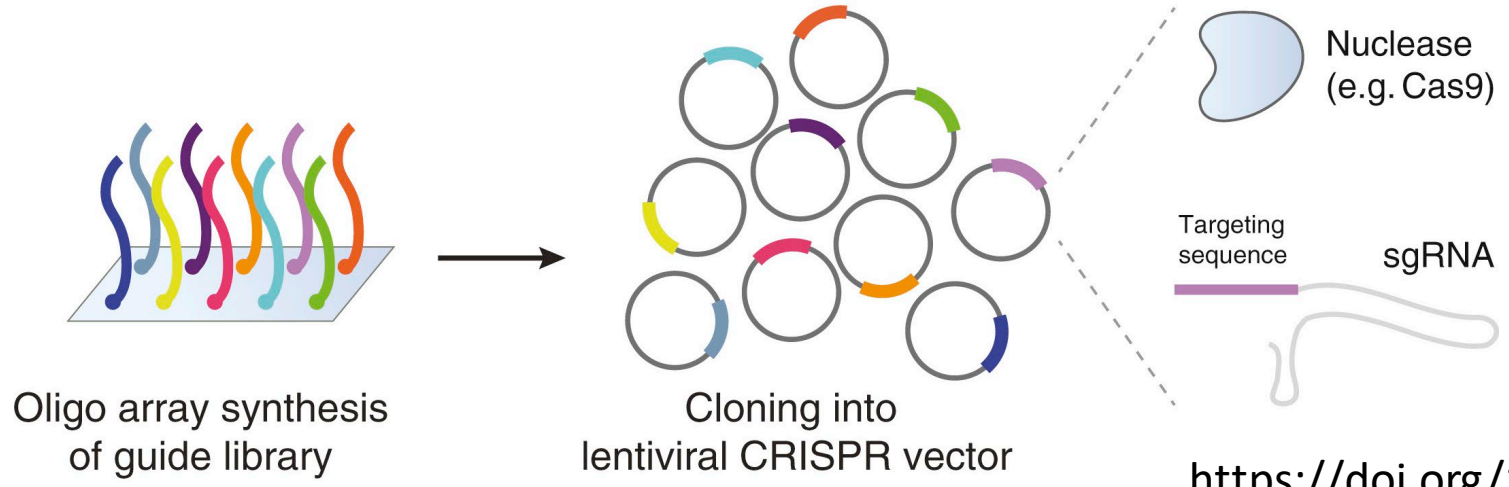



- **Model:** typically a cancer cell line
- **Perturbations:** Disrupt every gene in the genome (4 - 10 sgRNAs per gene)
- **Challenge:** outgrowth
- **Readout:** count number of cells with each sgRNA at each time point by sequencing

https://doi.org/10.1038/s41582-020-0373-z

# Where to get sgRNA libraries

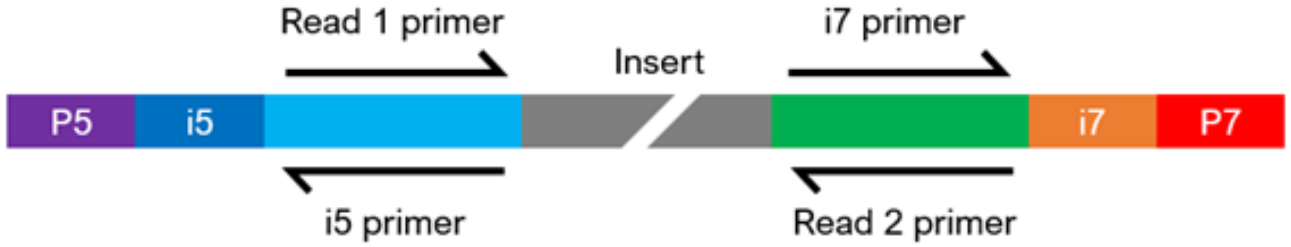
- For screening, a cell line is typically constructed containing the desired Cas9 effector, and sgRNAs are delivered separately by pooled lentiviral transduction.
- sgRNA libraries are synthesized and cloned in pooled format. Tens of thousands of sgRNAs possible.
- Efficacy of different sgRNAs against the same gene can be variable and hard to predict. To buffer, libraries usually include 4 – 10 sgRNAs per targeted gene.
- There are published libraries for many of the most common effectors. The Broad's CRISPick tool is popular, as are the CRISPRi/a libraries from the Weissman Lab.




https://doi.org/10.1016/j.ab.2016.05.014

# Reading out a CRISPR fitness screen by sequencing

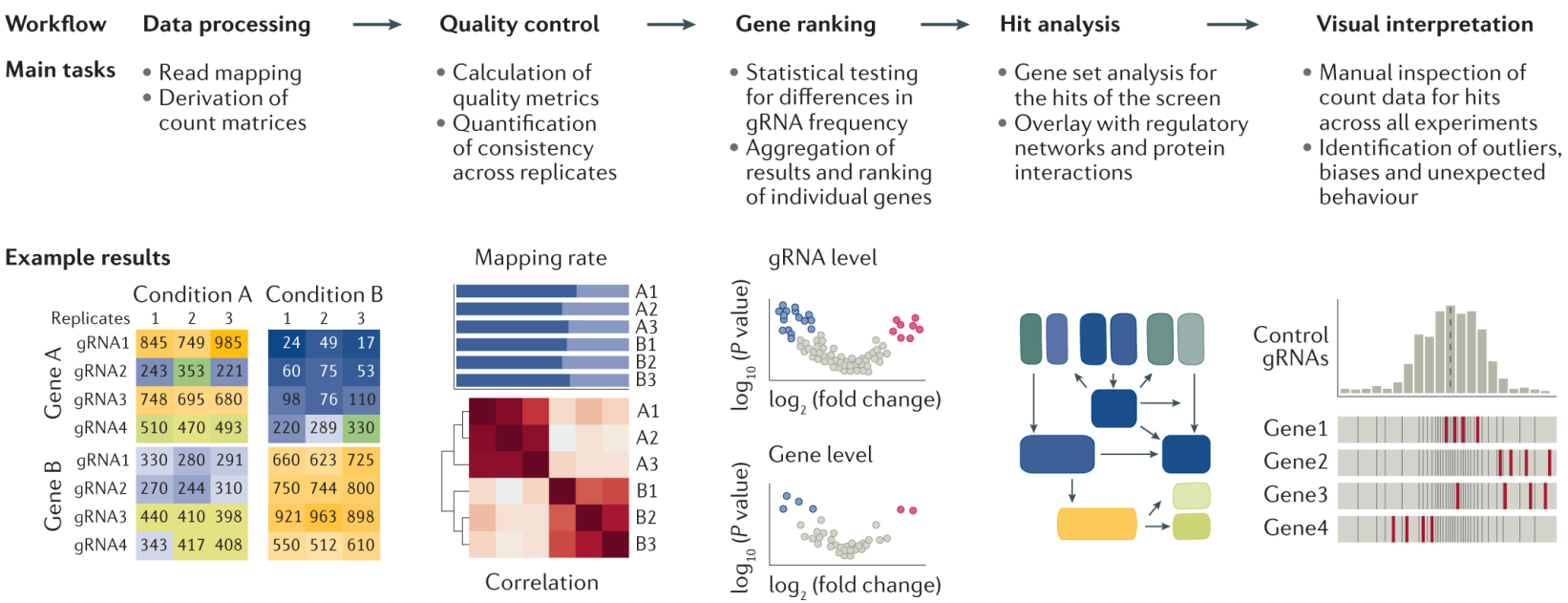
- Screens typically amplify the sgRNA protospacer from genomic DNA.
- This is a single-molecule assay—only one copy per cell. As a result, library prep must be done in "clean" conditions to avoid contamination (e.g. by ambient plasmid DNA).
- Library prep is essentially a careful PCR with primers that attach the sequences necessary for Illumina sequencing:




- Quick tour:
  - Two reads (can be up to 300 bp long and are what you typically think of as the result of sequencing)
  - Two index reads (read usually 8 bp barcodes that are used to distinguish different libraries on same sequencing flow cell)
  - P5 and P7 are sequences that are necessary for binding to Illumina flow cell during sequencing

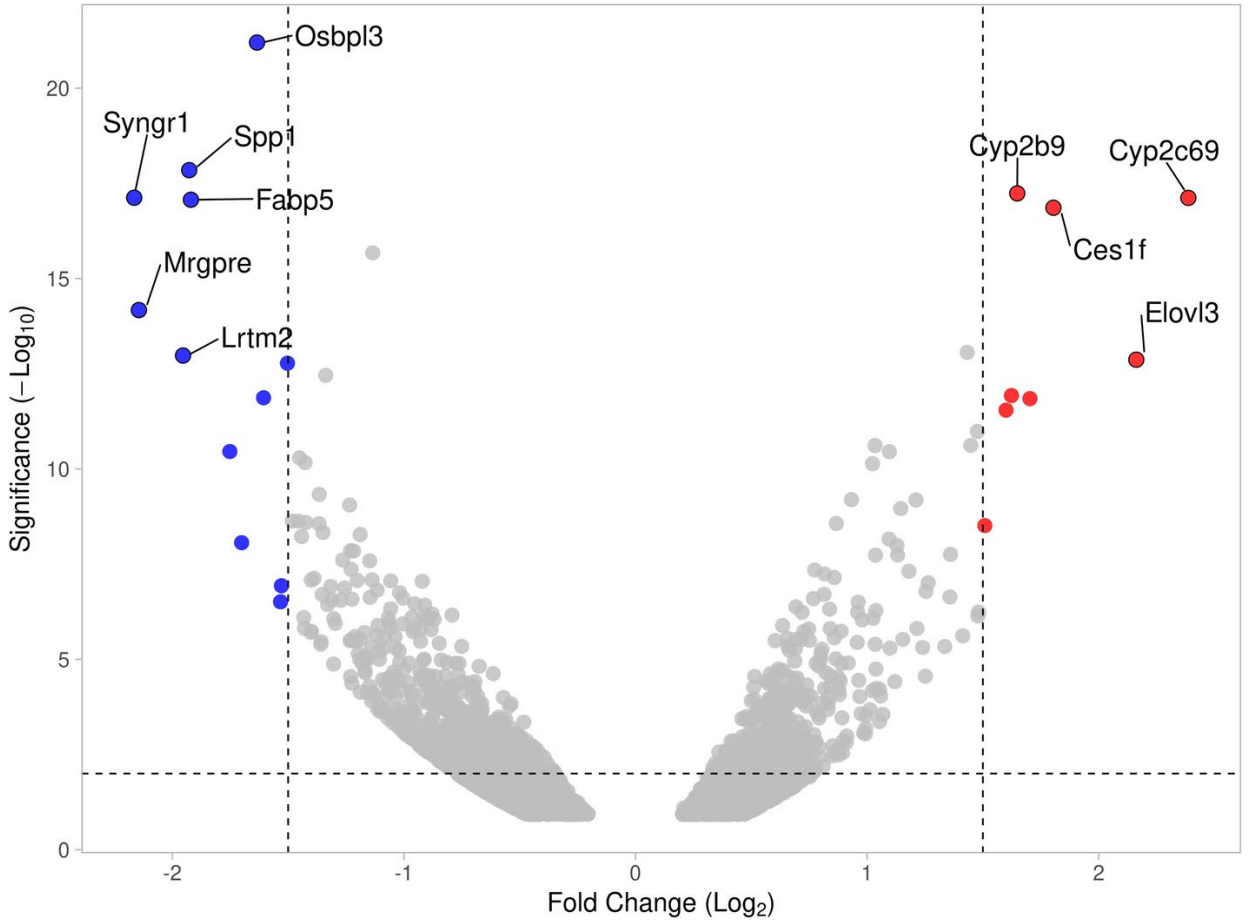
# What do the data look like?

- Result is a large matrix of counts for each sgRNA across replicates/conditions
- Fit a model to integrate effects across guides targeting the same gene while correcting for confounders such as variable guide representation and outgrowth during the screen
- Model output is an estimate of enrichment or depletion for each gene relative to cells with non-targeting control sgRNAs, plus some measure of statistical significance derived from how concordant behavior is among different sgRNAs targeting the same gene



# What do the data look like?

- Screen results are typically presented as volcano plots comparing effect size vs. statistical significance
- Genes that “explode out of the volcano” are hits
- Strive to be sensitive both to big and weak but reproducible phenotypes



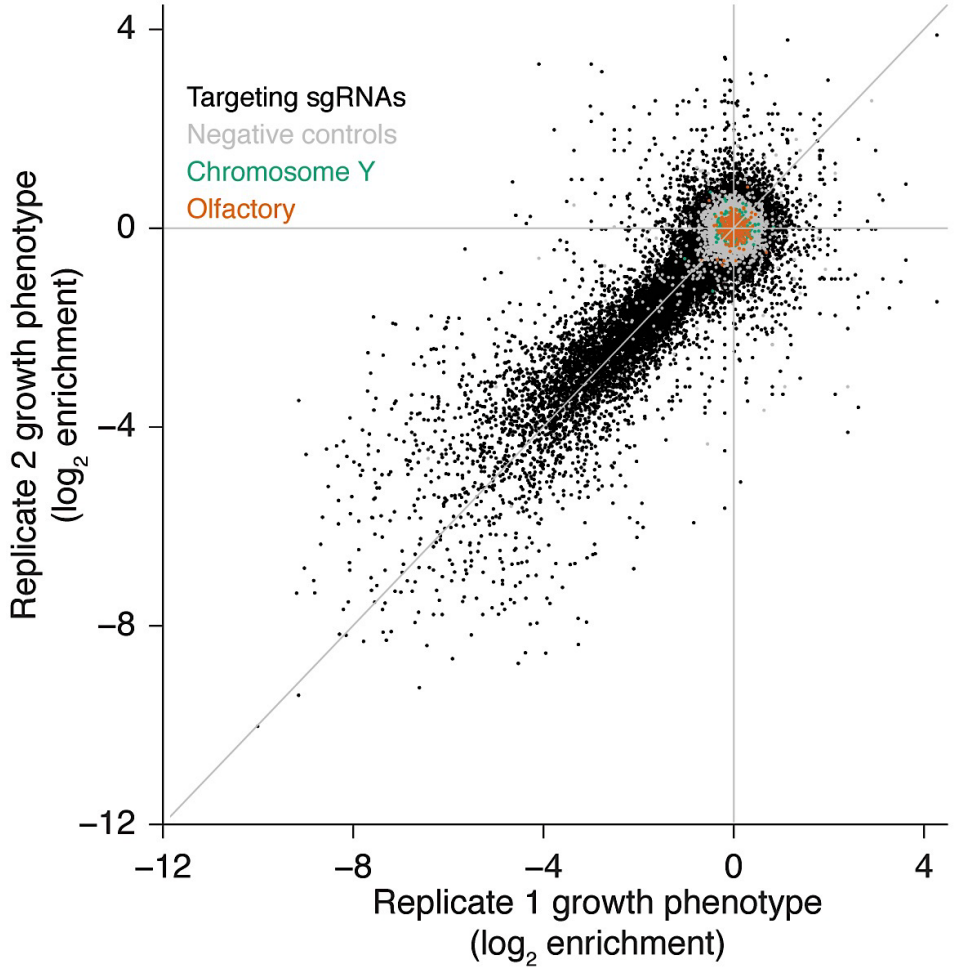

**Significance (e.g. -log p value)**

**Effect size (e.g. fold enrichment)**

# Example: genome-wide measurement of fitness effects

- Fitness screens can measure hundreds of thousands of quantitative phenotypes over orders of magnitude in a single pooled experiment
- A high-quality screen will show good correlation across replicates

- **Model:** K562 cells
- **Perturbations:** Knockdown every gene in the genome with CRISPRi (200k sgRNAs)
- **Challenge:** Growth
- **Readout:** growth defect relative to control




Gilbert & Horlbeck et al. *Cell* 2014

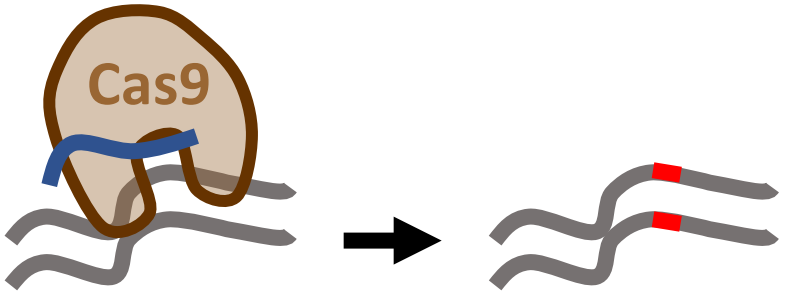
# Coverage (aka how many cells do I need?)

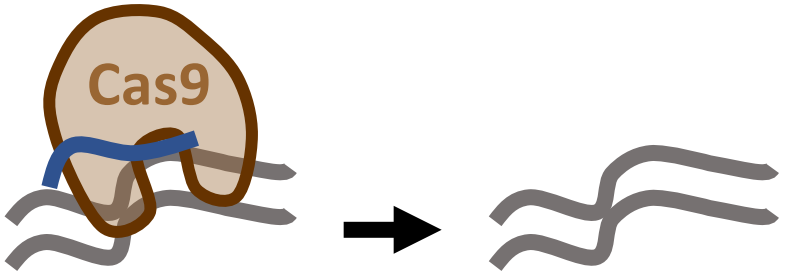
- Number of cells per sgRNA = coverage. This sets the quantitative resolution of the screen.
- In **negative selection (dropout) screens,** perturbed cells deplete relative to controls
  - Example: when using CRISPRi, essential genes will dropout over time
  - Here people often suggest 500× – 1000× coverage per sgRNA. E.g. for a 100k element genome-wide library, 50 – 100 million cells per replicate.
- In **positive selection screens**, perturbed cells enrich relative to controls
  - Example: a CRISPRa screen for genes that mediate drug resistance
  - Here lower coverage is possible due to the enrichment, so 200× may suffice

- If coverage is too low, sgRNAs or genes will become overrepresented by chance. The screen will then be **bottlenecked. <u>This is an important consideration for *in vivo* screens.</u>**
- Bottlenecking will result in poor correlation between replicates

# Incomplete outcomes are important confounders

- CRISPR perturbations do not always work
  - CRISPR cutting can produce in-frame indels or only edit one copy of a gene, creating a heterozygote
  - If knockdown of a gene by CRISPRi is toxic, there is selection for cells to escape it. These cells will outgrow the perturbed cells over time.
  - Some guides simply work better than others



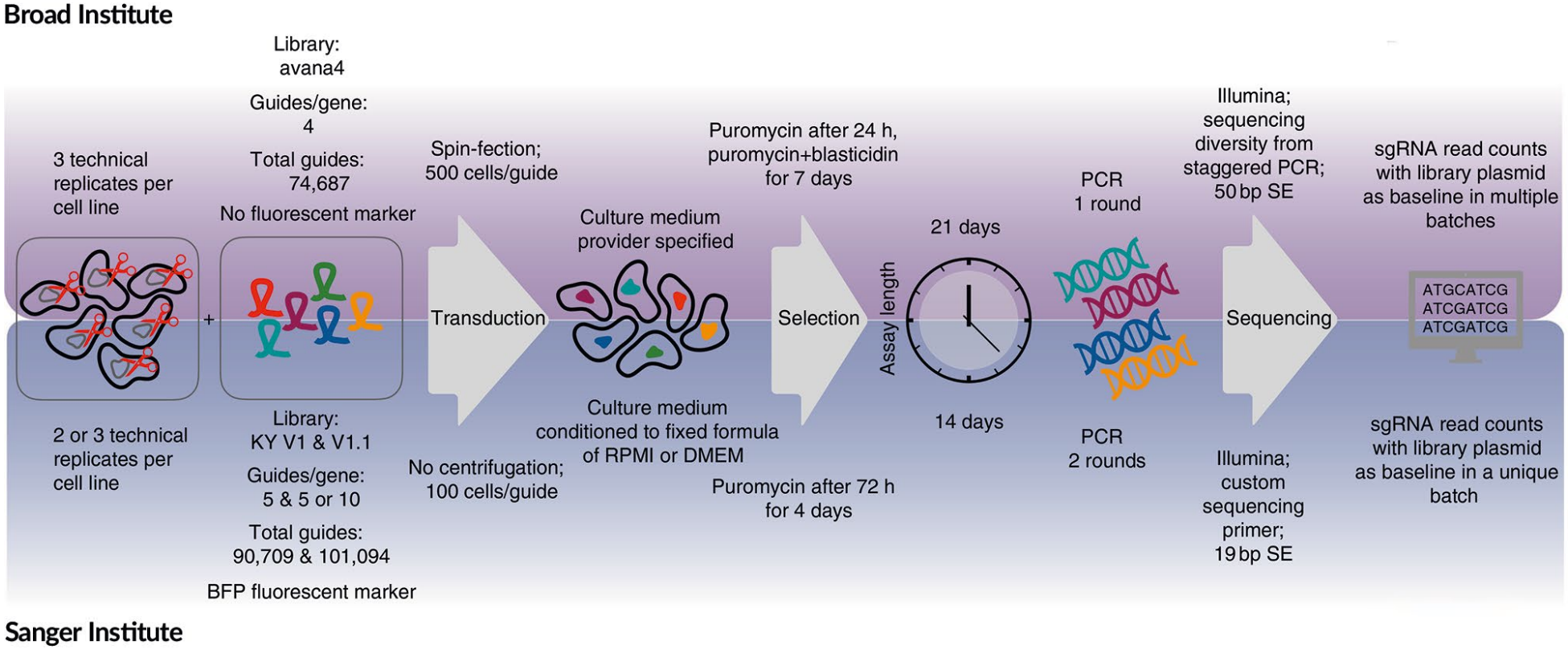

- Overall, critical to remember what is happening mechanistically when conducting analyses—screens are large evolution experiments

# Outline

1. Functional genomics background
   - Current CRISPR tools
   - The four elements of a CRISPR screen
2. **High-content readouts**
   - **The DepMap**
   - Development of Perturb-seq
   - Using genome-scale Perturb-seq to assign gene function
   - Using imaging CRISPR screens to assign gene function
   - *In vivo* screens and adding lineage information
3. High complexity perturbations
   - Genetic interaction mapping
   - Highly multiplexed studies of enhancers

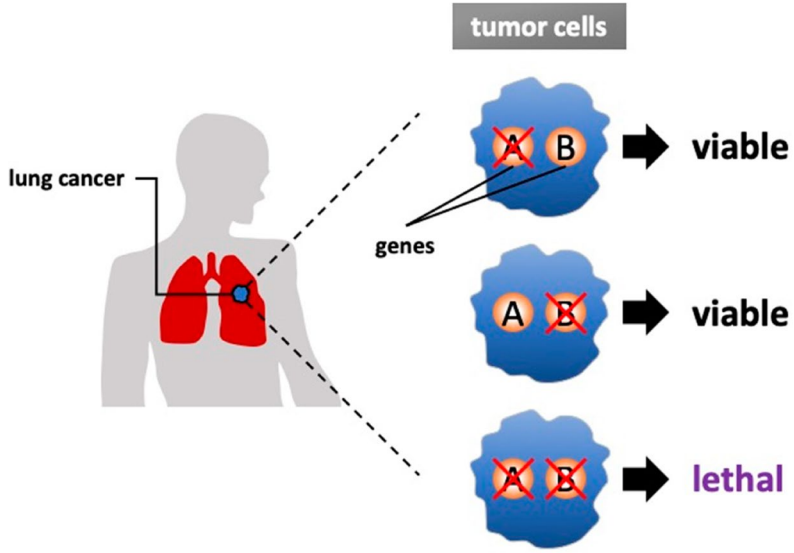
# Dependency Map (DepMap)/Project Score

- **Goal:** Find genes that are (conditionally) essential in different cancers
- Each project performed genome-wide CRISPR cutting screens in hundreds of cancer cell lines



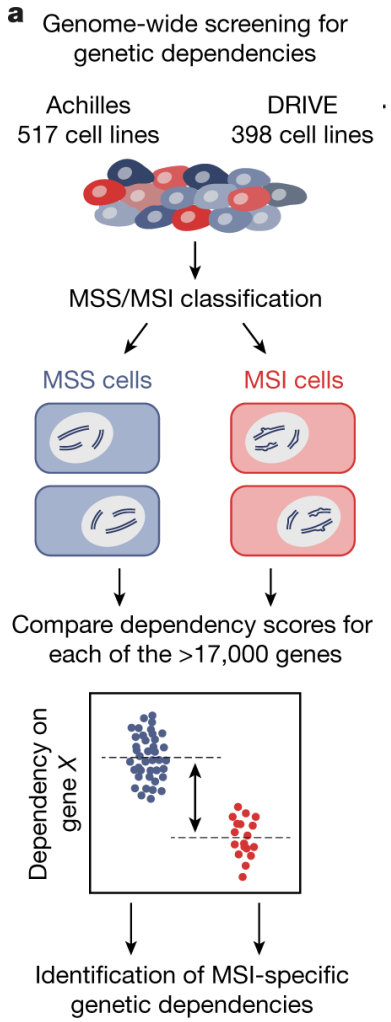
# The DepMap can be used to find synthetic lethal pairs

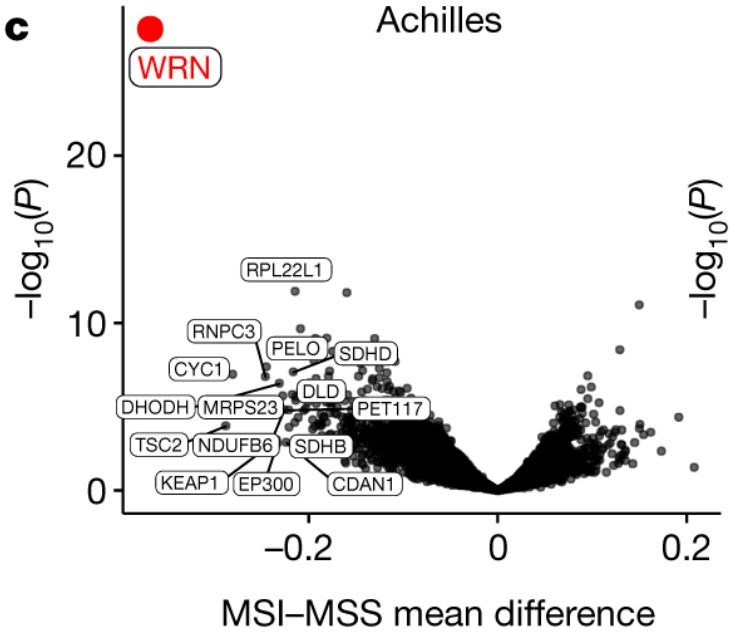
- Two genes A and B are **synthetically lethal** if cells can survive loss of each on its own, but the double mutant is lethal
- Potential route to targeted therapies: one gene might be mutated in a cancer and the other can be inhibited pharmacologically



# The DepMap can be used to find synthetic lethal pairs

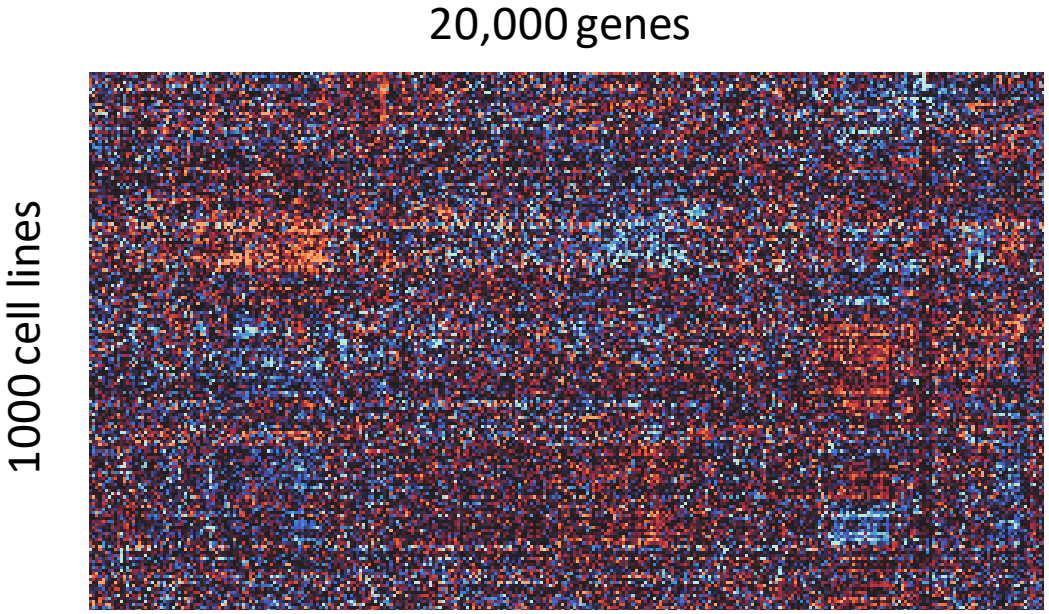
- Defects in mismatch repair lead to frequent mutations at nucleotide repeat regions called microsatellites
- Stratified cancers according to MSI (microsatellite instability) and looked for genetic drivers
- Werner helicase (WRN), a gene involved in DNA repair and replication, was an obvious outlier, and is a synthetic lethal target in MSI cancers
- Side note: this is also the causal gene for a premature aging syndrome called Werner syndrome



# The DepMap can be used to assign gene function

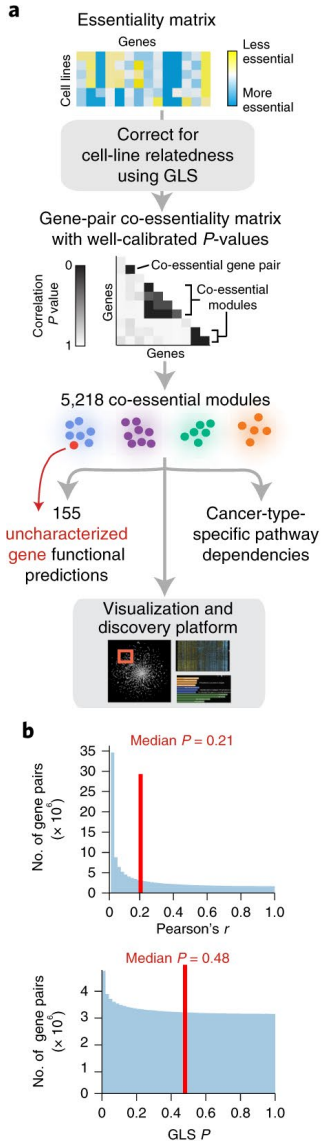
- A gene's fitness profile across diverse genotypes is a type of high-content phenotype that can be clustered to identify related genes
- Essentially we are using the phenotypic and genetic variation intrinsic to different cancer cell lines to reveal gene function

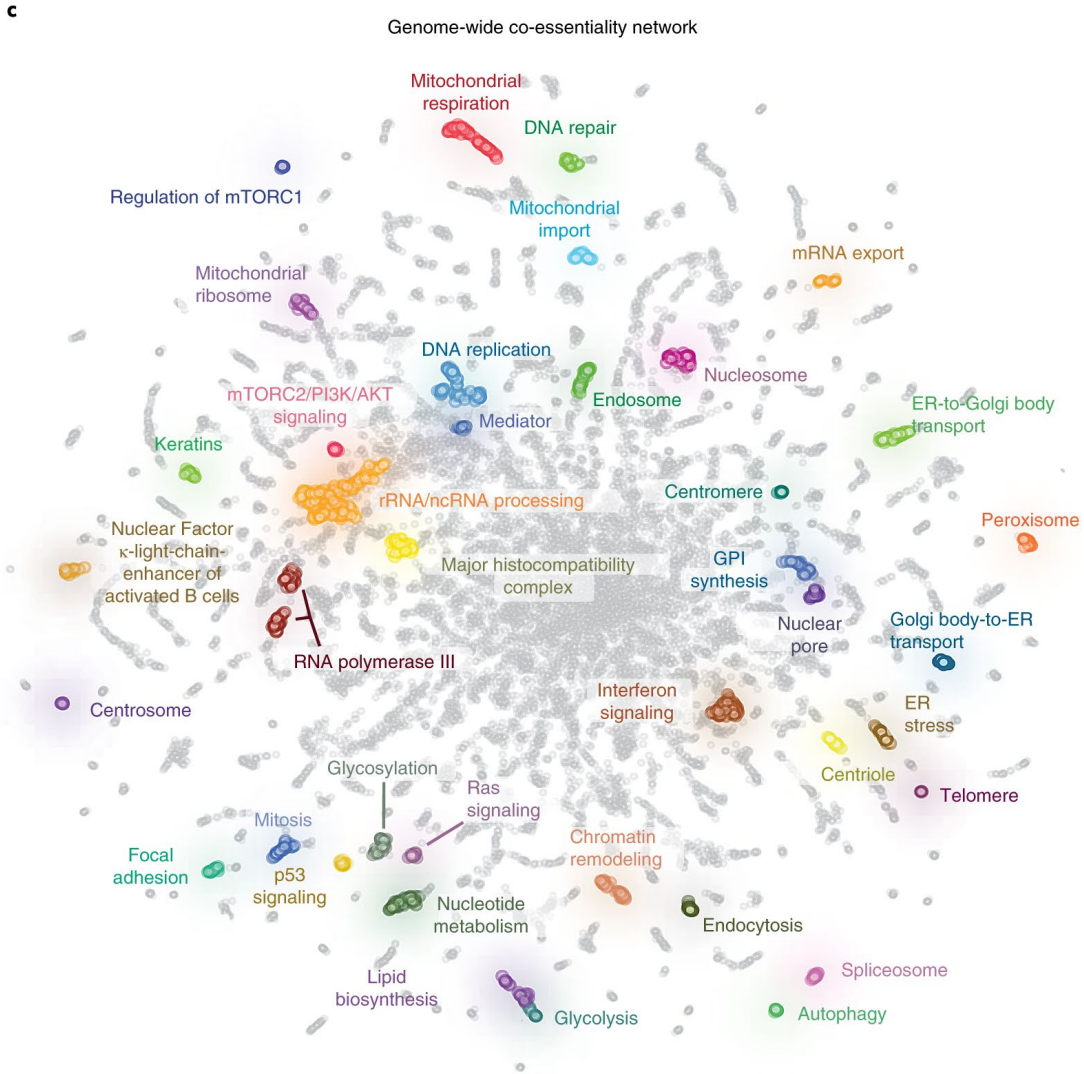



https://doi.org/10.1038/s41588-021-00840-z

# The DepMap can be used to assign gene function

- Not trivial: many cell lines in the DepMap are similar (e.g. same source tissue), so measurements are not independent
- To look for statistically significant associations, must correct for these dependencies
- Create nonlinear embedding of each gene based on these corrected profiles

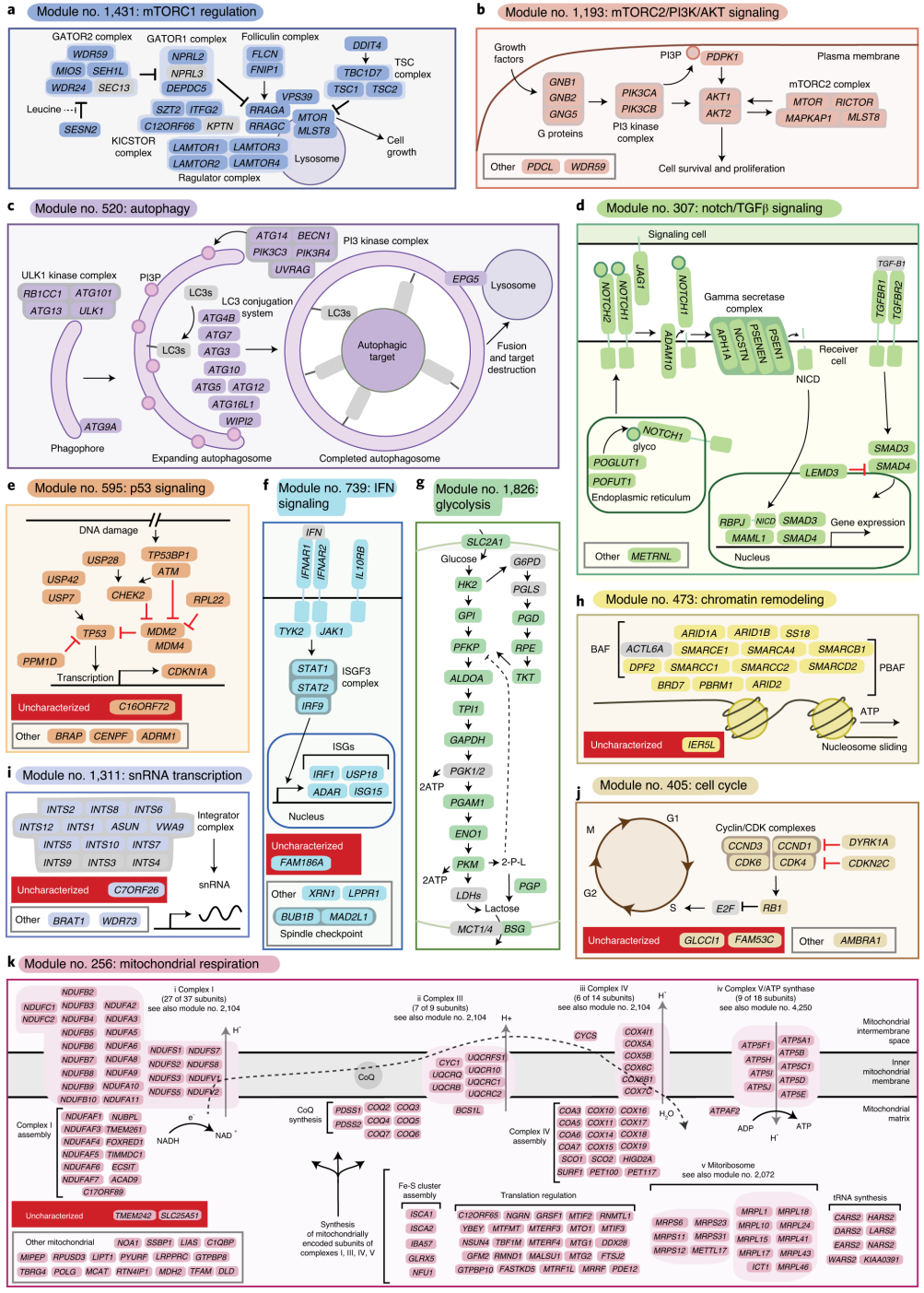



doi.org/10.1038/s41588-021-00840-z

# The DepMap can be used to assign gene function

- Lots of structure explained by known biology, can nominate new genes within known pathways based on the clustering
- Overall, a great resource for learning about possible interactors with your genes of interest

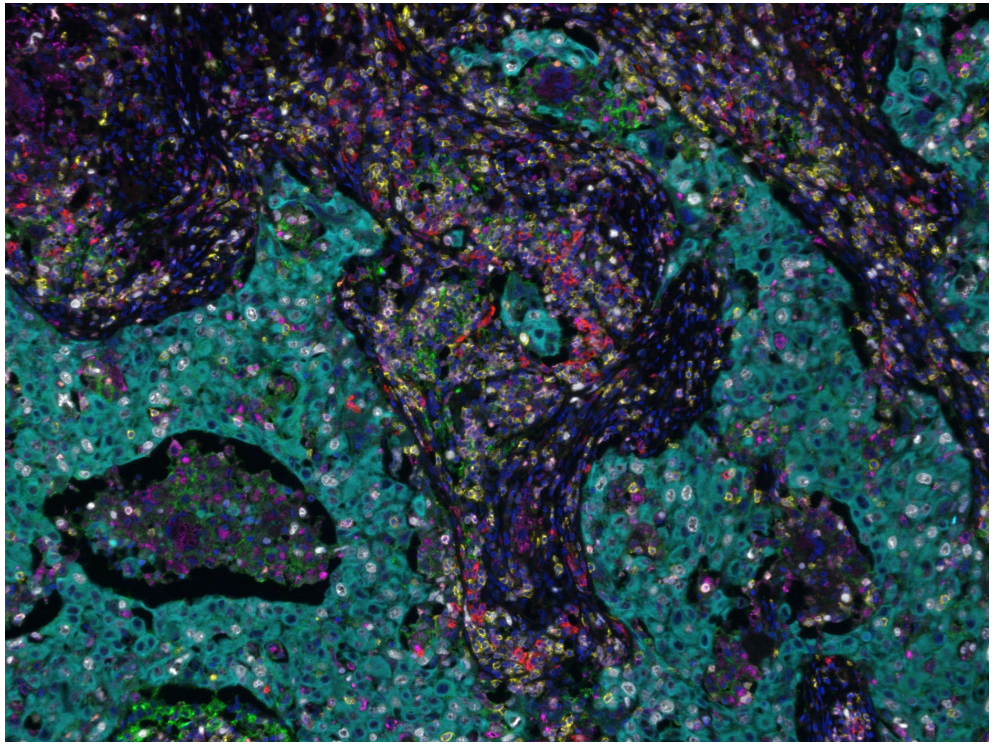



doi.org/10.1038/s41588-021-00840-z

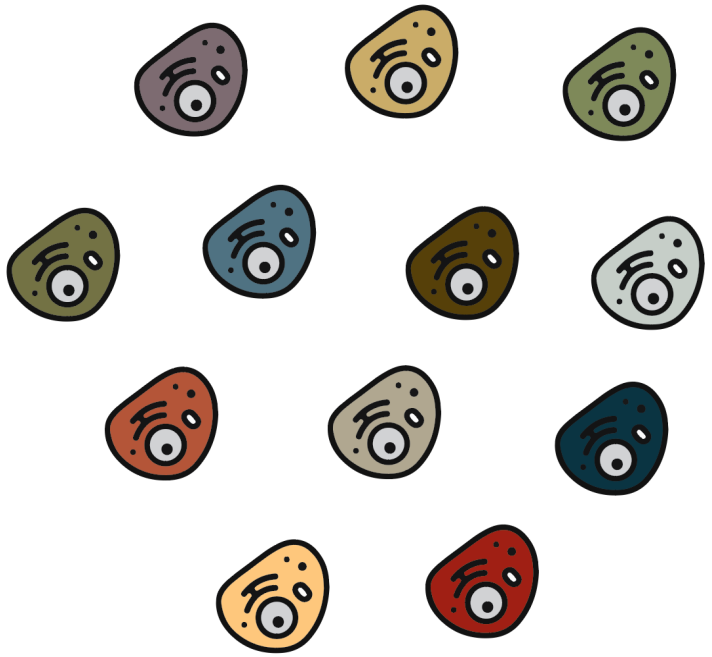
# Outline

1. Functional genomics background
   - Current CRISPR tools
   - The four elements of a CRISPR screen
2. High-content readouts
   - The DepMap
   - **Development of Perturb-seq**
   - **Using genome-scale Perturb-seq to assign gene function**
   - Using imaging CRISPR screens to assign gene function
   - *In vivo* screens and adding lineage information
3. High complexity perturbations
   - Genetic interaction mapping
   - Highly multiplexed studies of enhancers

# Two benefits of single-cell technologies



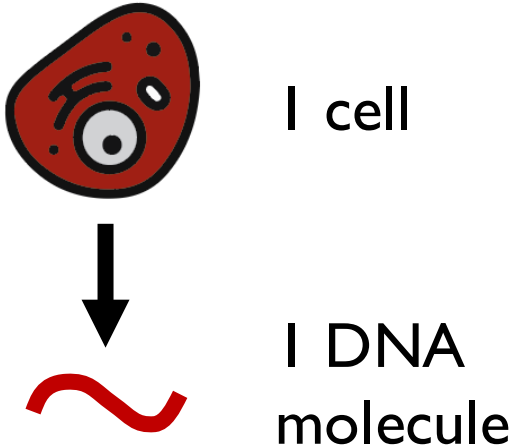
Understand how distinct cell types contribute to large-scale phenotype



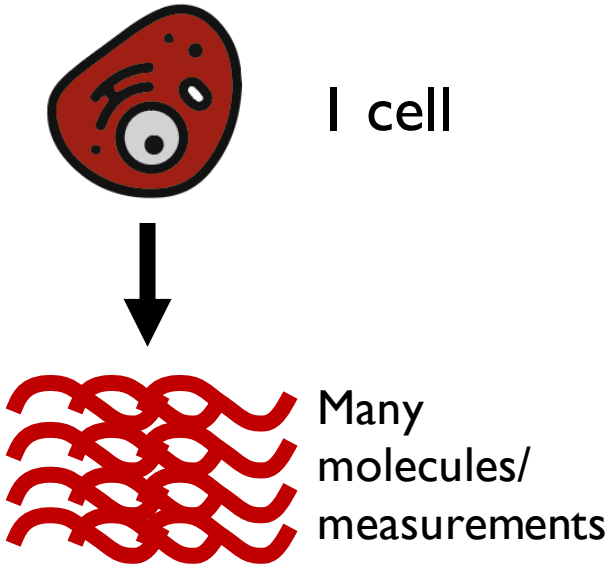
A means of achieving parallelization—each cell is an experiment

# Single-cell CRISPR screens

- All CRISPR screens are single-cell assays: each cell reports sgRNA contained within




- **Can we get more information from each cell?**
  - Richer phenotypes are more interpretable – e.g. cell cycle arrest or cell death might be conflated as “fitness defects”
  - More information per cell means fewer cells needed, more applicable to precious cell types
  - Can expand screening to systems where cells are not actively growing and dividing, such as *in vivo*



# Single-cell CRISPR screens

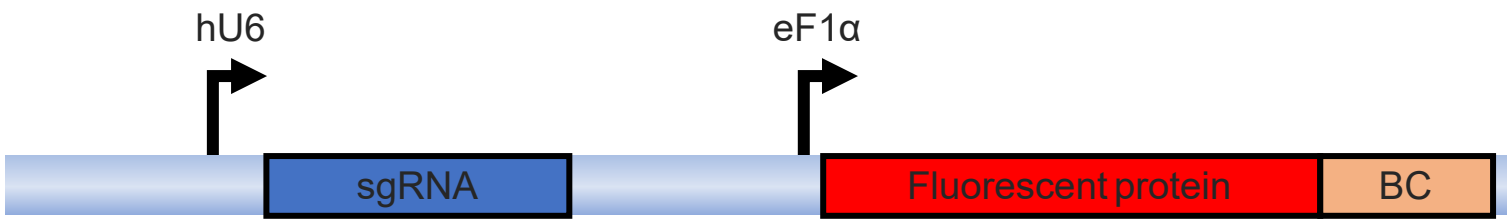
- Natural synergy with the many emerging methods for gathering rich phenotypes of single cells—single-cell genomics and imaging

| Target | Technique |
|---|---|
| Transcriptome | Single-cell RNA sequencing<br>Multiplexed RNA FISH (MERFISH, seqFISH) |
| (Surface) Proteome | Barcoded antibodies (CITE-seq/REAP-seq)<br>Flow/Mass cytometry (Pro-Codes) |
| Chromatin accessibility | Single-cell ATAC-seq |
| Morphology/localization | Imaging |

# Single-cell CRISPR-screens: the devil is in the details

We will start by looking at the original Perturb-seq approach

- **Idea c. 2015:** want to somehow read out how each cell is perturbed genetically, along with some rich measure of its state
- **Problem:** most single-cell RNA sequencing methods only capture polyadenylated transcripts, sgRNAs are expressed by Pol III and are not captured
- **One solution:** Encode identity in an auxiliary barcode transcript that is captured



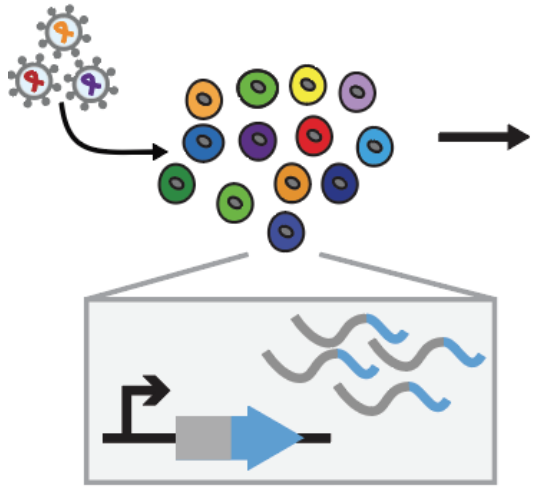
# Example: Perturb-seq: a screen with RNAseq as the readout

**Introduce barcoded pool of CRISPR perturbations**

**Droplet single-cell RNAseq**

**Single-cell transcriptomes tagged with perturbation identity**

**Rich characterizations of hundreds to thousands of perturbations**



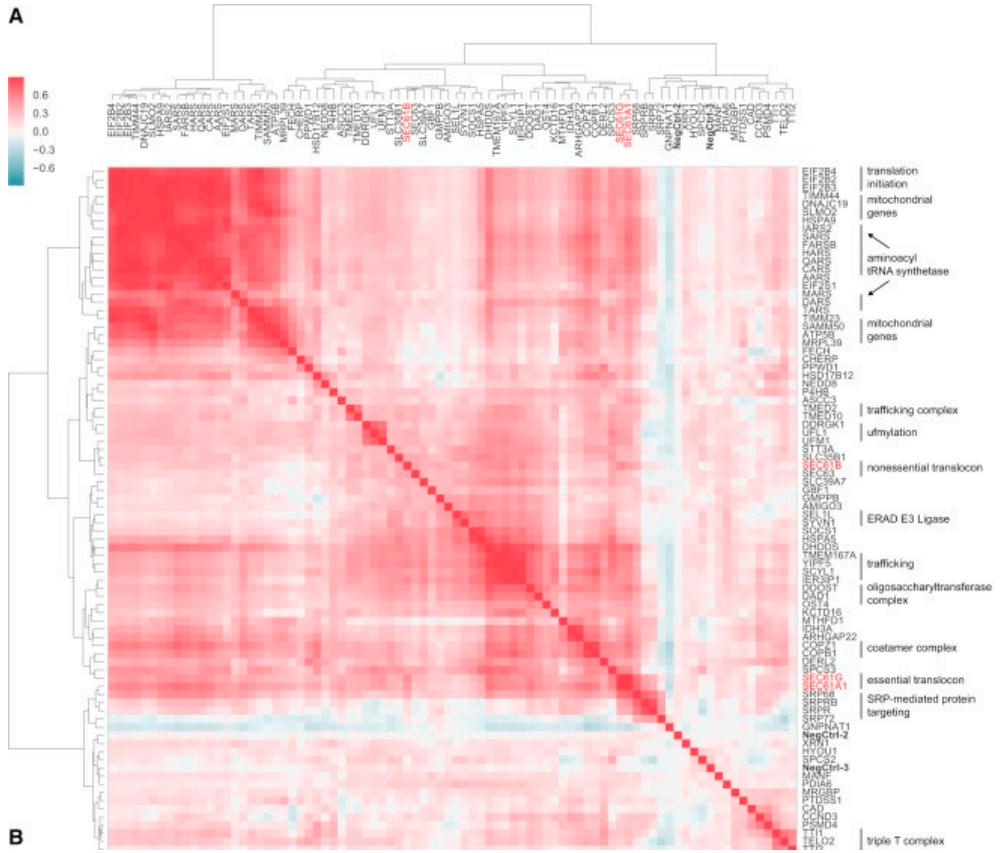
perturbation identity encoded in transcript

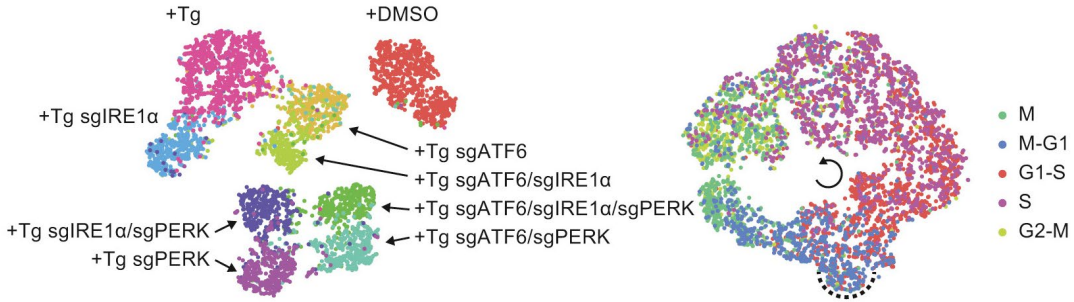
Adamson & Norman et al., *Cell* 2016, Dixit & Parnas et al., *Cell* 2016, Jaitin et al., *Cell* 2016, Datlinger et al., *Nature Methods* 2017

# Interpreting results

- **Comprehensive:** Thousands of "virtual screens" performed in parallel. Phenotypes don't need to be prespecified—can revisit data with a new hypothesis later.

- **Informative:** can discover genes associated with a process *and* reason about why they appear
  - E.g. Cluster RNAseq profiles to identify genes with similar functions
  - Can identify and ignore "boring" hits
  - But can also do more sophisticated machine learning. E.g. Separate independent effects like cell cycle and genetic perturbations.

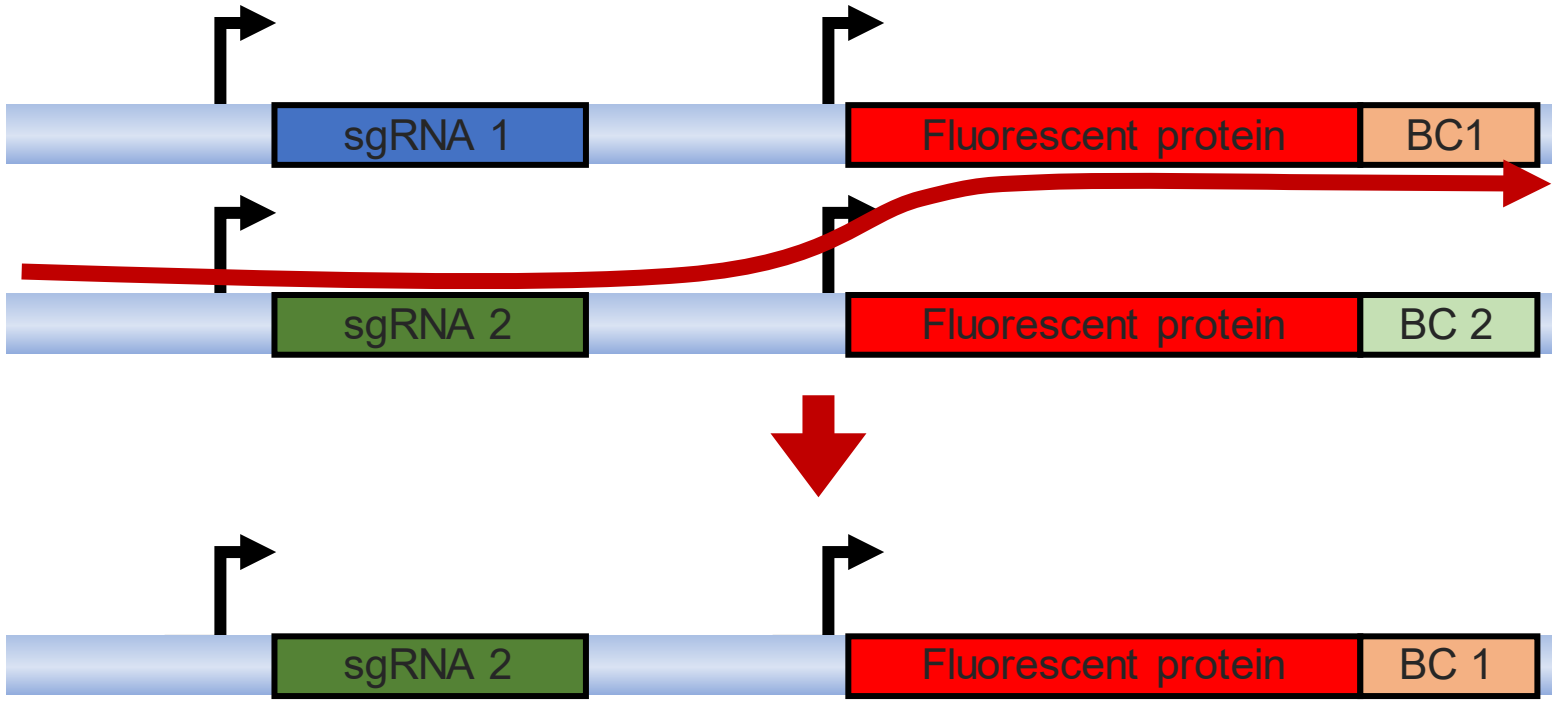



https://doi.org/10.1016/j.cell.2016.11.048

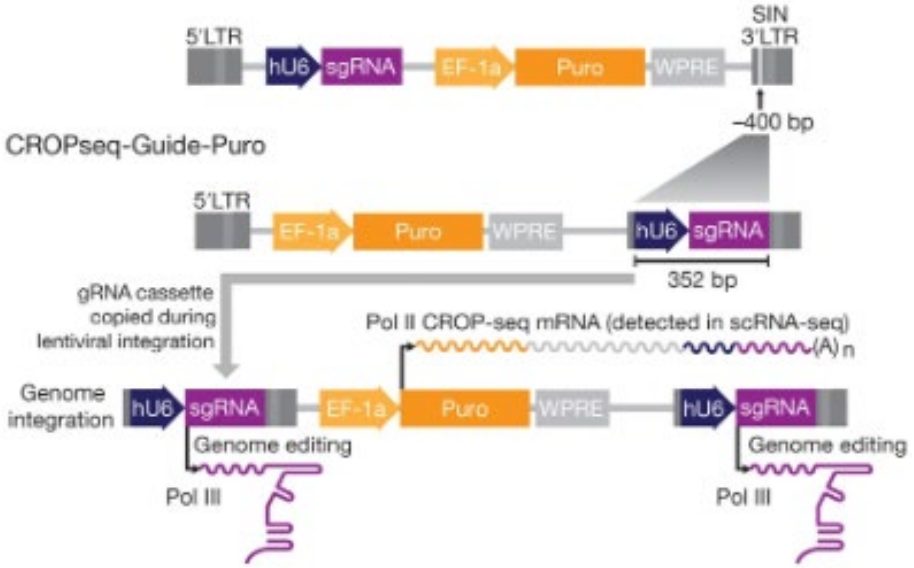
# Subsequent improvements to avoid template switching

**Lentiviral template switching:** lentiviruses are pseudodiploid and randomly hop between genomes during transduction, can lead to recombinant viruses with scrambled association between sgRNA and barcode when libraries are packaged in pooled format. Surprisingly pernicious: two markers separated by 1 kb can lead to ~30% rate of uncoupling. Common problem in the literature!

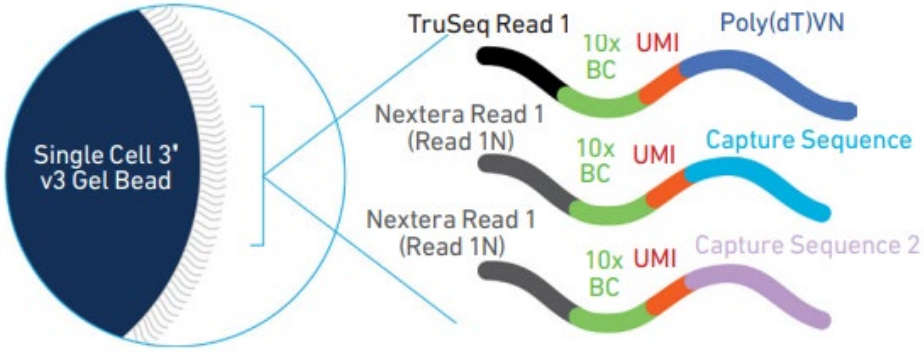


# Subsequent improvements to avoid template switching

- **CROPseq vector:** exploits a trick of lentiviral replication to embed the sgRNA during integration into the target cell

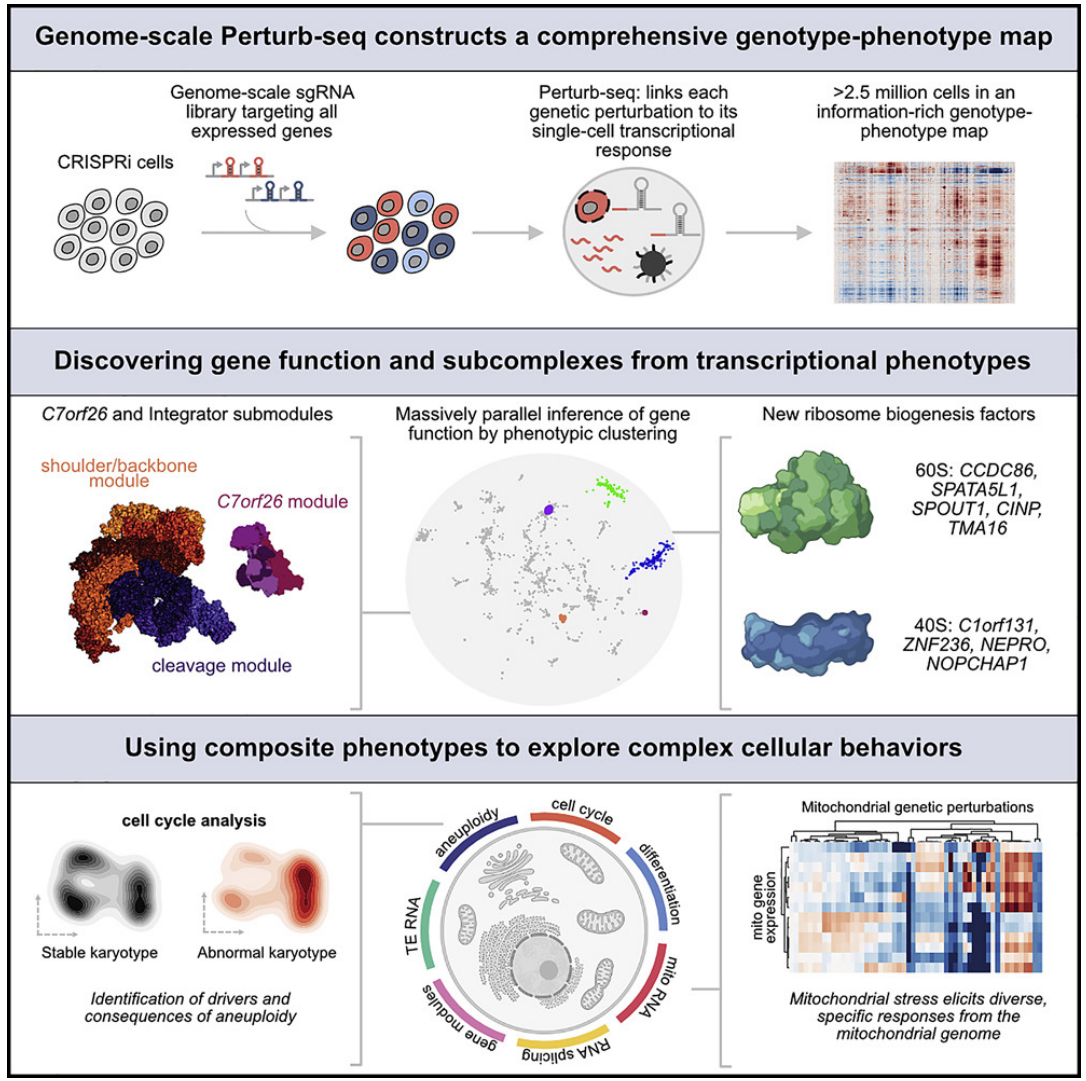



- **Direct sgRNA capture:** manipulate single-cell assay to capture sgRNAs as well as mRNAs (e.g. 10x Feature Barcoding)




doi: 10.1038/s41587-020-0470-y doi: 10.1038/nmeth.4177

# Genome-scale Perturb-seq to understand cell biology




Joseph Replogle

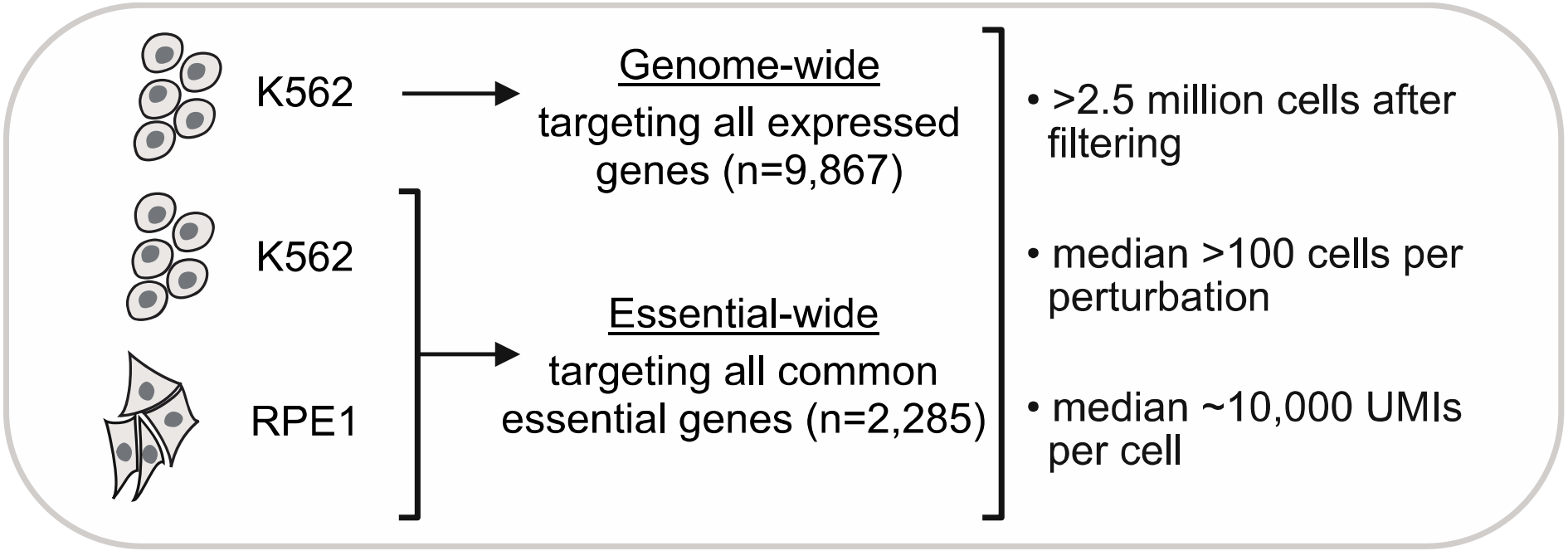
Reuben Saunders

Jonathan Weissman

https://doi.org/10.1016/j.cell.2022.05.013

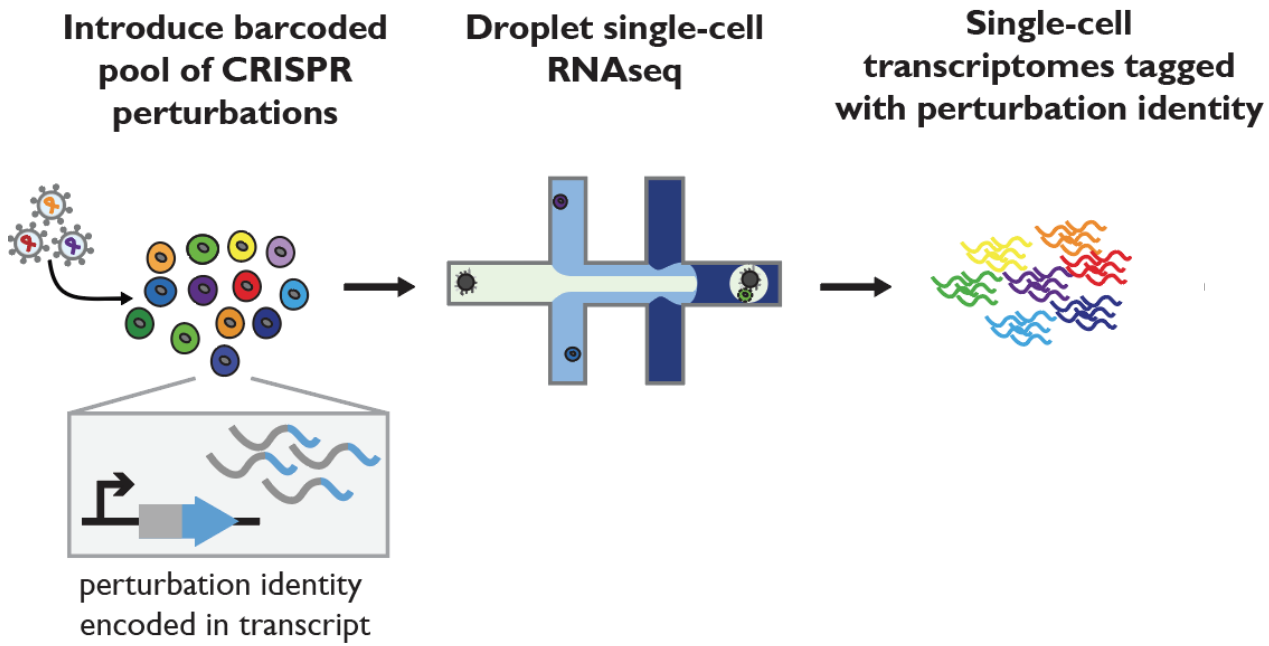
# Genome-scale Perturb-seq to understand cell biology

- Until this study Perturb-seq had always been applied to targeted gene sets (usually the hits from other screens)
- We wanted to see what happened when it was applied in unbiased fashion. In the largest dataset, we knockdown every expressed transcript using CRISPRi.



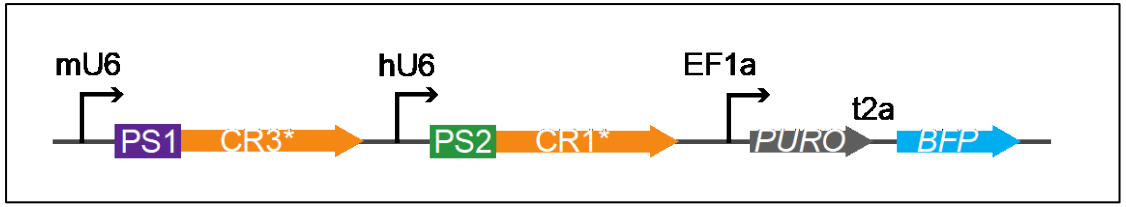
# Genome-scale Perturb-seq

- **Model:** **K562 cells (womp womp)**
- **Perturbations:** **CRISPRi knockdown (dual sgRNA library) of 9,867 transcripts**
- **Challenge:** outgrowth
- **Readout:** **single-cell RNA sequencing**

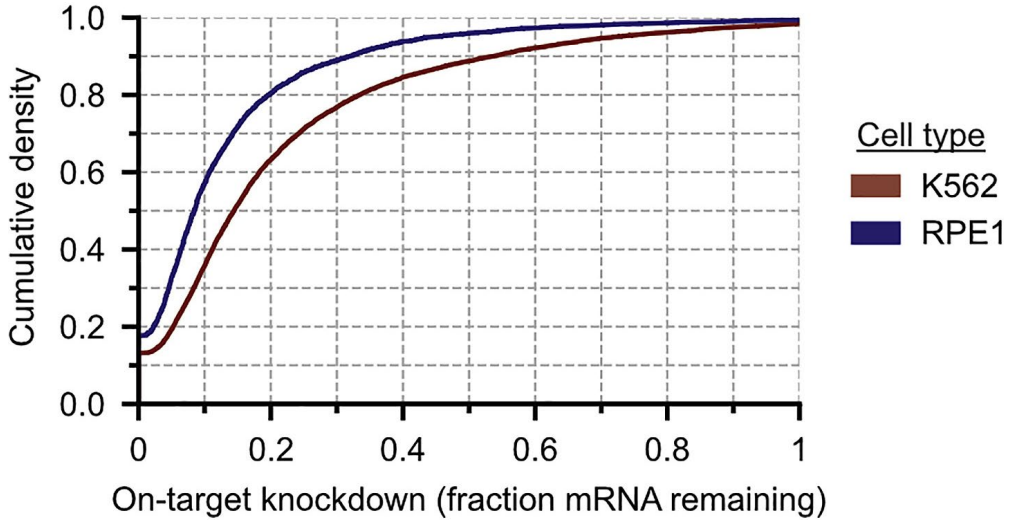


# Optimized CRISPRi yields median ~90% knockdown

- We use a dual sgRNA library to improve the chances of knocking down target genes

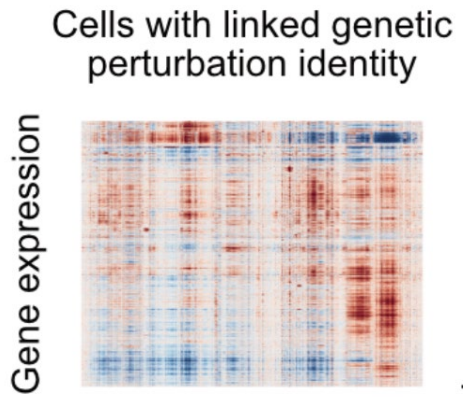



- We developed an optimized ZIM3-dCas9 CRISPRi effector that we used in later experiments
- An advantage of CRISPRi in single-cell contexts is that you can assess on- and off-target activity
- Optimized effector leads to median 90% knockdown of targeted genes



# A global view of the data structure

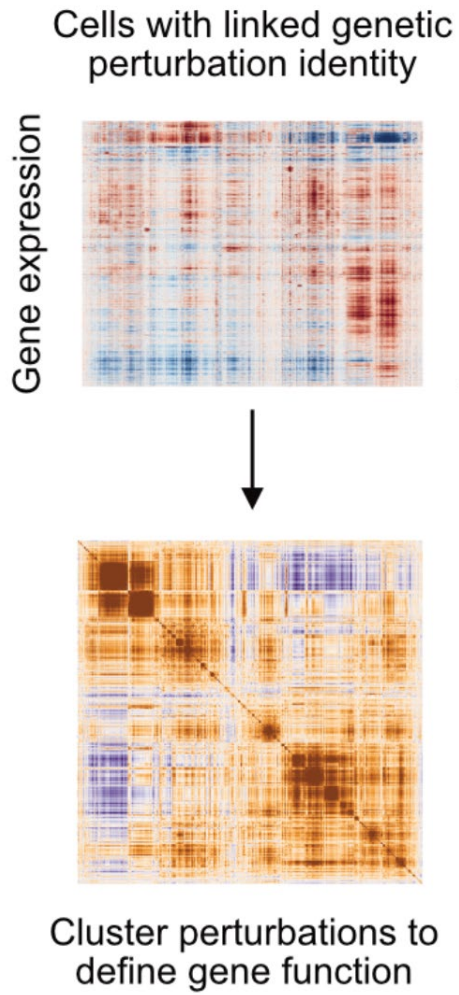
- Most analysis is done at the pseudo-bulk level—average single-cell RNAseq profiles of cells with same perturbation (sgRNA) together
- Experiment is overall a large matrix of gene expression × perturbations



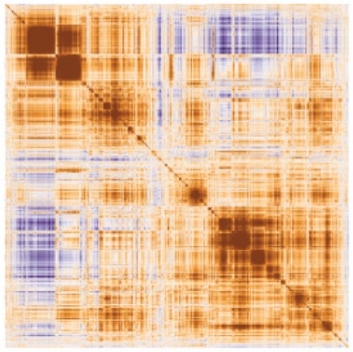
# Annotating gene function from transcriptional phenotypes

- Most analysis is done at the pseudo-bulk level—average single-cell RNAseq profiles of cells with same perturbation (sgRNA) together
- Experiment is overall a large matrix of gene expression × perturbations

- To assign function to perturbations, we can cluster this matrix based on gene expression profile correlations
- We can visualize these similarities using a nonlinear embedding that places correlated profiles close to each other



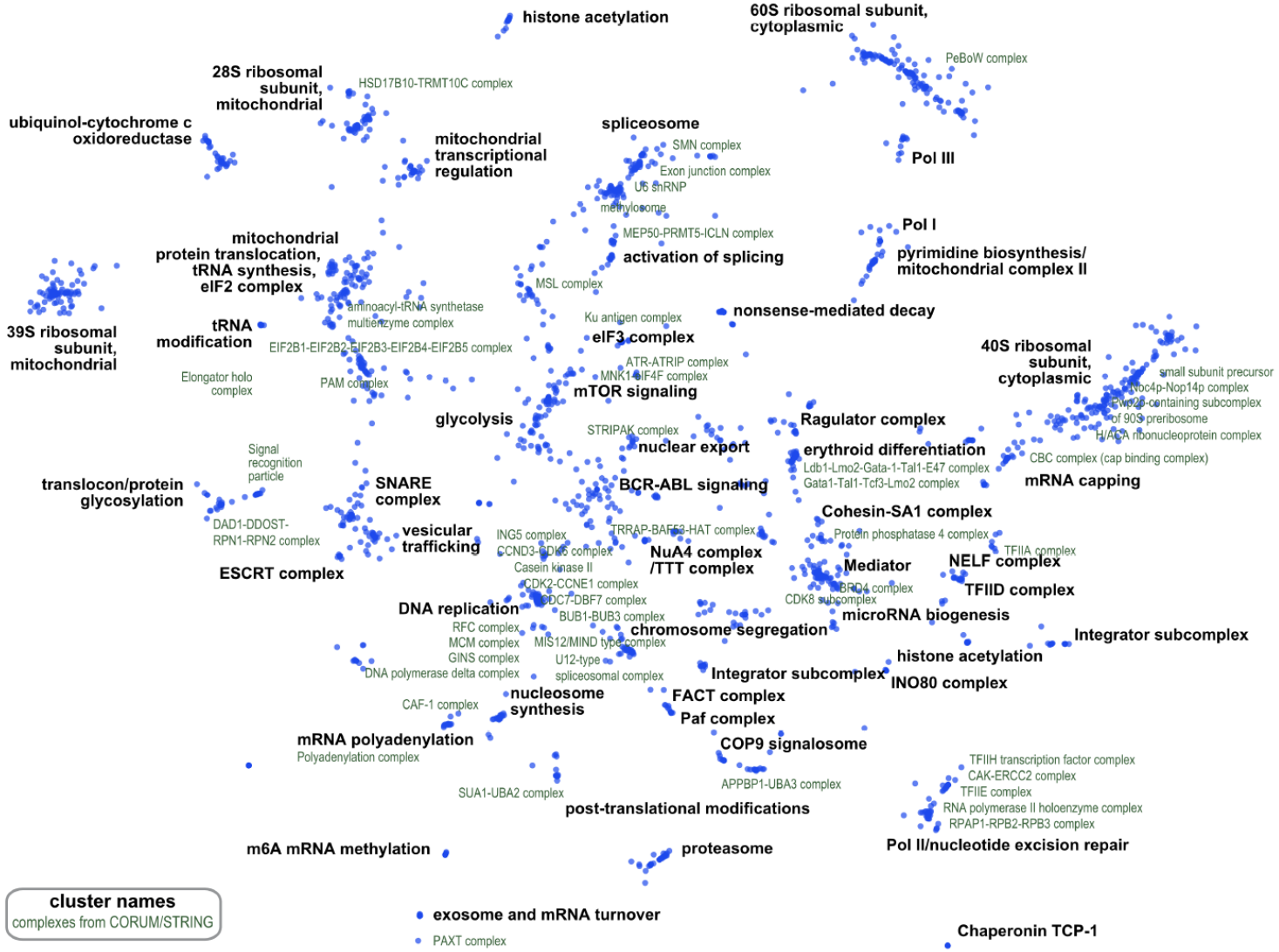
# Annotating gene function from transcriptional phenotypes



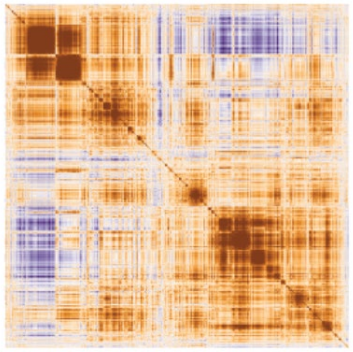
Cluster perturbations to define gene function

Embed perturbations so that similar profiles are nearby

Annotate manually and using databases



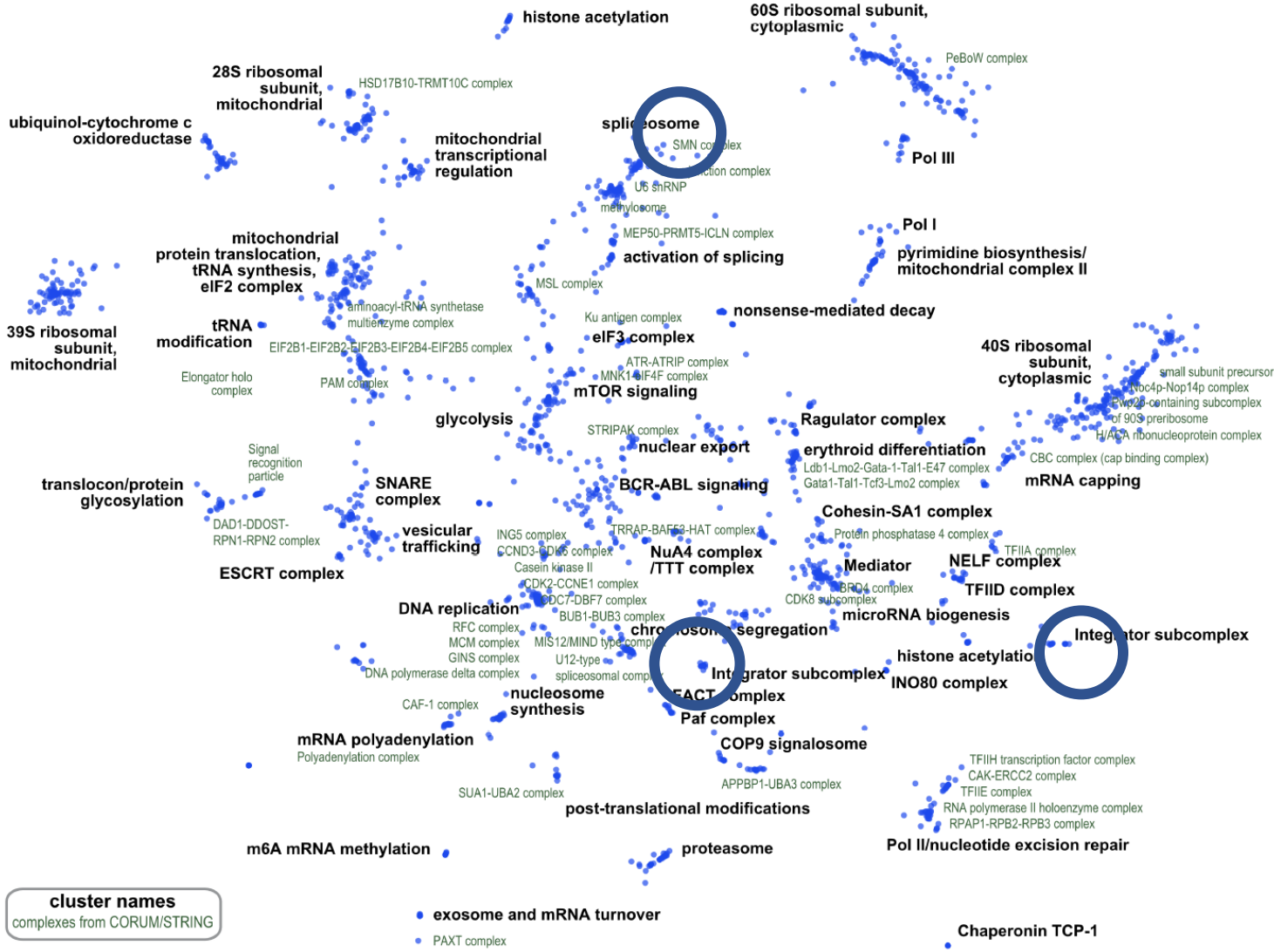
# Annotating gene function from transcriptional phenotypes



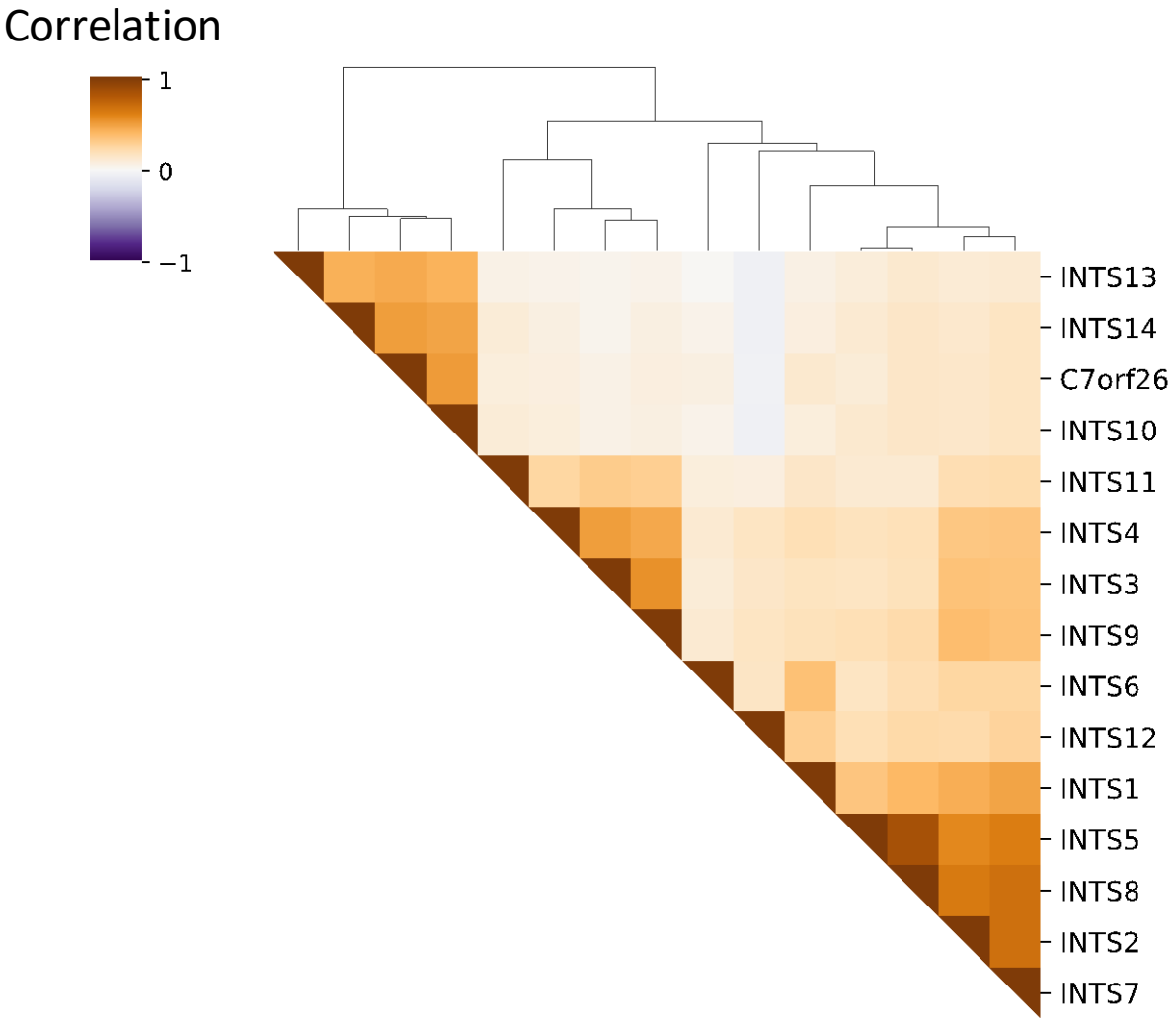
Cluster perturbations to define gene function

Embed perturbations so that similar profiles are nearby

Annotate manually and using databases

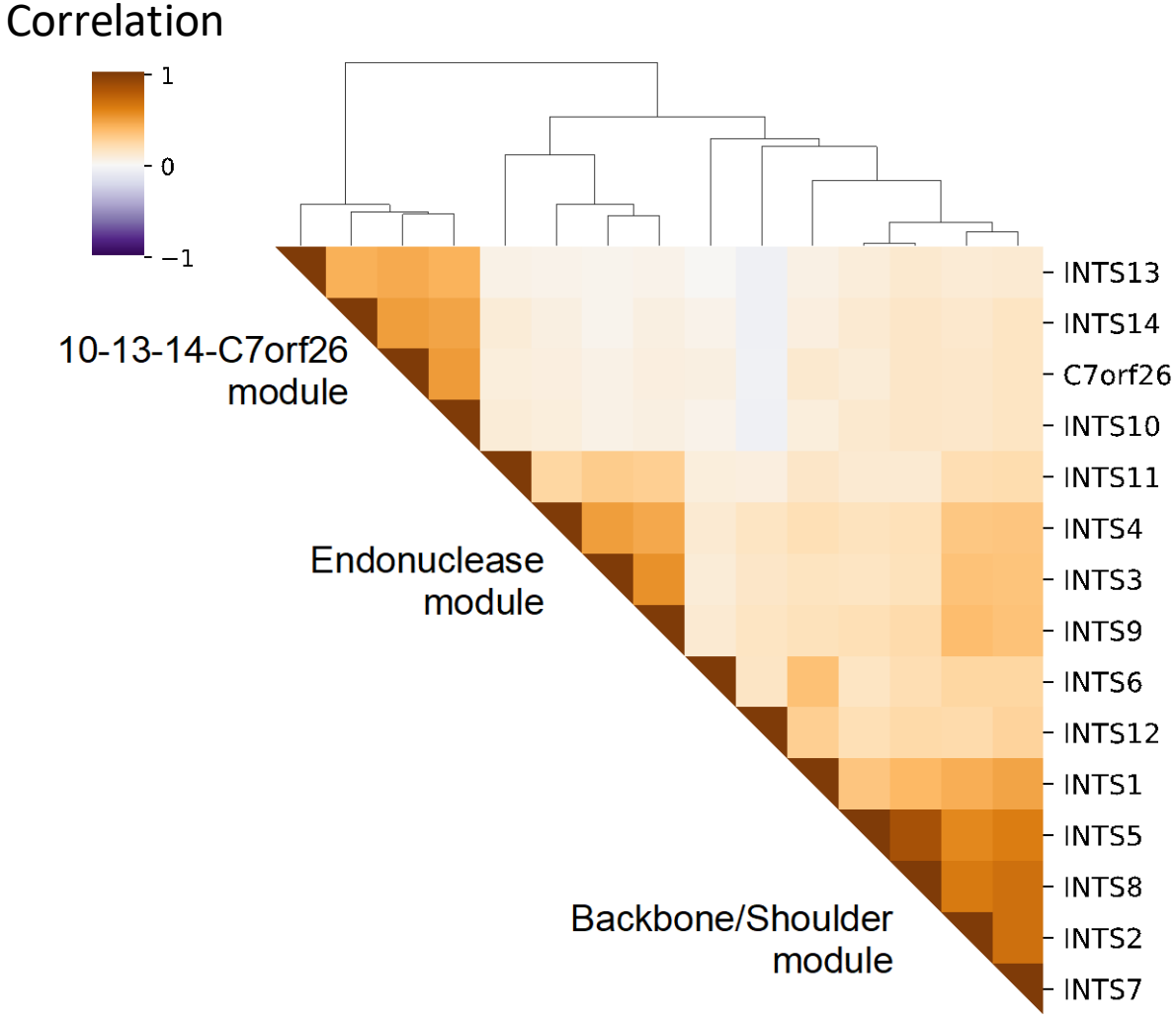


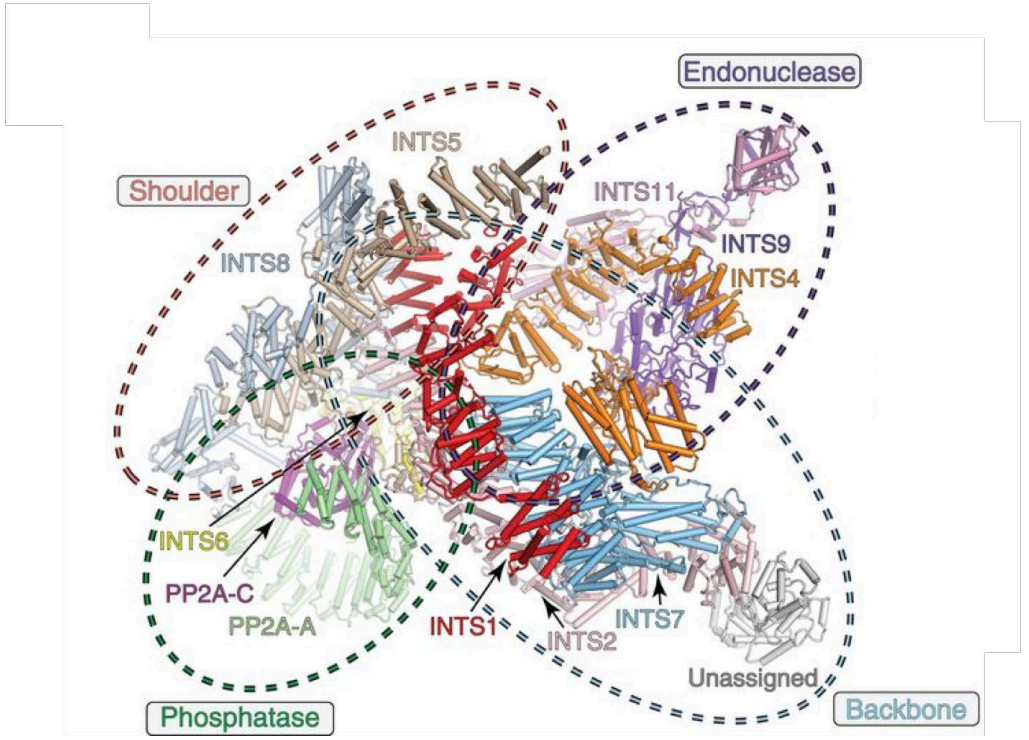
# Defining functional submodules through transcriptional phenotypes




- Integrator complex is a global regulator of transcription:
  - Processes enhancer transcripts and some small nuclear RNAs into their mature forms by cleavage
  - Regulates transcription of protein coding genes by controlling transcription termination
- Functions of most subunits are unknown

# Comparing structural and functional submodules

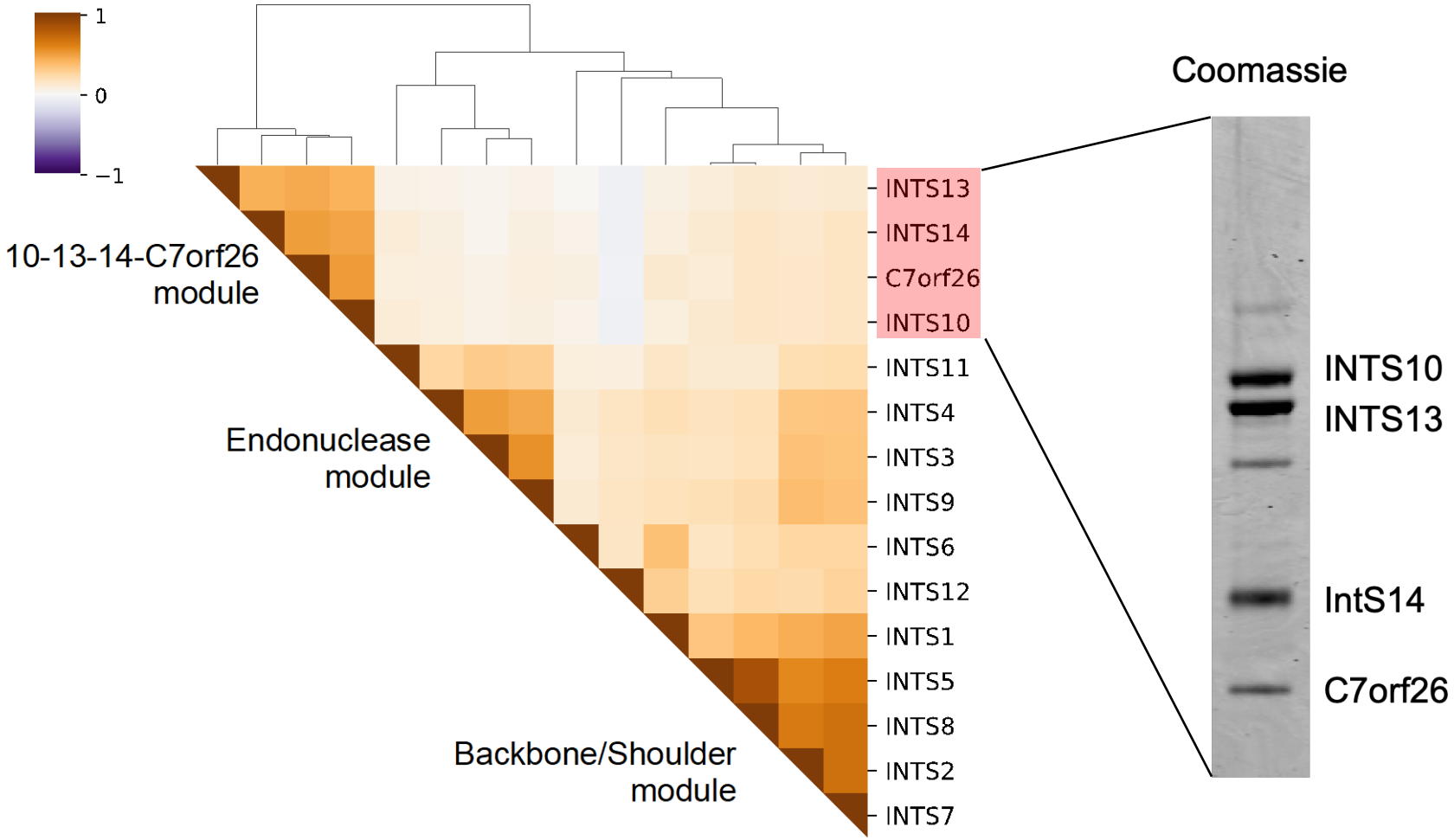



Zheng *et al*, 2020

IntS10,13,14 were present in prep but not resolved in CryoEM maps

# C7orf26, which we rename INTS15, is a core component of the INTS10-13-14-15 subcomplex

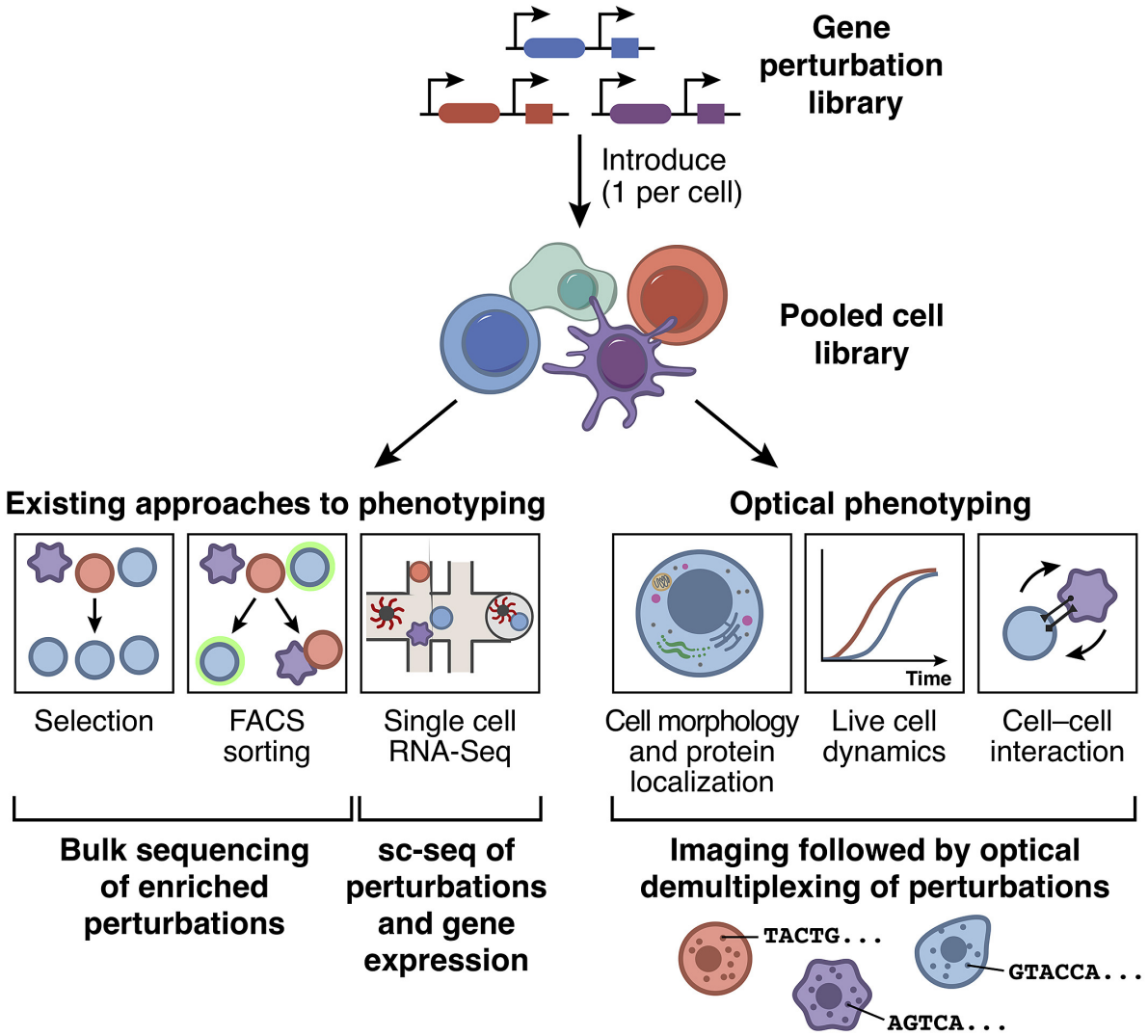


# Outline

1. Functional genomics background
   - Current CRISPR tools
   - The four elements of a CRISPR screen
   - How is single-cell RNA sequencing performed?
2. High-content readouts
   - Development of Perturb-seq
   - Using genome-scale Perturb-seq to assign gene function
   - **Using imaging CRISPR screens to assign gene function**
   - *In vivo* screens and adding lineage information
3. High complexity perturbations
   - Genetic interaction mapping
   - Highly multiplexed studies of enhancers

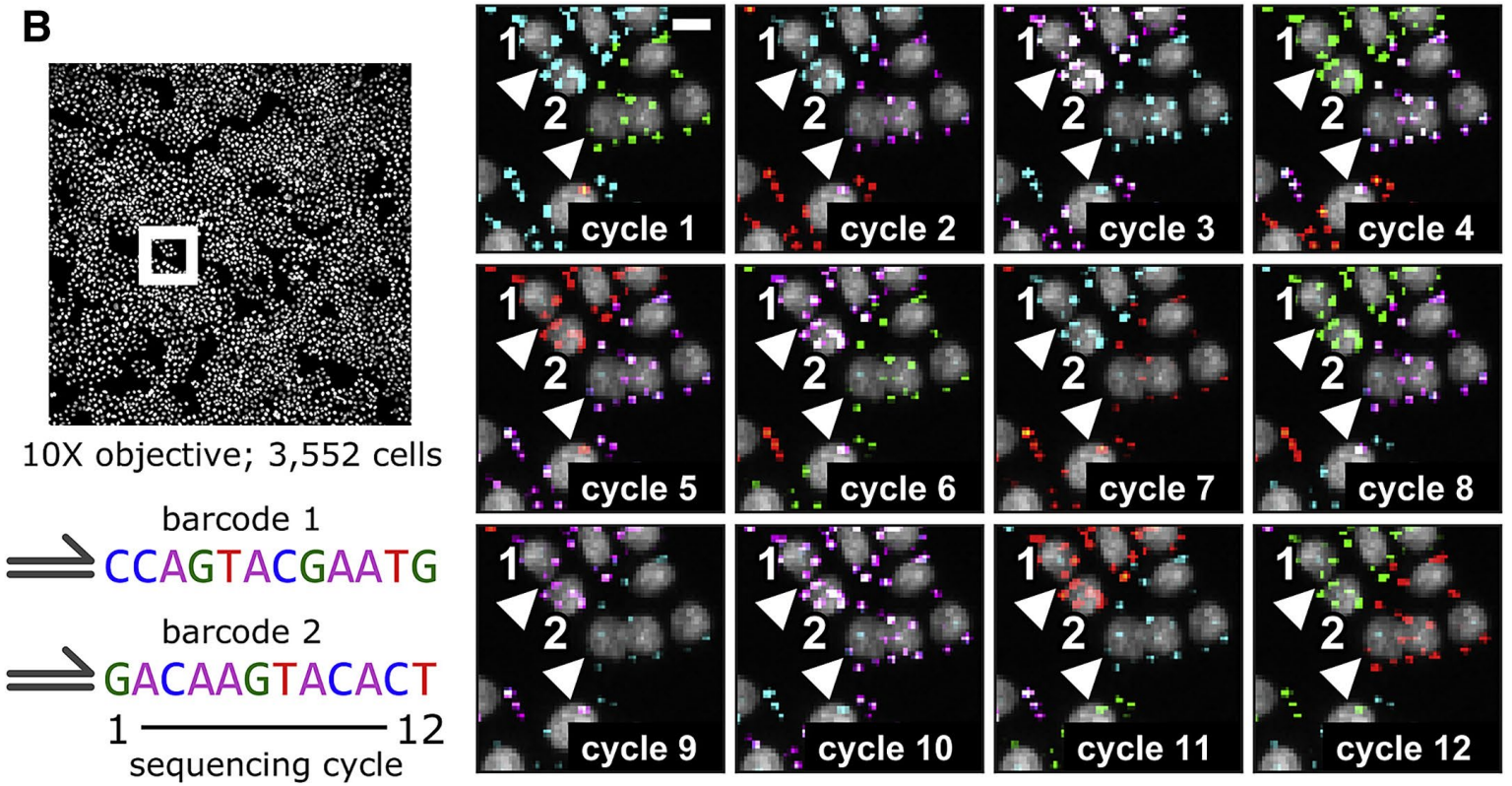
# Optical pooled screening

- An alternative to Perturb-seq: introduce genetic perturbations using CRISPR, but use cell morphology as a high-content phenotype instead of the transcriptome



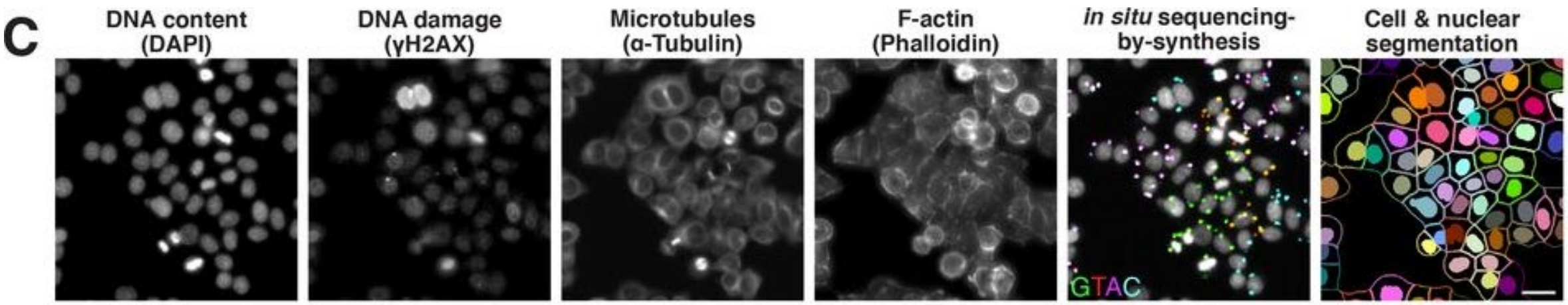
# Reading out the guide by *in situ* sequencing

- Cas9 target is encoded in protospacer part of guide – only 20 nts
- An excellent target for *in situ* sequencing, which is still quite finicky



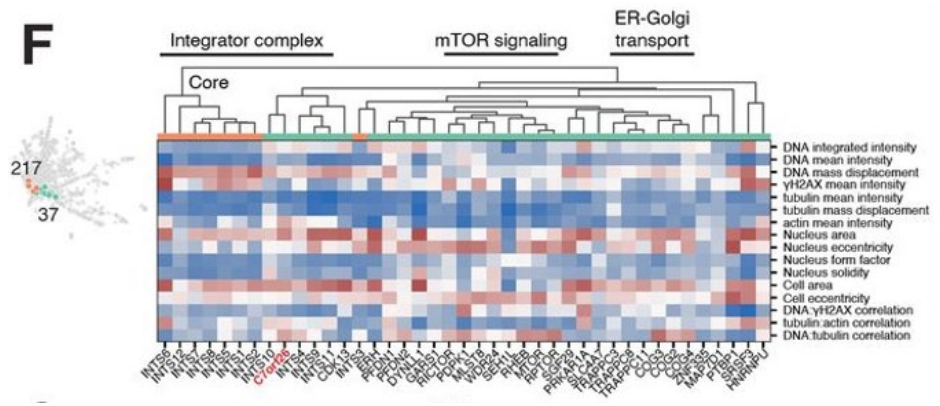
# Screening cell morphology in response to knockouts of essential genes

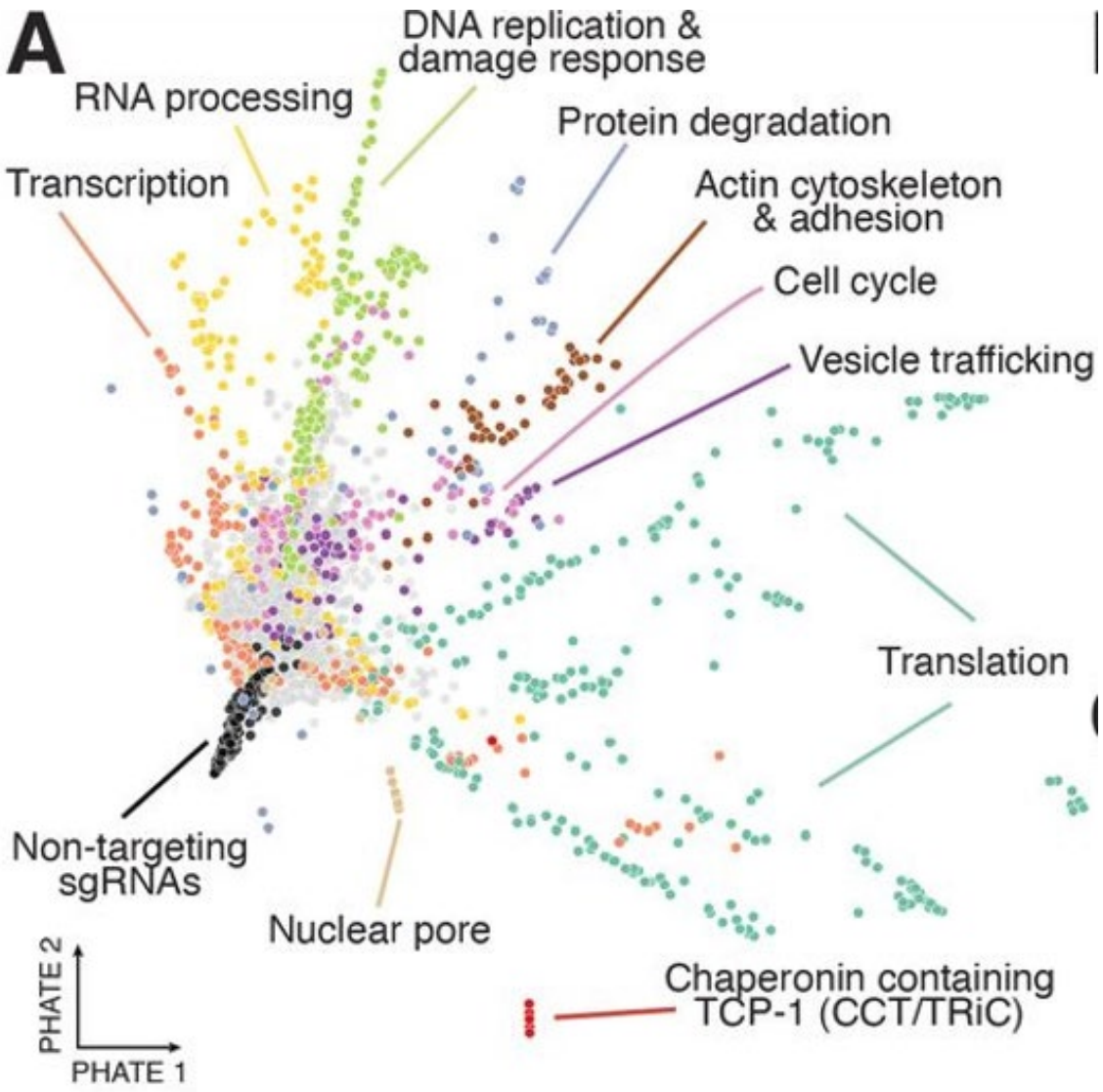
- Cells stained for:
  - DNA (DAPI)
  - DNA damage γH2AX (anti-phospho-Ser139 H2AX antibody)
  - microtubules (anti-α-tubulin antibody)
  - filamentous actin (phalloidin)
- Total 31,884,270 individual cells with a median of 6,119 cells per gene target across each set of four sgRNAs (i.e. ~1500 per sgRNA)!



# Imaging phenotypes can cluster genes by function

- For each cell, extract 1084 morphological features using CellProfiler and Python image processing
- Use the PHATE dimensionality reduction algorithm to produce a visualization of how genes relate
- Amusingly, they identify the same gene of unknown function that we did: C7orf26 is a member of the Integrator complex



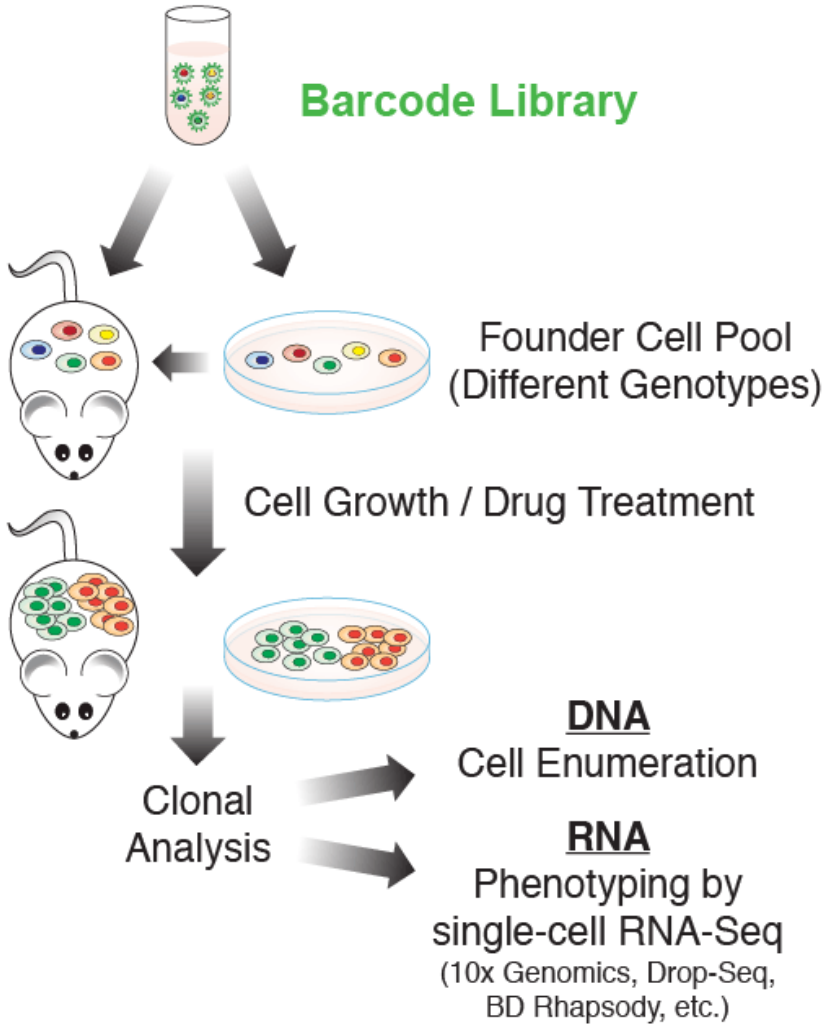



# Outline

1. Functional genomics background
   - Current CRISPR tools
   - The four elements of a CRISPR screen
   - How is single-cell RNA sequencing performed?
2. High-content readouts
   - Development of Perturb-seq
   - Using genome-scale Perturb-seq to assign gene function
   - Using imaging CRISPR screens to assign gene function
   - ***In vivo* screens and adding lineage information**
3. High complexity perturbations
   - Genetic interaction mapping
   - Highly multiplexed studies of enhancers
   - Opportunities for machine learning

# *In vivo* CRISPR screens

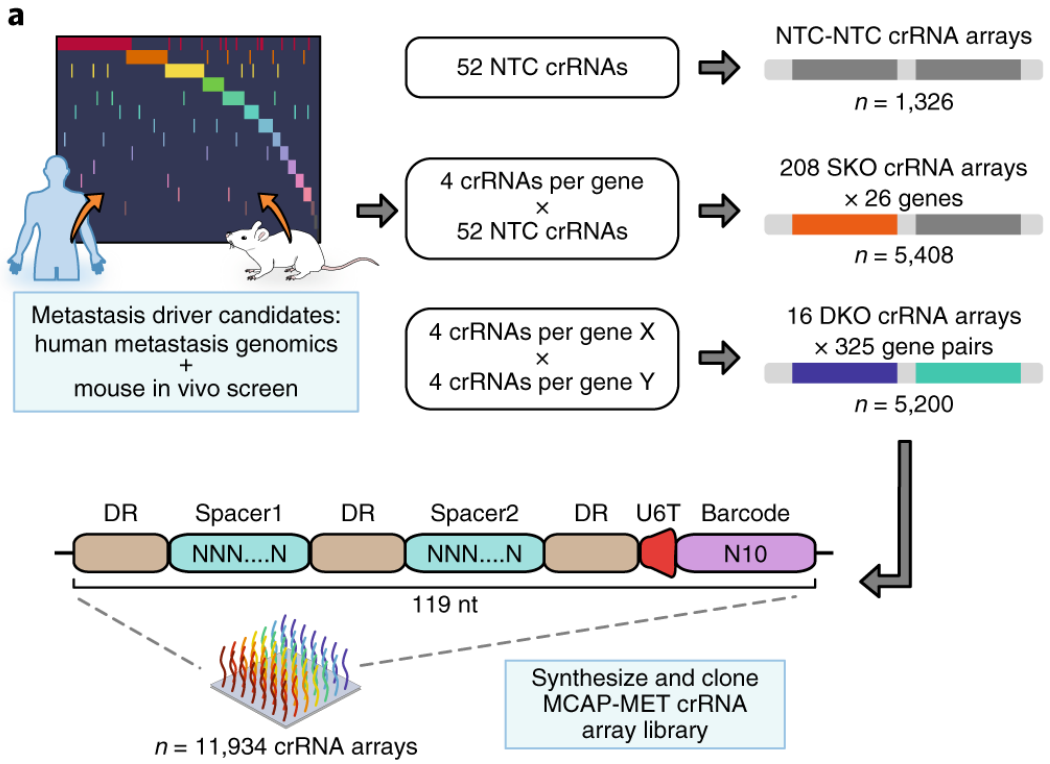
- A current frontier, and interesting for obvious reasons
- In many tissues, number of cells that are feasible to infect or engraft is limited → limit number of perturbations to avoid bottlenecking
- Some cells may reside in environments that are more or less permissible, leading to strong clonal effects → use clone barcodes and replicates
- Cells are likely not actively growing and dividing, so different target proteins may deplete on very different time scales because of variability in turnover rates




https://manuals.cellecta.com/clonetracker-xp-lentiviral-barcode-libraries/v1b/en/topic/clonetracker-xp-lentiviral-barcode-libraries

# *In vivo* CRISPR screens

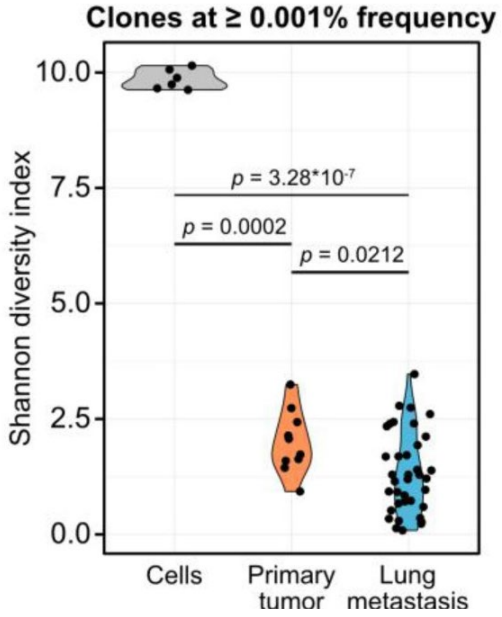
- Use a NSCLC cell line called KPD (*Kras*$^{G12D/+}$;*p53*$^{-/-}$;*Dicer1*$^{+/-}$), which is not naturally metastatic. Had previously conducted screens to identify gene knockouts that were pro-metastatic in this background.
- Wanted to study whether there was synergy among these hit genes
- Used Cas12a (aka Cpf1), which has ability to self-process arrays of guides. Have two targeting positions for each gene in the pair, plus a third position containing a random barcode used to follow clones.

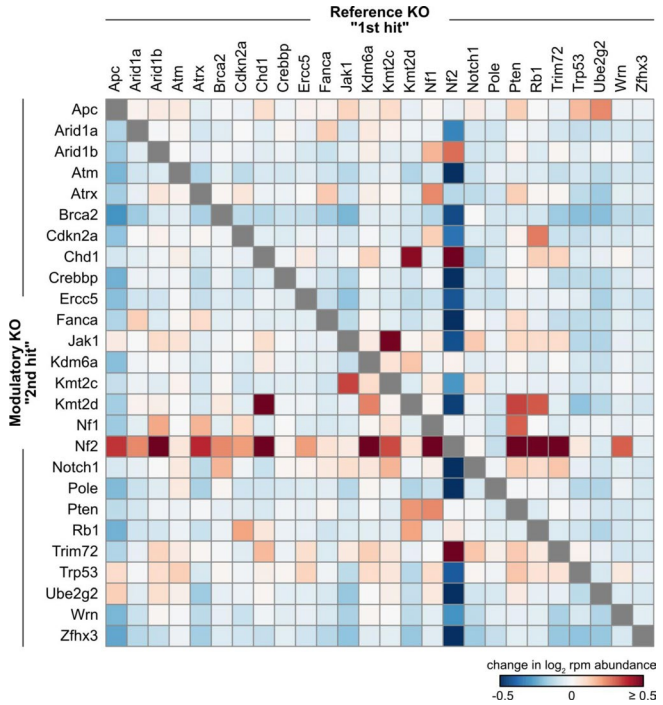



https://doi.org/10.1038/s41592-019-0371-5

# *In vivo* CRISPR screens

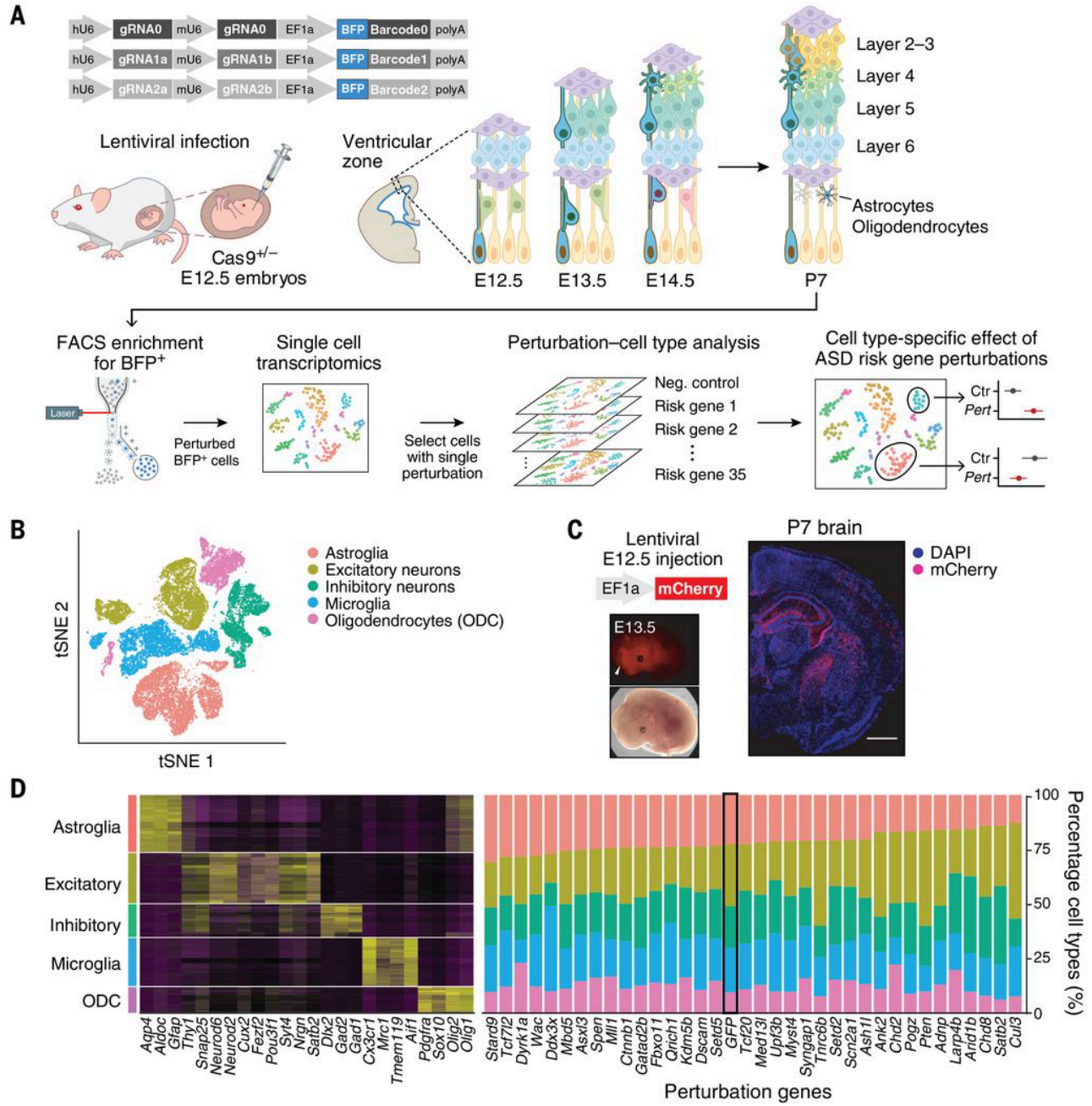
- Inject 4 million cells subcutaneously into 10 replicate mice. Using their clone barcodes, can infer this gives 350× coverage of each element.
- Challenges:
  - Cas12a is not as good at cutting as Cas9 – many library elements inactive
  - Weak activity + strong selection means only positive selection screens are feasible. Intrinsically less reproducible.
- See strong evidence of progressive selection of clones from injected cell pool, to primary tumor, to lung metastasis
- Also see evidence for synergy in enhancement of metastasis, particularly with *Nf2/Trim72*



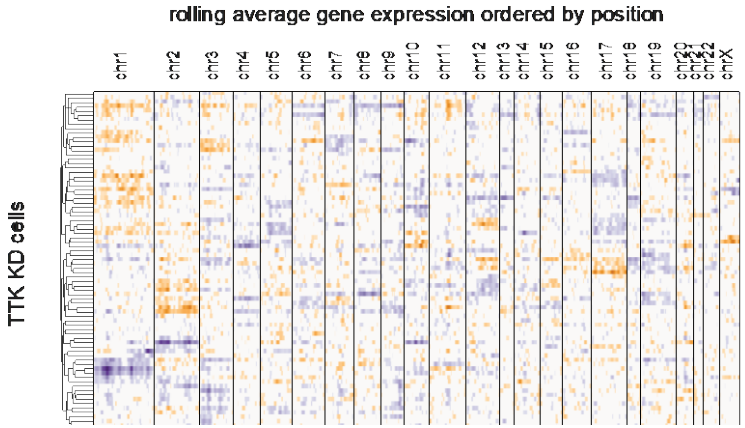
# *In vivo* Perturb-seq

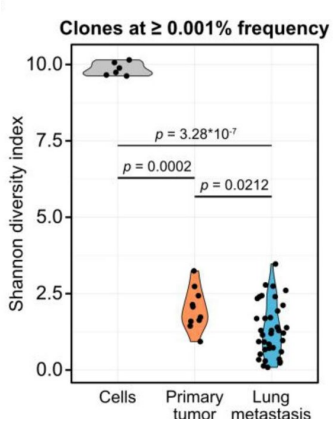
- GWAS studies have identified risk alleles for diseases such as autism, but phenotypes likely manifest through action in specific cell types
- Targeted 35 genes in the mouse brain. Mouse expresses Cas9, and guides were introduced to E12.5 embryos by injection. Used two guides per gene to increase efficacy.
- Guides could be introduced into several different cell types, enabling analyses of whether guides altered cell type composition. (They did not.)

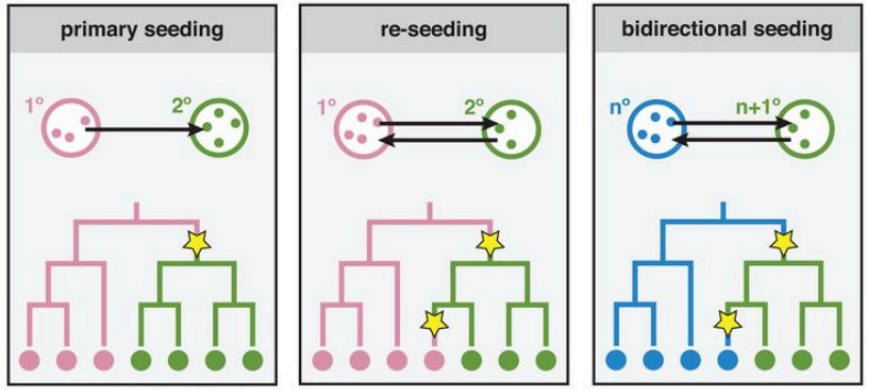



https://doi.org/10.1126/science.aaz6063

# Impact of lineage on screens



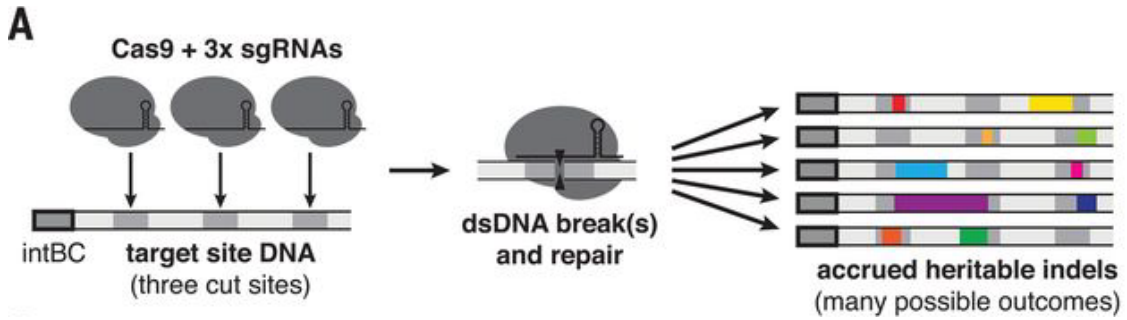



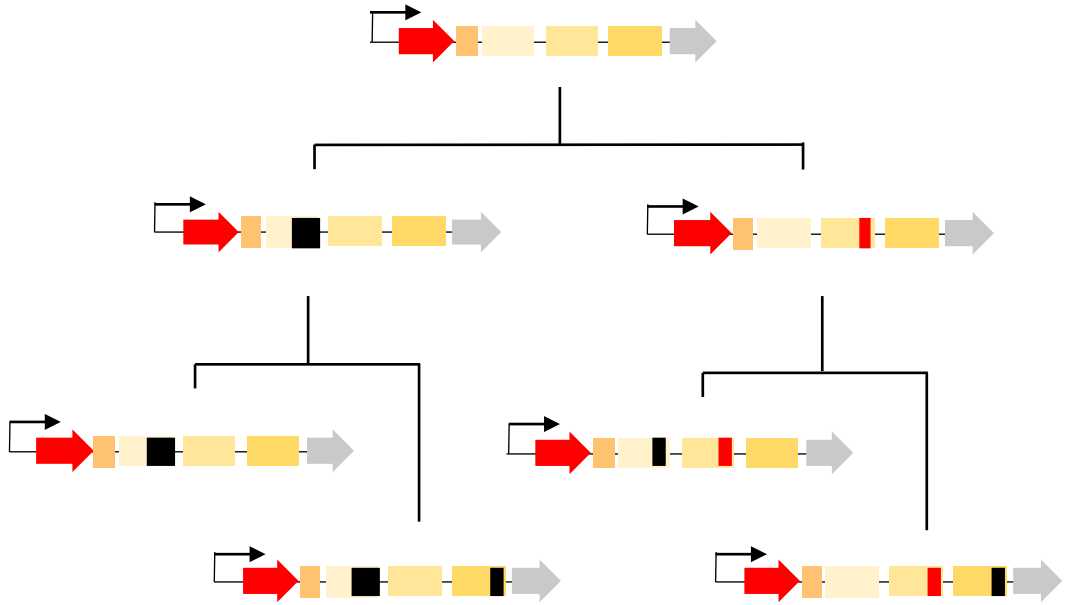

Clonal expansion is an important, sometimes underacknowledged, confounder in CRISPR screens

Clonal expansion may be an important driver of phenotype in development or dynamic processes such as metastasis

# Evolving barcodes using lineage tracers

- Much interest in evolving barcoding systems: these leverage the variable indels formed by Cas9 cutting to create diverse barcodes over time
- E.g. One design has 3 synthetic “target sites” per construct, each designed to have variable cutting rates (slow to fast) and highly variable indels (more labels)
- Integrate many copies of this into target cells using PiggyBac
- Expressed in 3’ UTR of mCherry, so capturable using same approach as original Perturb-seq

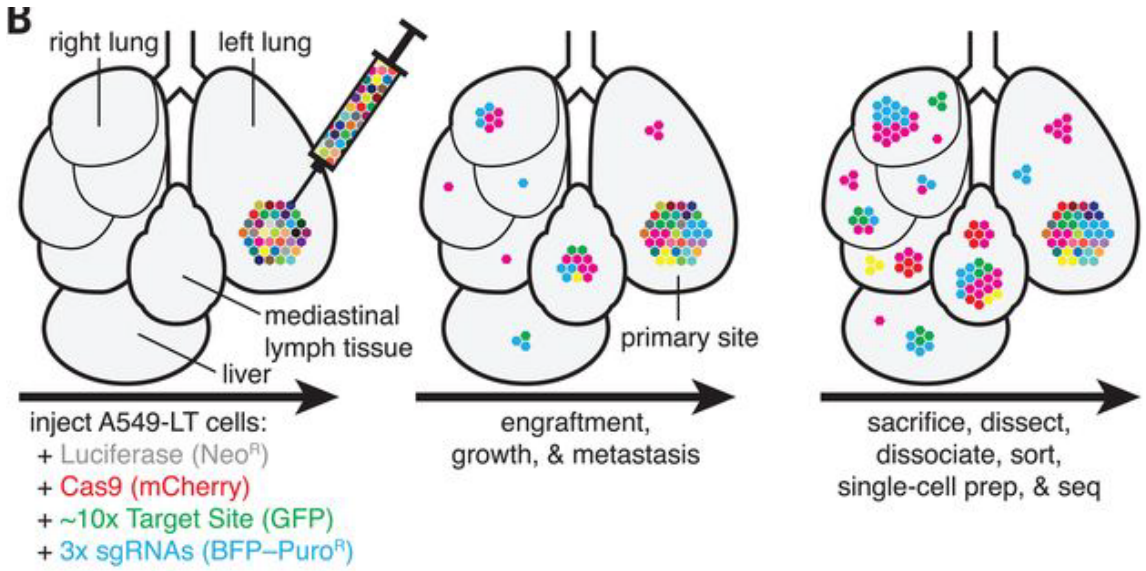


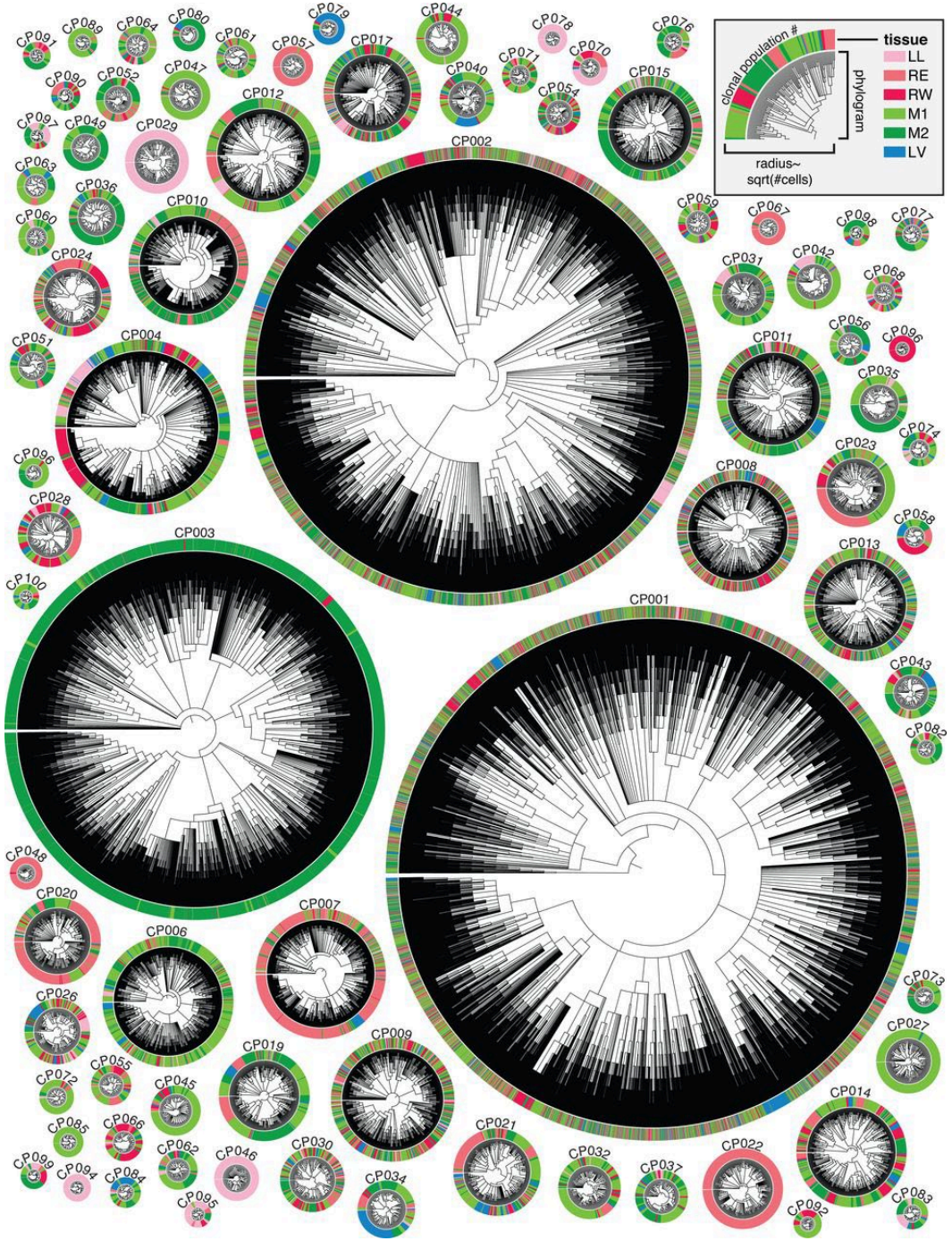



DOI: 10.1126/science.abc1944

# Tracing lineage in a model of metastasis

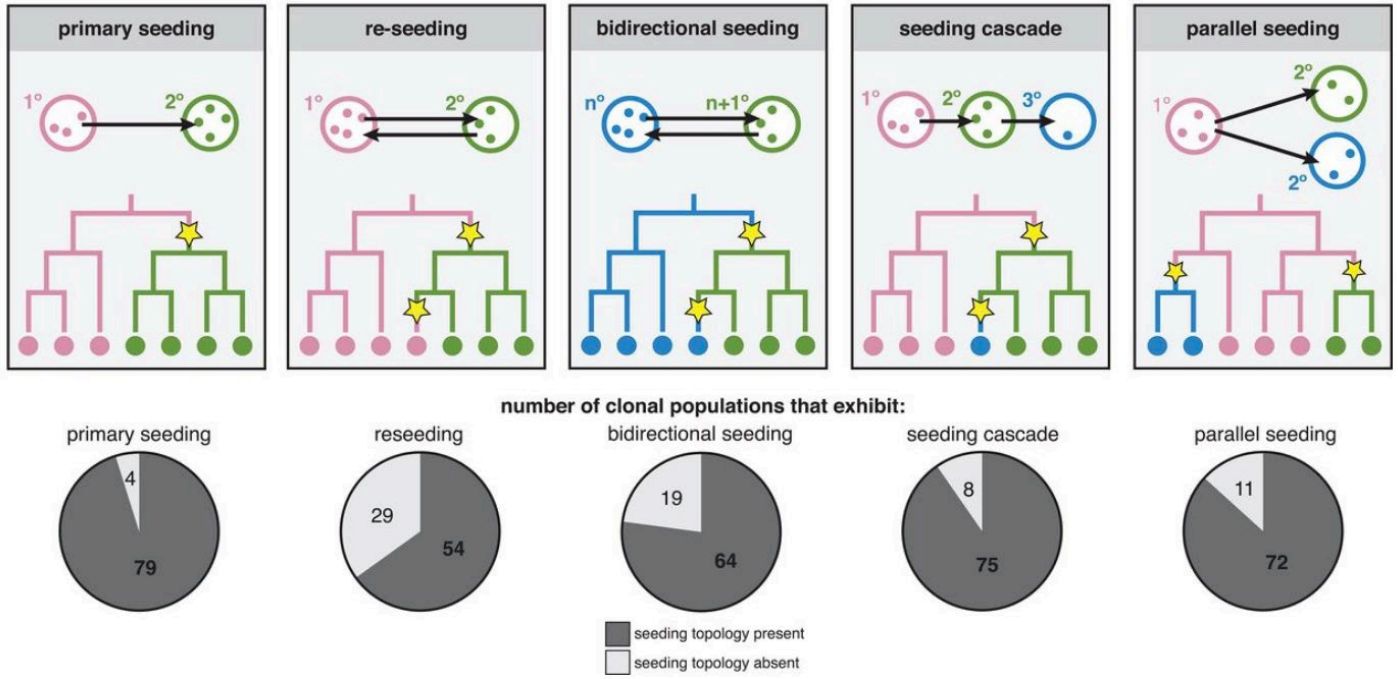
- Use A549 cells (human *KRAS*-mutant lung adenocarcinoma), implant into left lung, monitor metastases to other parts of the body
- Using evolving barcode, can reconstruct phylogeny of how cells relate to each other. E.g. Many clones primarily metastasized via the mediastinal lymph tissue, indicating it is a favorable niche.



# Tracing lineage in a model of metastasis

- Phylogeny can detect and quantify five different modes of metastasis that are happening concurrently in the mouse.
- As noted earlier, this example highlights how important it is to remember the evolutionary dynamics that are occurring within screens
- Combining lineage tracing with CRISPR-mediated perturbations is a major current direction



# Outline

1. Functional genomics background
   - Current CRISPR tools
   - The four elements of a CRISPR screen
   - How is single-cell RNA sequencing performed?
2. High-content readouts
   - Development of Perturb-seq
   - Using genome-scale Perturb-seq to assign gene function
   - Using imaging CRISPR screens to assign gene function
   - *In vivo* screens and adding lineage information
3. **High complexity perturbations**
   - **Genetic interaction mapping**
   - **Highly multiplexed studies of enhancers**

# Looking beyond genome-scale

1. **Genetic interactions**
   - Instead of looking at single gene perturbations, look at how genes work together to realize phenotype
   - Grow combinatorially: among 1000 genes there are $\binom{1000}{2} = 499{,}500$ possible pairwise interactions
   - We already saw some instances of genetic interactions when discussing *in vivo* screens, and they are a major target of the DepMap

2. **Regulatory elements**
   - Millions of candidate regulatory elements in the genome

Both areas are the target of many current efforts

# Genetic interactions (GIs)

Genes work together to generate emergent phenotypes:

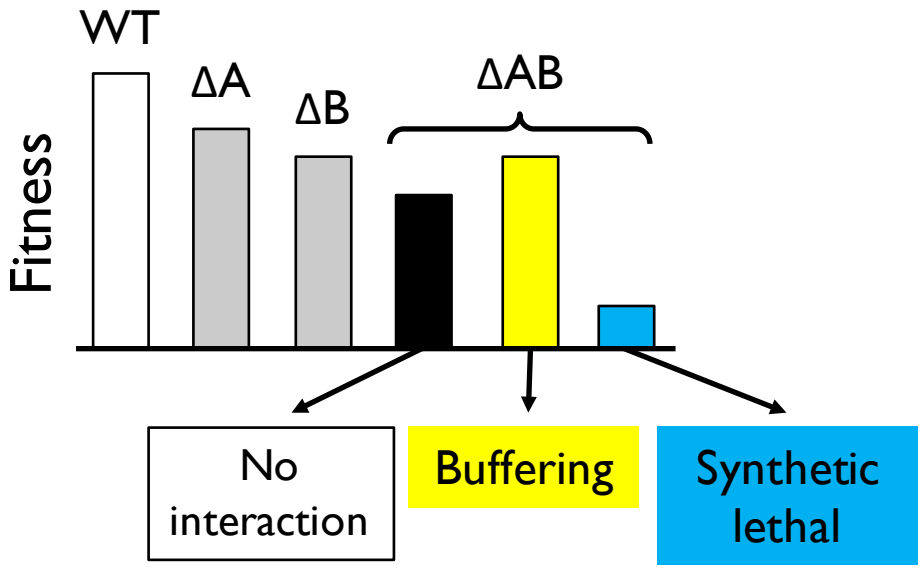
- **Development:** transcription factors define cell type
- **Cancer:** synthetic lethal interactions
- **Genetic disease:** modifier genes alter severity

**Core task:** Does something “surprising” happen when I perturb genes together vs. separately?

# A fitness measure of genetic interactions

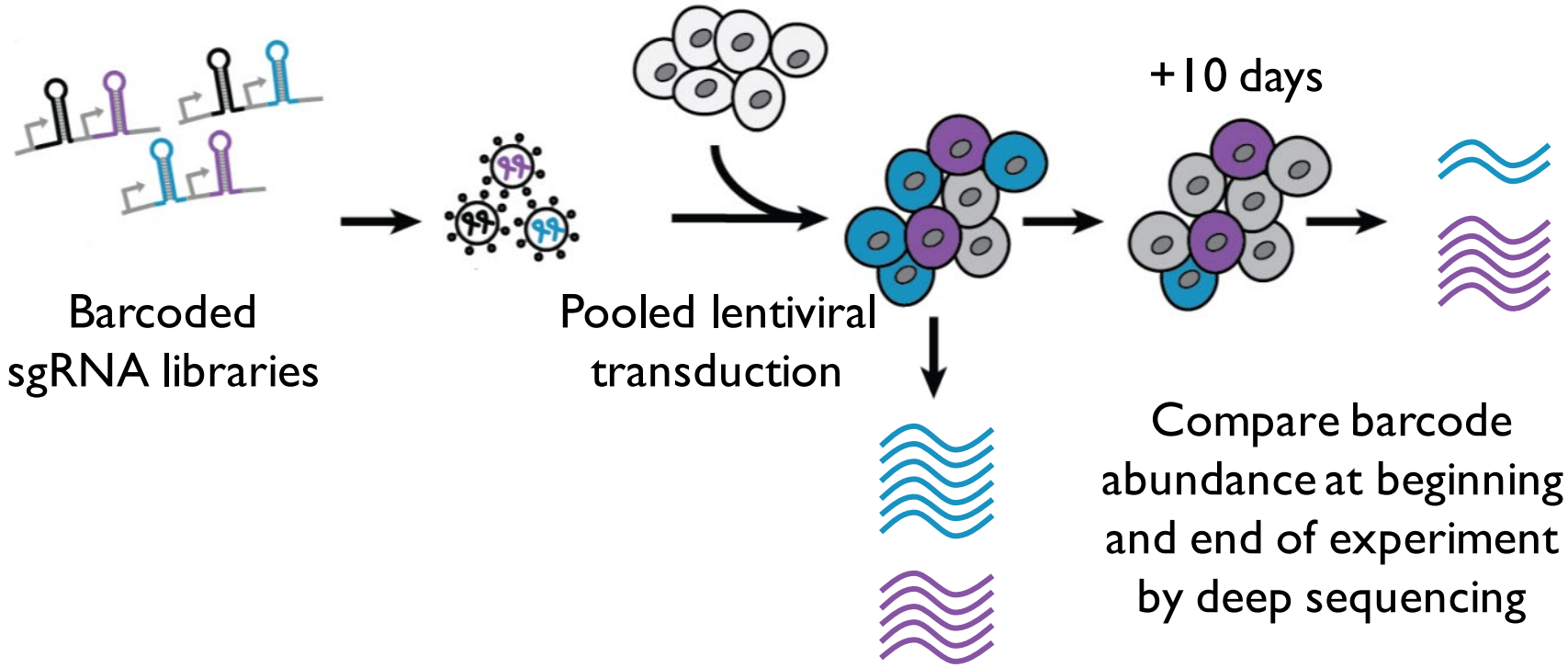
**Genetic interaction (epistasis)**: when the phenotype of the double mutant deviates from the expectation given by the single mutants

- Really measures how "surprised" we are by the outcoming of combining two effects
- Weaker than expected = buffering
- Stronger than expected = synthetic lethal



# CRISPR genetic interaction mapping

- Fairly trivial to multiplex CRISPR experiments – simply deliver two sgRNAs (or more) in each lentivirus and proceed as usual



# CRISPR genetic interaction mapping

Gene B

Fix Gene A and measure fitness
with all possible partner Genes B

Gene A

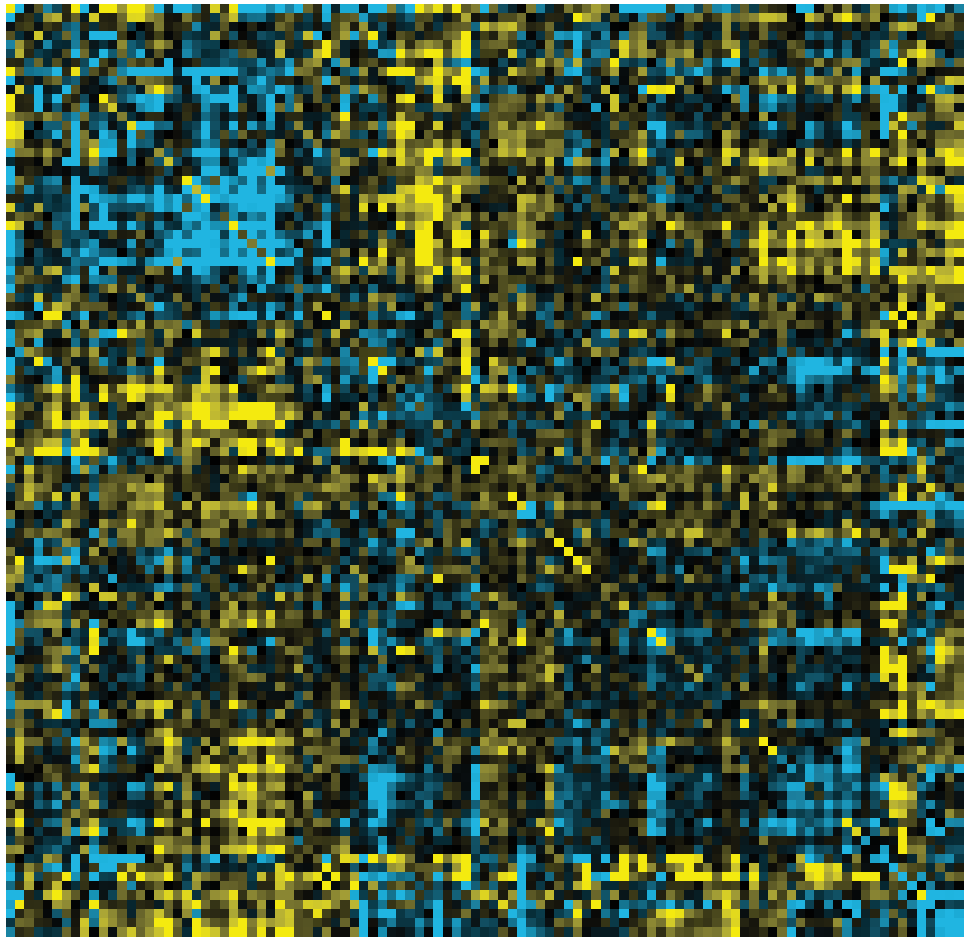
Buffering

Synthetic
lethal

# CRISPR genetic interaction mapping

Gene B

Gene A



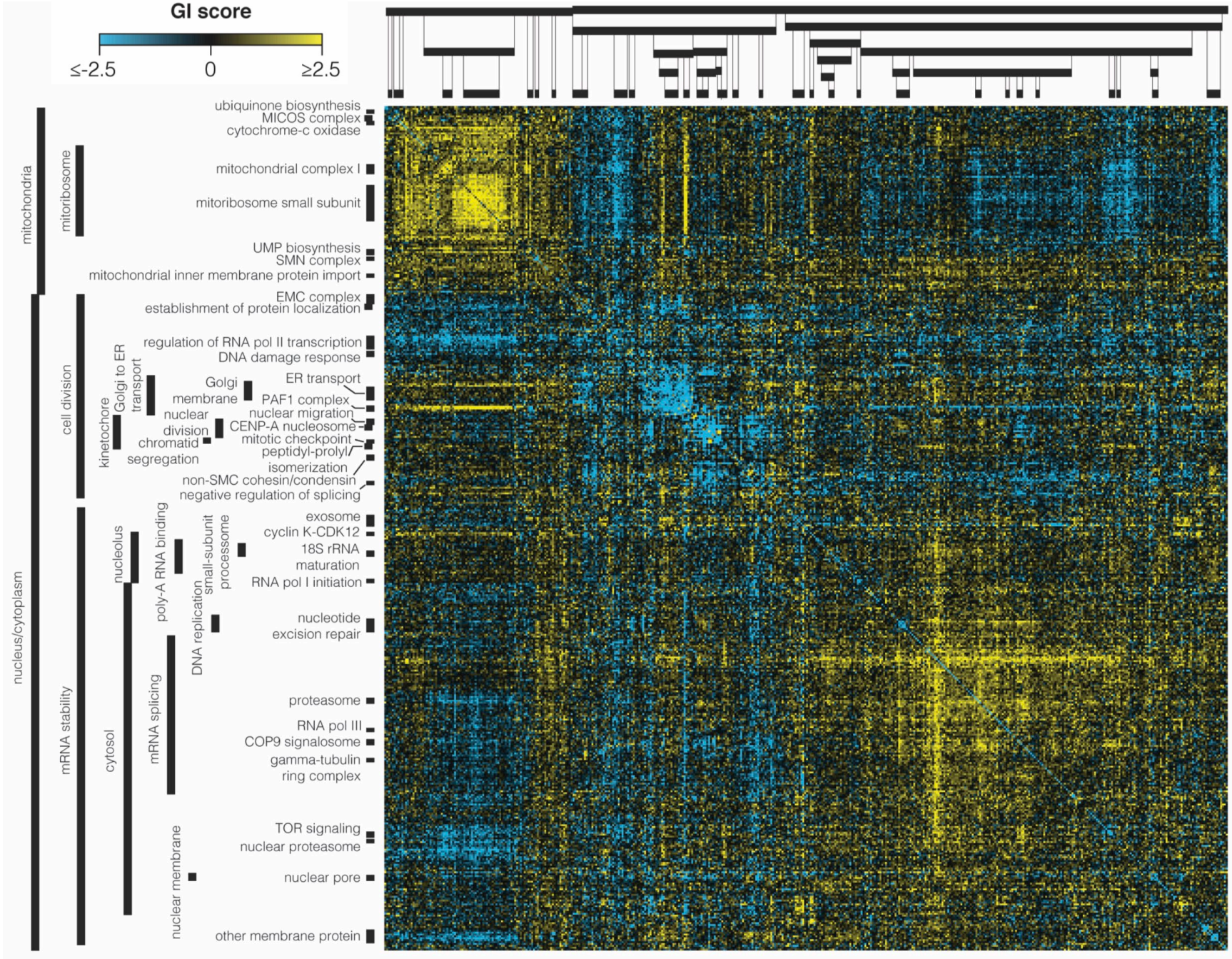
Buffering

Synthetic lethal

Similar interactions
=
Similar function

## Example: CRISPRi GI Map

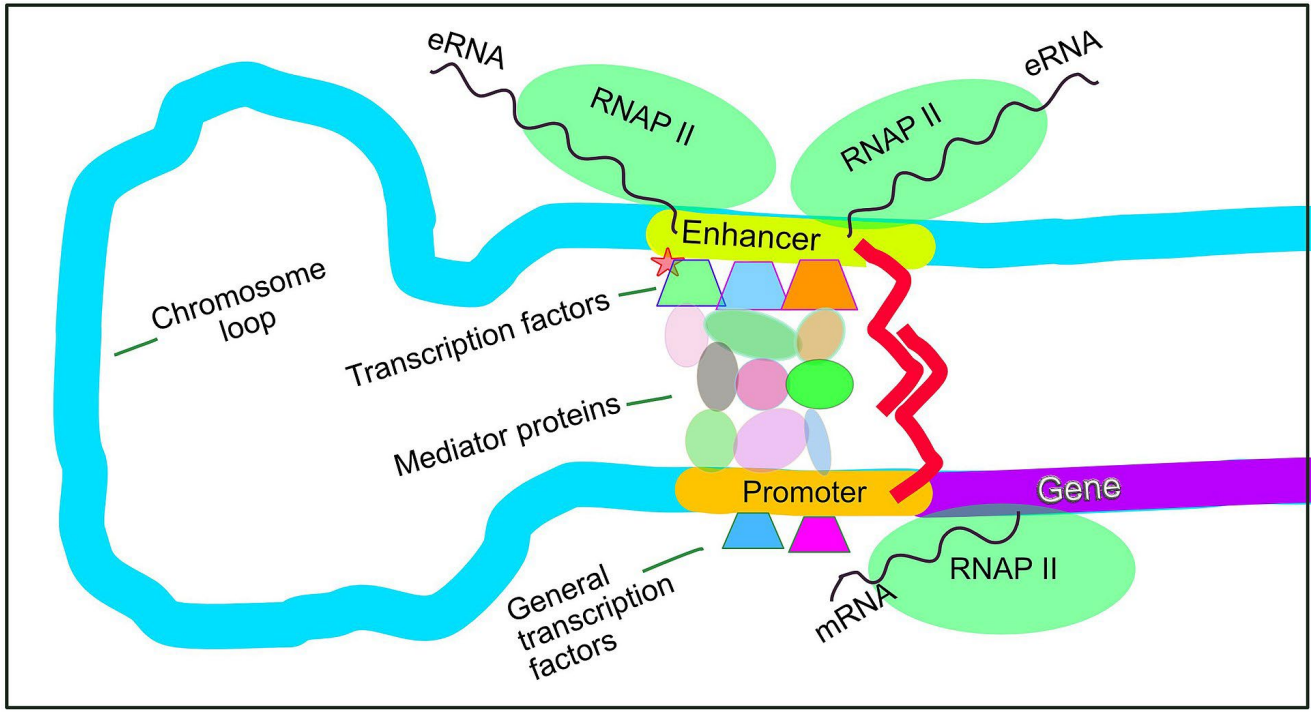
- Largest measurement of genetic interactions in mammalian system: ~500 × 500 genes, ~1 million sgRNA pairs
- A way fitness can be used for high-content phenotyping
- Unlike the DepMap, requires only a single cell line, and combinations are programmed and so more interpretable




Horlbeck et al. *Cell* 2018

# Studying enhancers

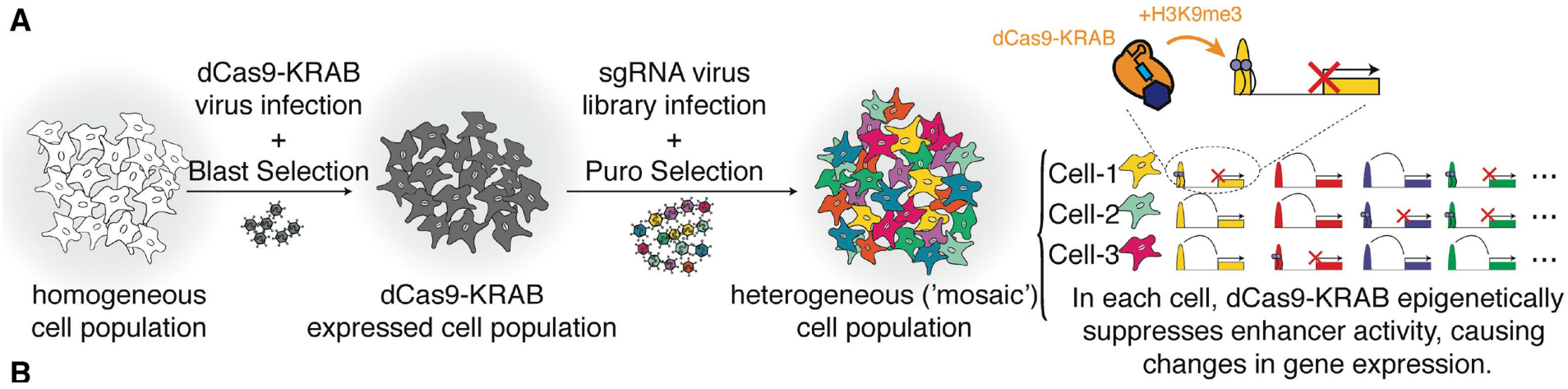
- ENCODE (Encyclopedia of DNA Elements) identified ~400,000 putative enhancers
- **Challenges:**
  - Many enhancers have weak or redundant effects, complicating analysis
  - Diverse mechanisms of action and non-coding nature make them harder to identify and study than protein-coding genes
  - Interactions between enhancers and their target genes can occur over long distances




https://en.wikipedia.org/wiki/Enhancer_%28genetics%29

# Probing enhancer function by scRNA-seq

- Mosaic-seq is like Perturb-seq but targeting enhancers instead of promoters




- Can tie an enhancer to the genes it regulates by looking at gene expression changes in its vicinity
- **Problem:** scaling number of enhancers queried is expensive

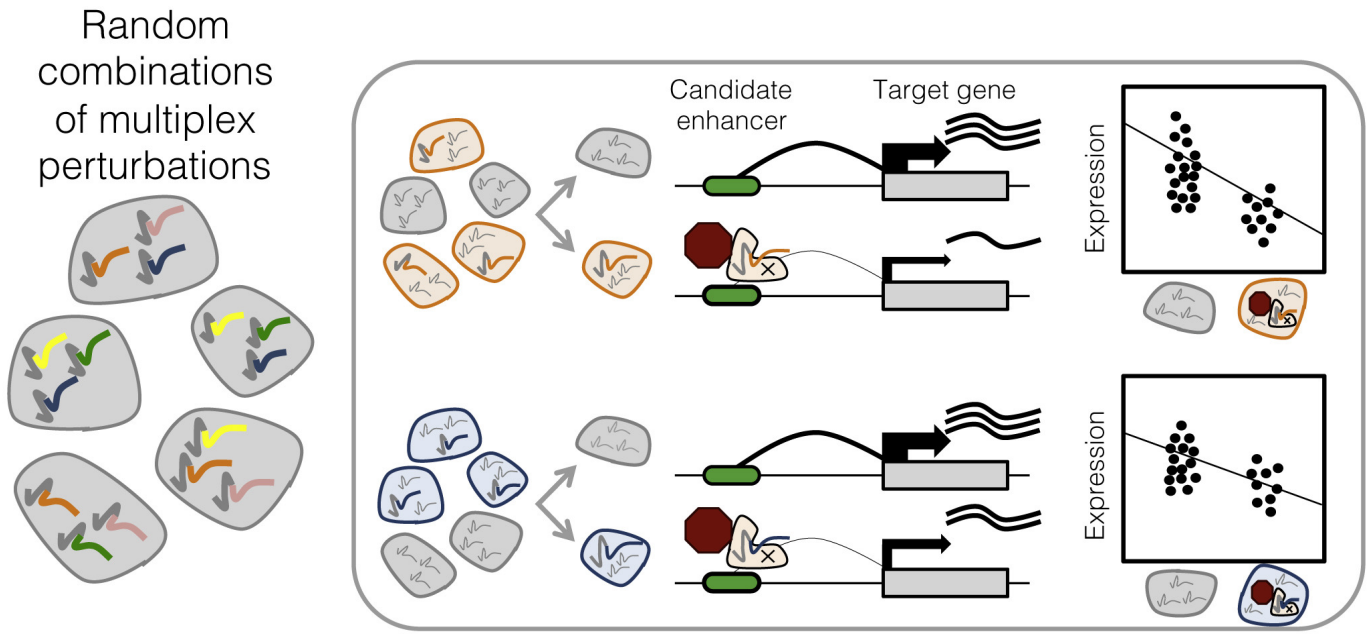
doi.org/10.1016/j.molcel.2017.03.007

# Overcoming cost by multiplexing

- There are many enhancers to test:
  - Many are expected to have weak or absent phenotypes
  - Enhancers are expected to act locally, meaning interactions or far-reaching effects are unlikely (though not impossible!)

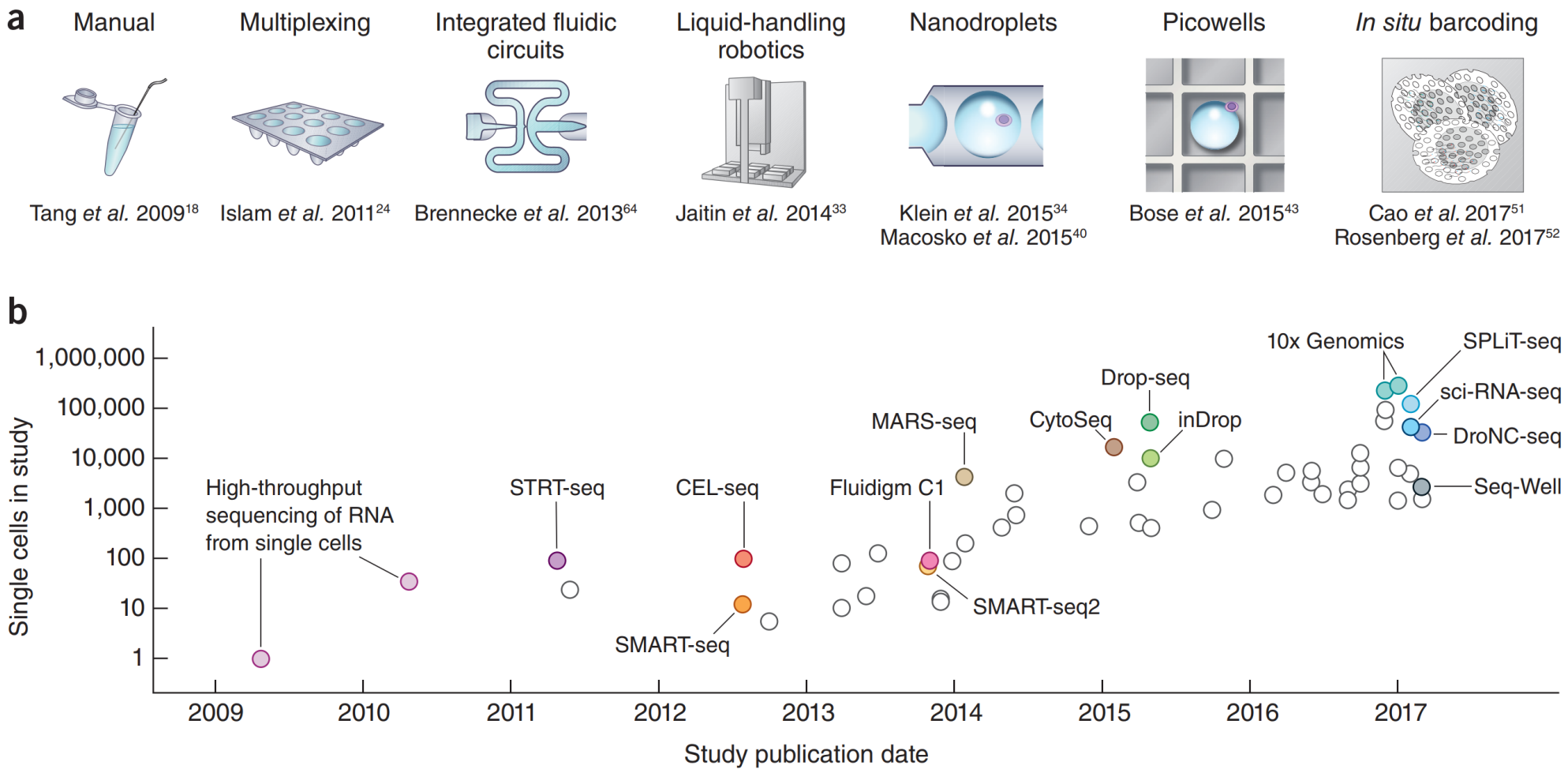
# Overcoming cost by multiplexing

- One idea: superinfect cells with lentivirus such that each cell contains a random mixture of 15 CRISPRi perturbations on average
- Data are amenable to computational “demultiplexing” by a regression model. Same cell can be used in 15 different analyses on average, letting them test ~6,000 enhancers in a single screen




doi.org/10.1016/j.cell.2018.11.029

# An exciting frontier to think about




doi.org/10.1038/nprot.2017.149

# Outline

1. Functional genomics background
   - Current CRISPR tools
   - The four elements of a CRISPR screen
   - How is single-cell RNA sequencing performed?
2. High-content readouts
   - Development of Perturb-seq
   - Using genome-scale Perturb-seq to assign gene function
   - Using imaging CRISPR screens to assign gene function
   - *In vivo* screens and adding lineage information
3. High complexity perturbations
   - Genetic interaction mapping
   - Highly multiplexed studies of enhancers