



Single-Cell Sequencing in Molecular Recording and Natural Time Tracking

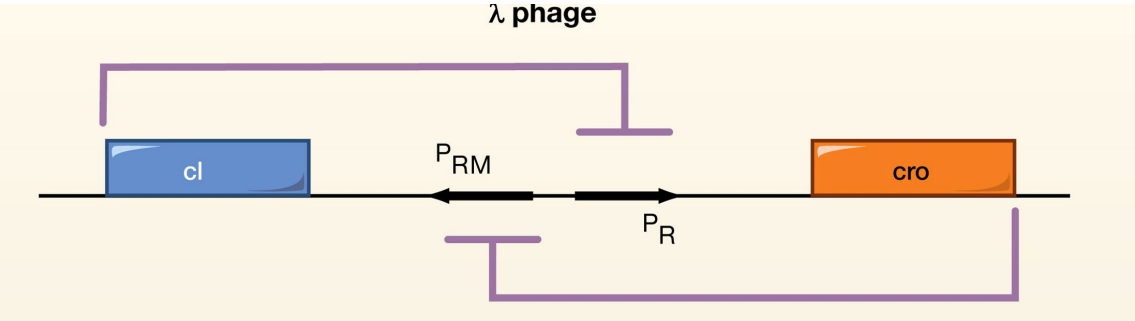
2 0 2 5

Group1: Lewei He

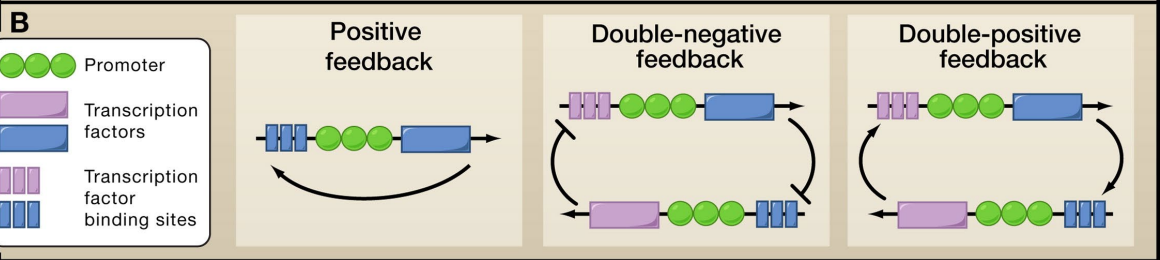
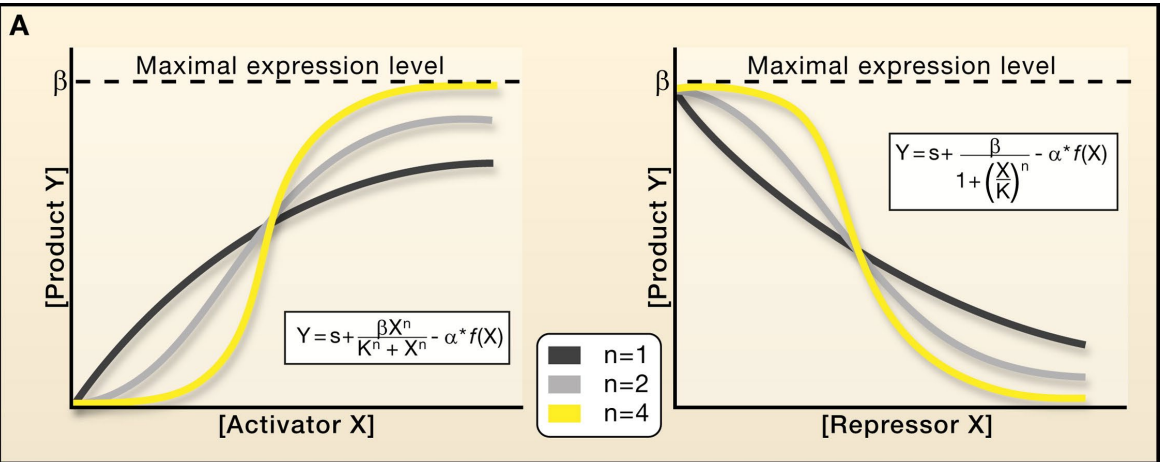
Making Cellular Memories

Biological memory can be defined as a sustained cellular response to a transient stimulus

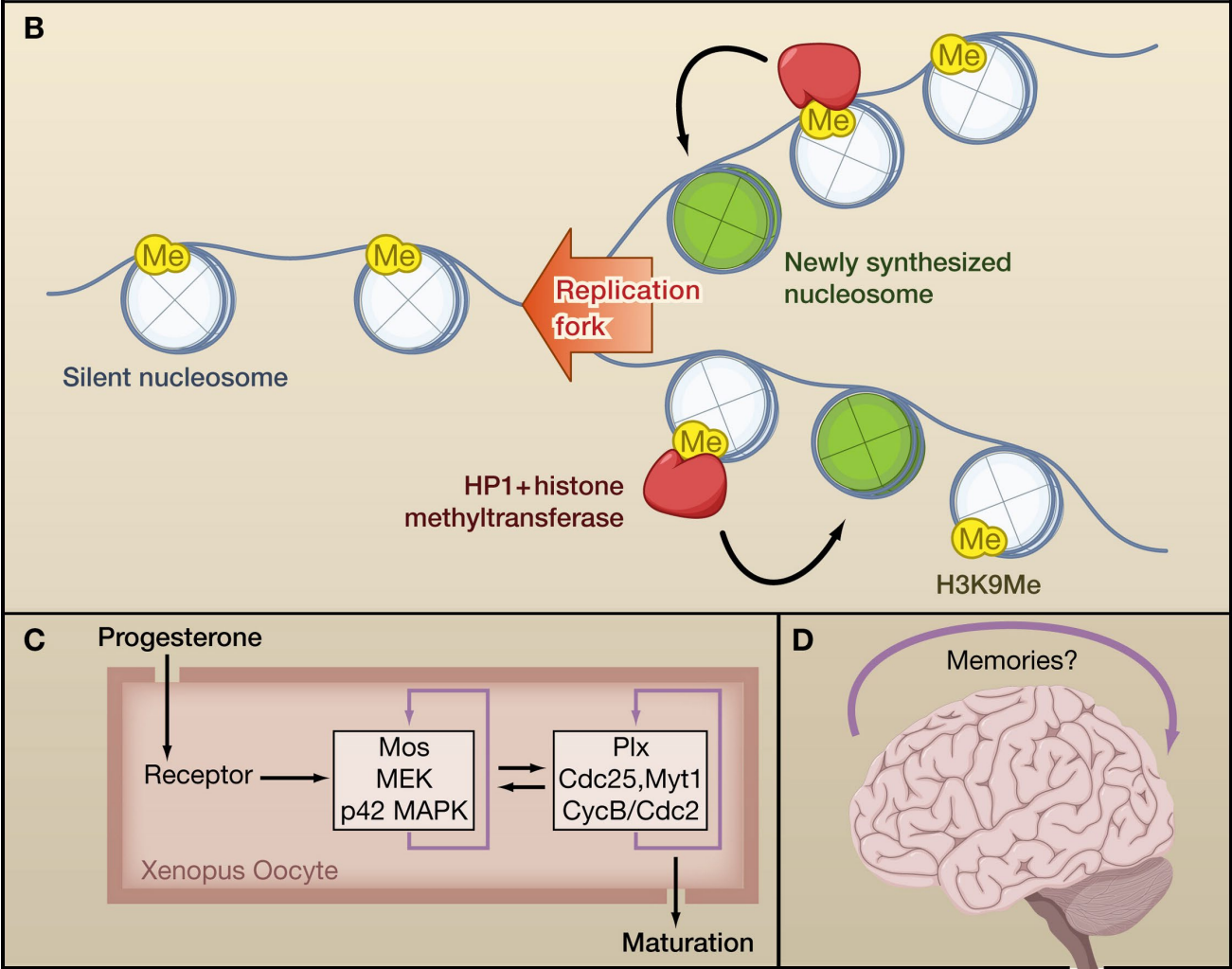
The phage lambda system



Gene Circuits for Cellular Memory

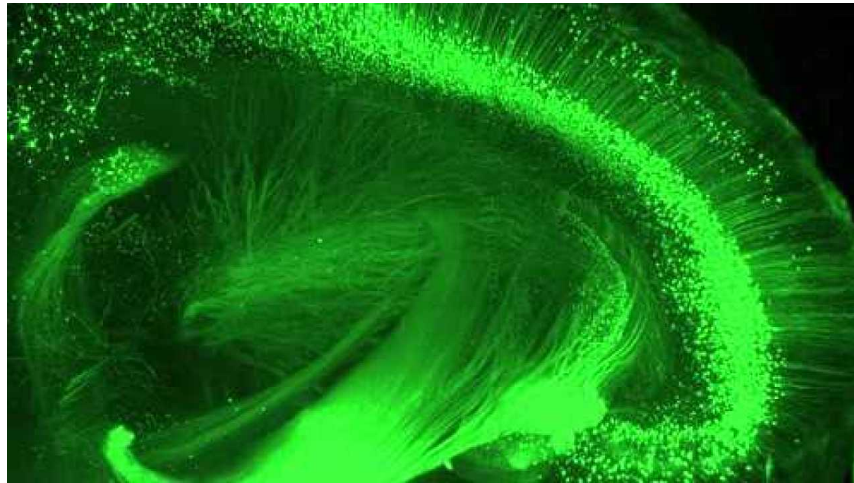


Replication-coupled nucleosome

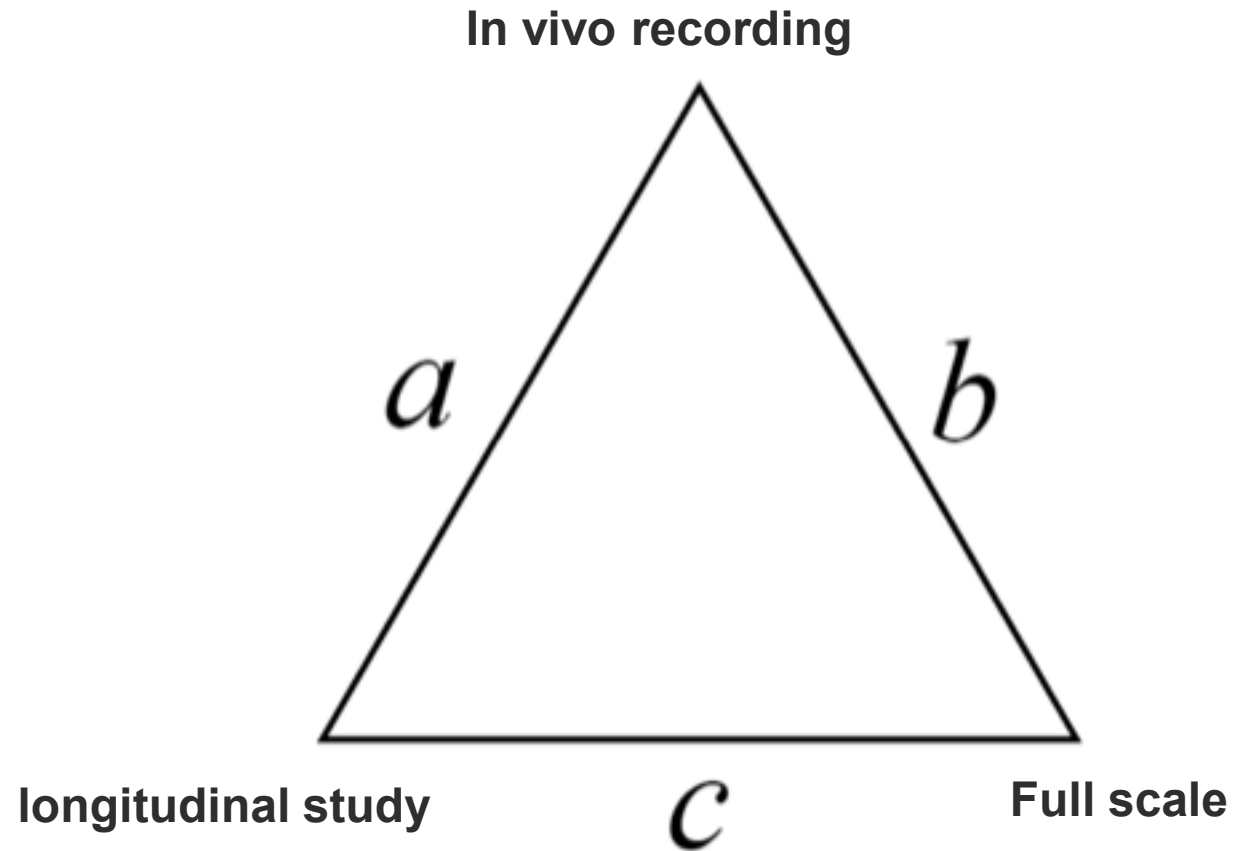


Questions in Recording Cellular Memory

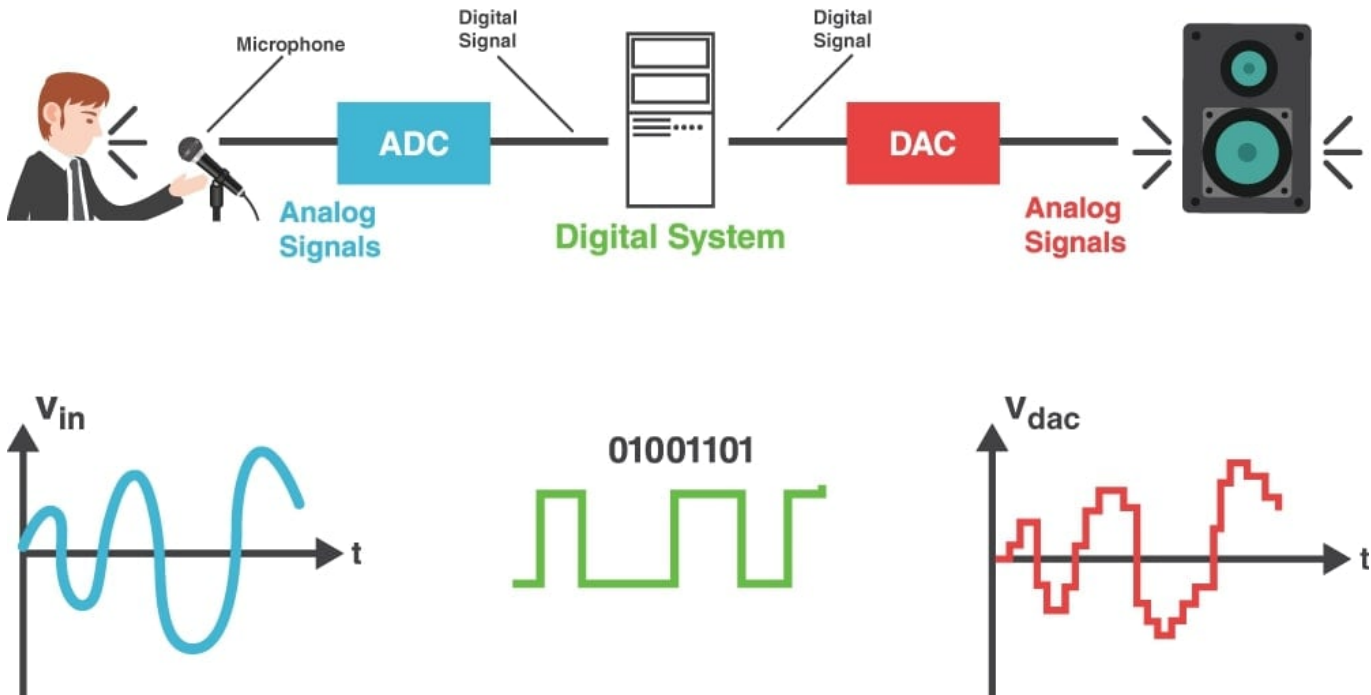
Impossible trinity in molecular recording:



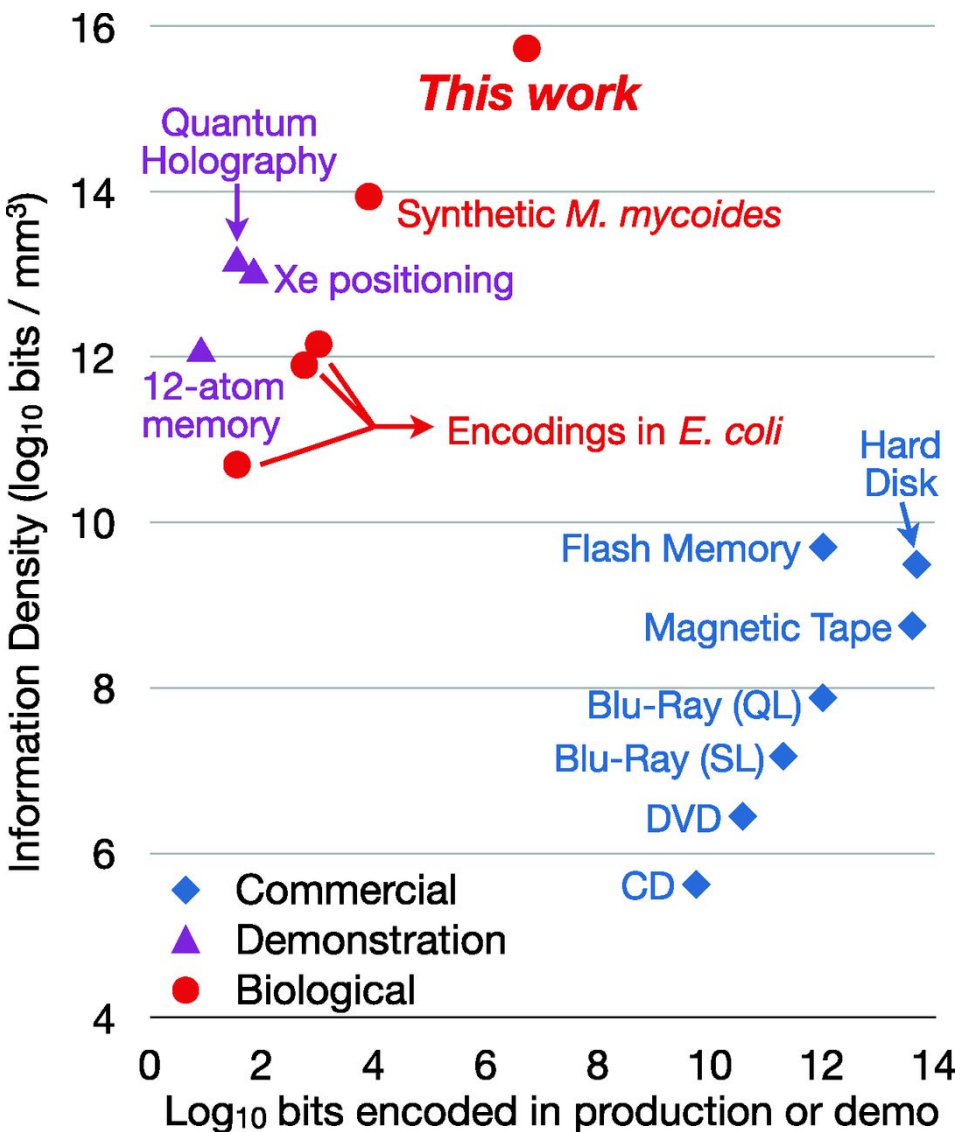
reporter signal



Digital recording in model life



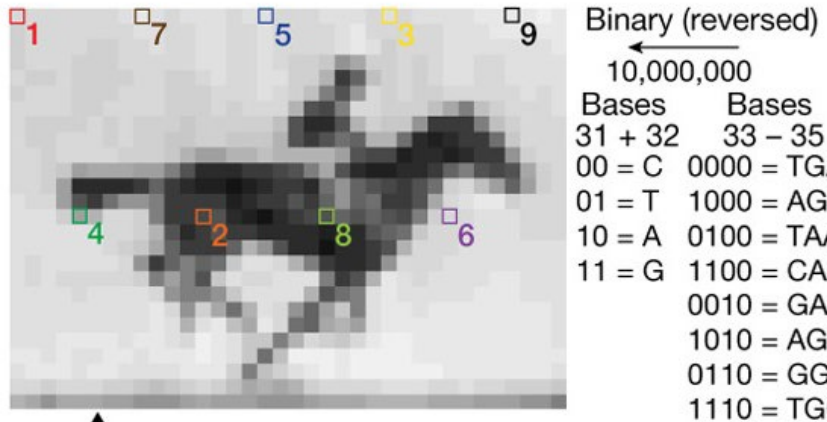
Next-Generation Digital Information Storage in DNA



Encoding a digital movie into the genomes of bacteria

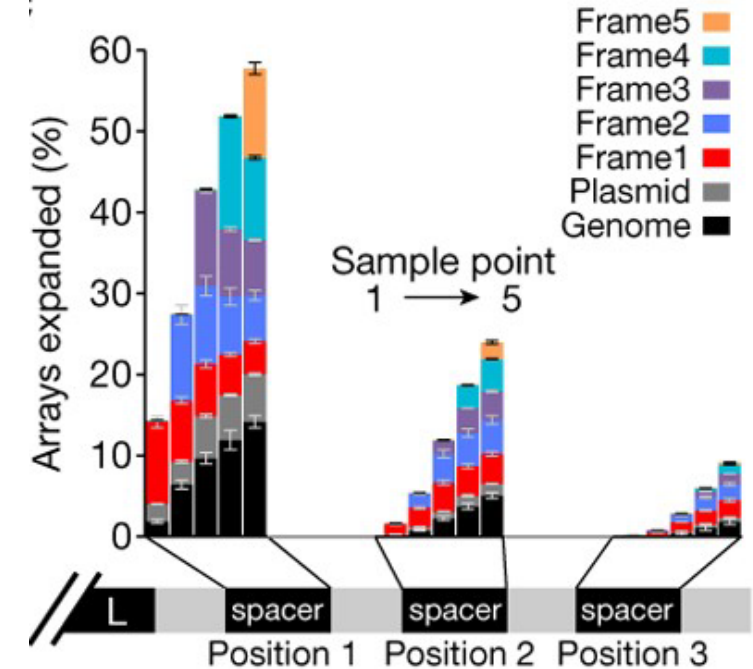
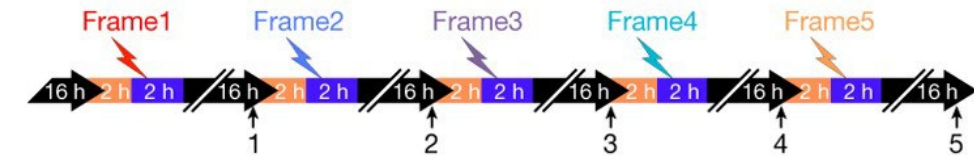
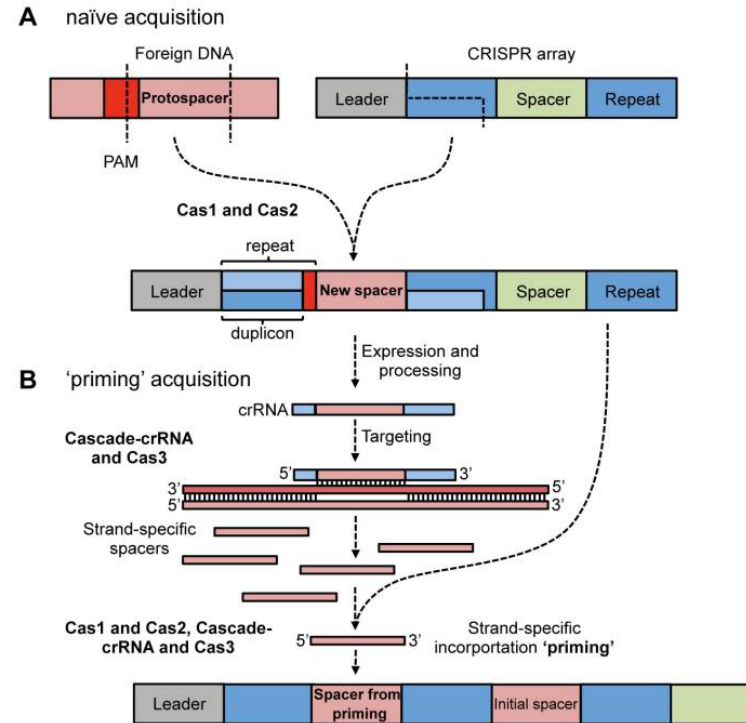
GC ~50%, no mononucleotide repeats >3 bp, no internal PAMs

Pixel
AAGCGACGTAGACTCTCTCGACAATAGTTACTGA
1 2 3 4 5 6 7 8 9 1

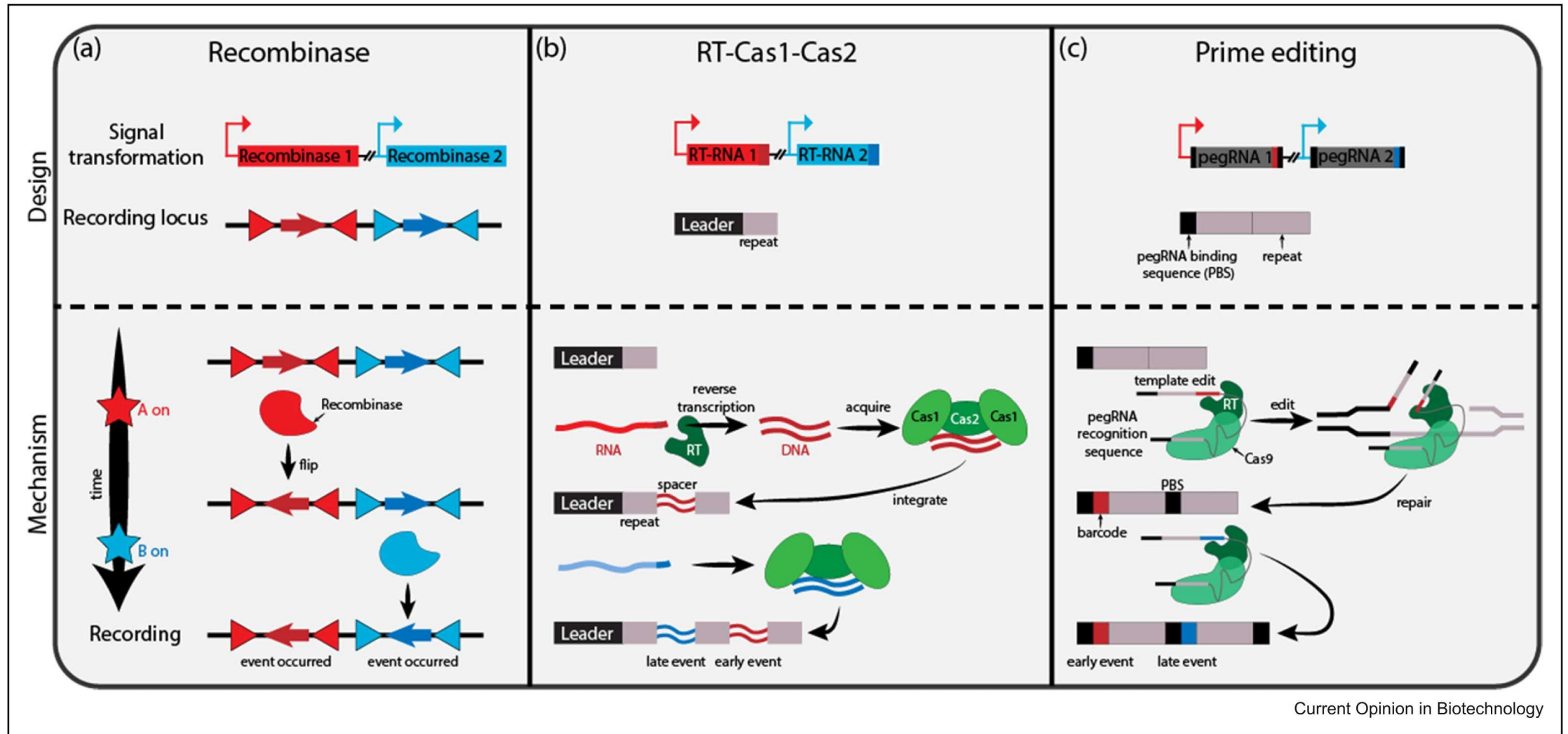


Human and Animal Locomotion

Cas1-Cas2 system

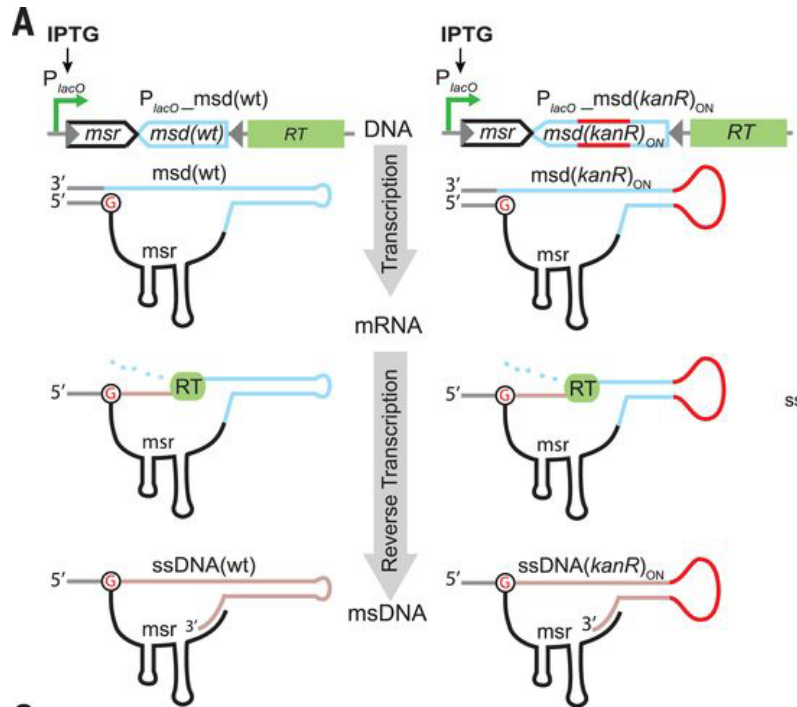


Strategy in Molecular Recording

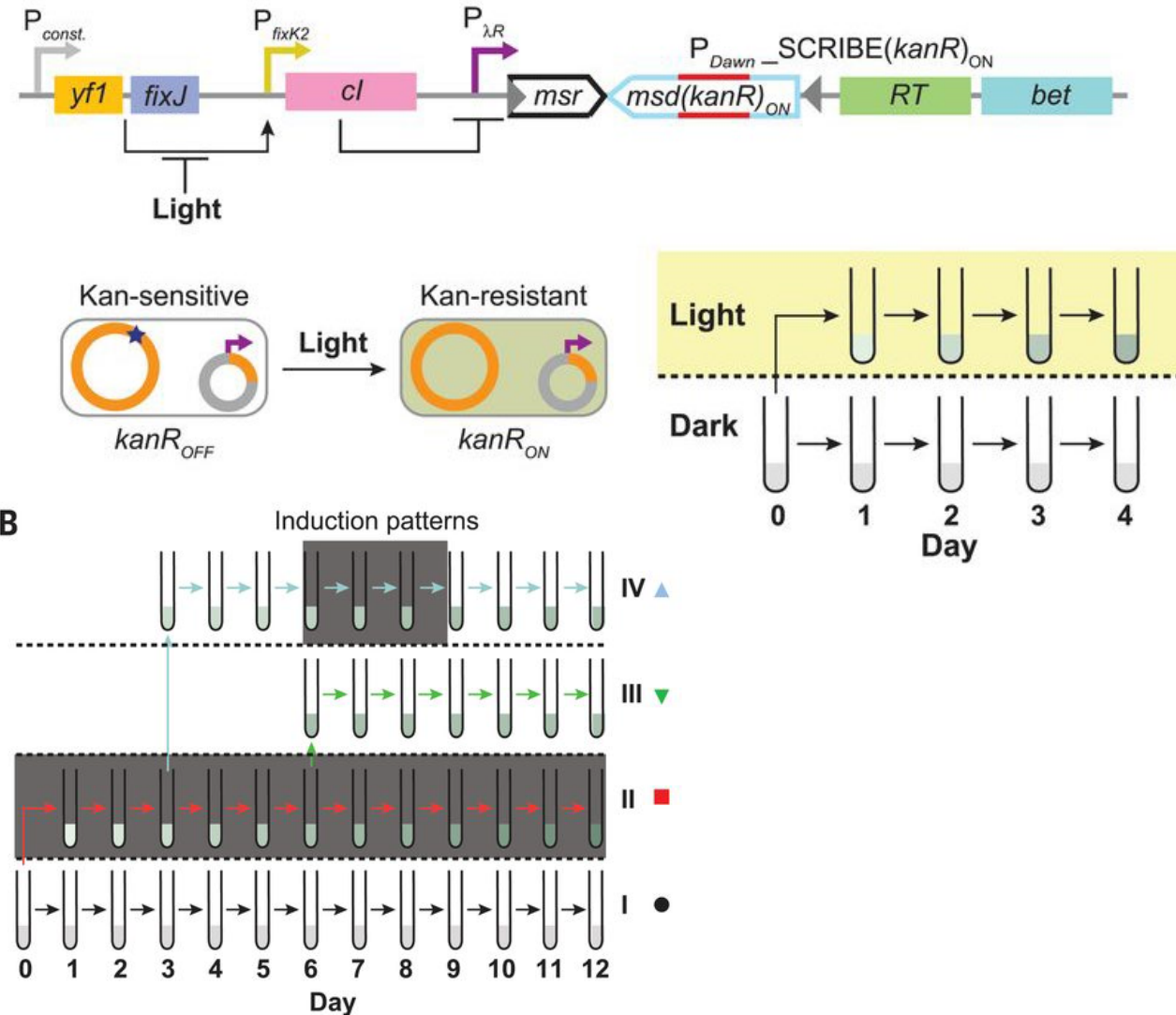


Recombinase + Recording in living populations

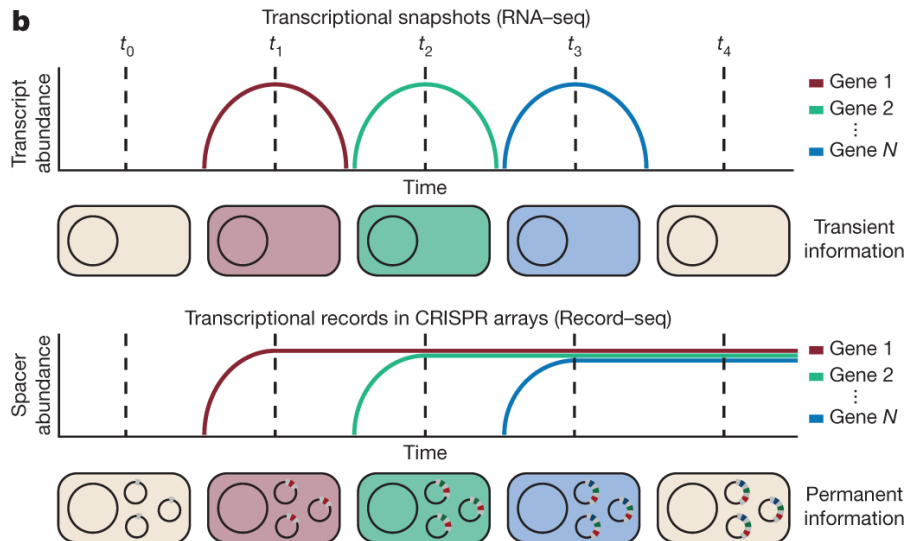
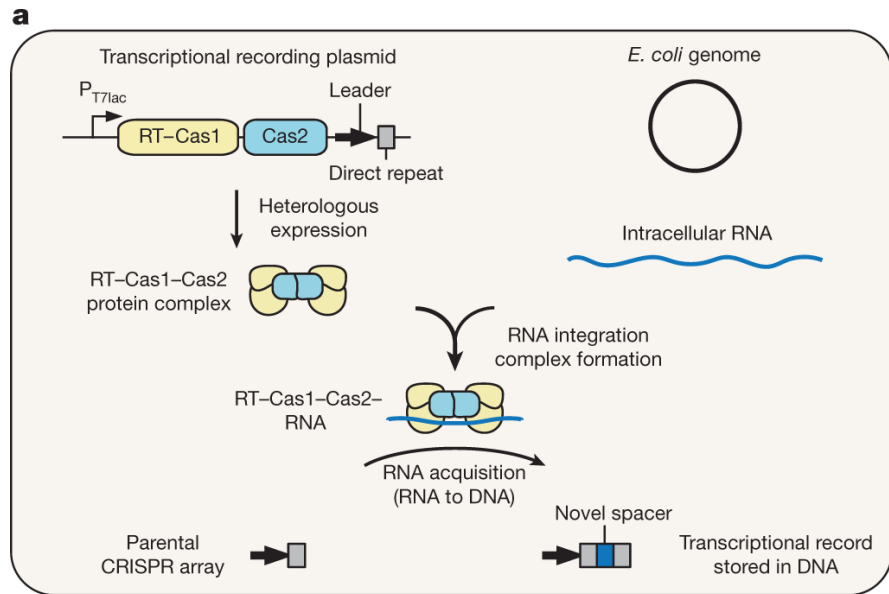
Synthetic ssDNA



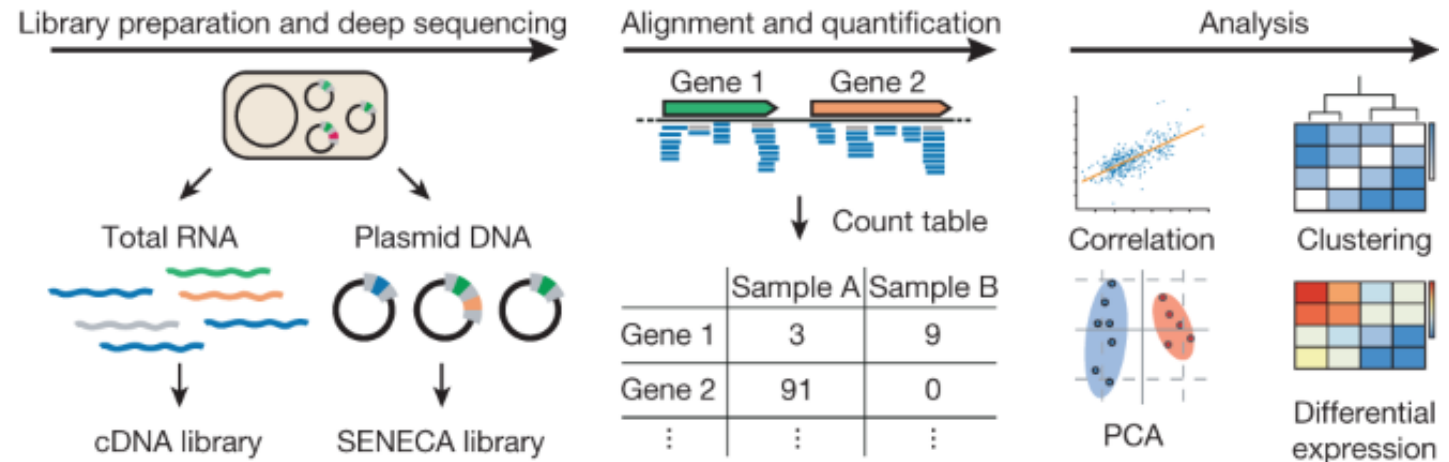
Optogenetic genome editing



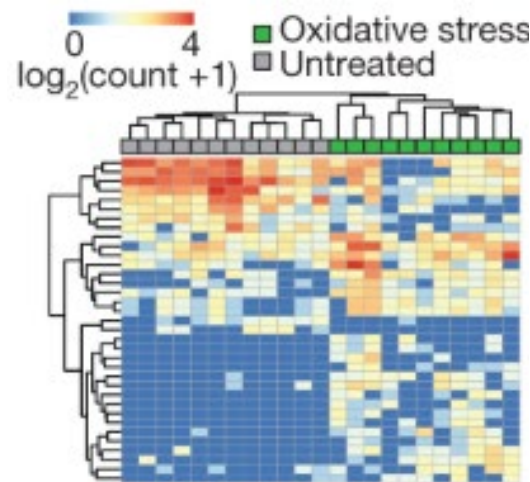
FsRT-Cas1-Cas2+ Recording in living populations



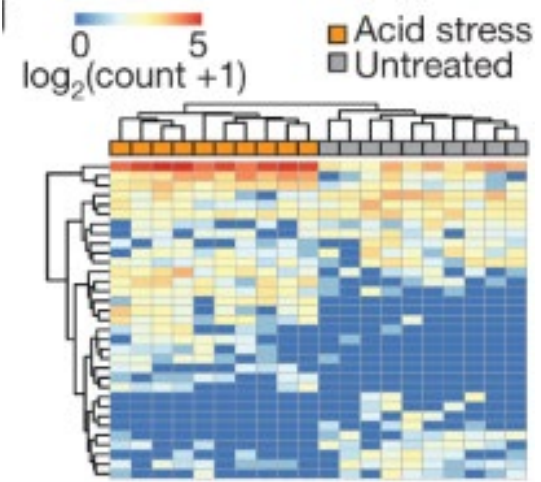
Record-seq



under oxidative stress

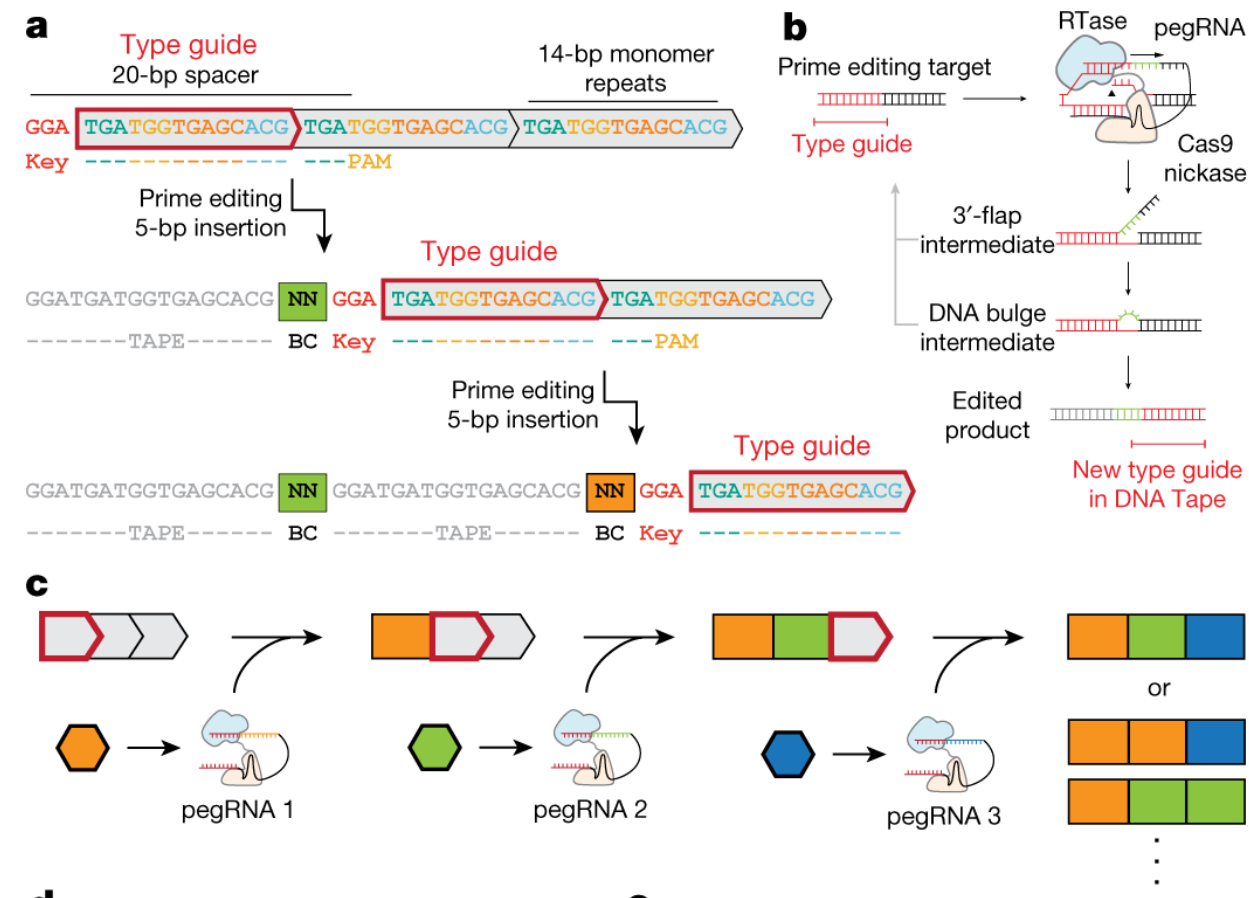


under acid stress

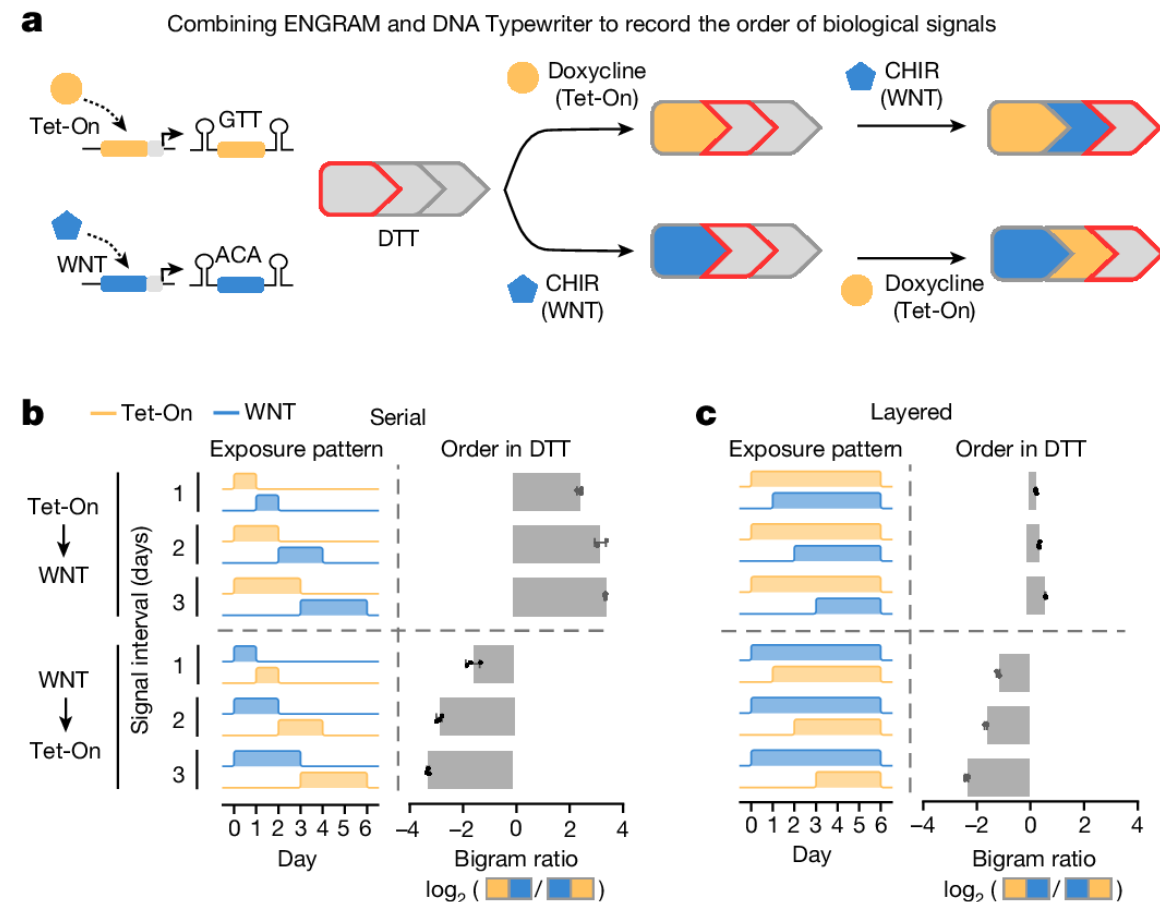


DNA Typewriter: Prime editing + recording in single cell

Sequential genome editing with DNA Typewriter



Multiplex recording of intensity and duration of activity

Choi, J. *Nature* (2022)

Chen, W. *Nature* (2024)

Liao, H. *Nat Protoc* (2024)

SRR15584112(175) → SRA Toolkit(fasta) → samtools (BAM)

```
=====
                        系统信息
=====
操作系统： Darwin 24.4.0
Python 版本： 3.9.23
R 版本： R version 4.3.3 (2024-02-29) -- "Angel Food Cake"

=====
                        命令行工具检查
=====
✅ 命令行工具： R           已安装
✅ 命令行工具： python       已安装
✅ 命令行工具： jupyter      已安装
✅ 命令行工具： conda        已安装
✅ 命令行工具： pip          已安装

=====
                        PYTHON 包检查
=====
✅ Python 包： numpy         已安装 (版本： 2.0.2)
✅ Python 包： pandas        已安装 (版本： 2.3.0)
✅ Python 包： scipy         已安装 (版本： 1.13.1)
✅ Python 包： scikit-learn   已安装 (版本： 1.6.1)
✅ Python 包： plotly         已安装 (版本： 6.1.2)
✅ Python 包： ipywidgets     已安装 (版本： 8.1.7)
✅ Python 包： rpy2           已安装 (版本： 3.5.11)
✅ Python 包： zarr           已安装 (版本： 2.18.2)
✅ Python 包： dask           已安装 (版本： 2024.8.0)
✅ Python 包： joblib         已安装 (版本： 1.5.1)

=====
                        R 包检查
=====
✅ R 包： ape                 已安装 (版本： packageVersion('ape'))
✅ R 包： phangorn            已安装 (版本： packageVersion('phangorn'))
✅ R 包： tidyverse           已安装 (版本： packageVersion('tidyverse'))
✅ R 包： ggdendro            已安装 (版本： packageVersion('ggdendro'))
✅ R 包： dendextend          已安装 (版本： packageVersion('dendextend'))
✅ R 包： devtools            已安装 (版本： packageVersion('devtools'))

=====
                        JUPYTER 集成检查
=====
✅ R Jupyter 内核已安装
[1] "rpy2 集成测试成功"
✅ rpy2 集成工作正常

=====
                        文件权限检查
=====
✅ 文件系统写入权限正常

=====
                        环境检查汇总
=====
🌟 所有依赖已正确安装！环境准备就绪。
您现在可以运行单细胞谱系分析流程。
提示：启动 JupyterLab 使用命令：jupyter lab
```

Step 1: TAPE_10X_read2fromBAM

Objective: Extract raw data from CellRanger-generated BAM files to generate a table linking cell barcodes, TAPE barcodes, edit sites (Sites 1–5), and molecule counts.

Input: Position-sorted BAM file (10X Genomics data).

Output: A cell-TAPE edit association table

Step 2: TAPE_text_sorting
Objective: Process the output from Step 1 to generate structured statistics for downstream analysis.

Input: The cell-TAPE edit association table from TAPE_10X_read2fromBAM.

Output: Bigram matrix; Unigram frequencies; Transfected barcode order.

	Cell	TargetBC	Site1	Site2	Site3	Site4
1	AGCCATAGCGCCTAC-1	ATTTGGTT	CCGGGA	CCGGGA		
2	AGCCATAGCGCCTAC-1	TTCACGTA	CCGGGA	TATGGA		
3	AGCCATAGCGCCTAC-1	TAGATTTT	CTAGGA	GCCGGA		
4	AGCCATAGCGCCTAC-1	TTGTTTAC	CATGGA	CCGGGA		
5	AGCCATAGCGCCTAC-1	ATTTATAT	ATAGGA	CATGGA	CATGGA	CATGGA
6	AGCCATAGCGCCTAC-1	TTAGATTG	GATGGA	CTAGGA	ACAGGA	
7	AGCCATAGCGCCTAC-1	TGGACGAC	GCCGGA			
8	AGCCATAGCGCCTAC-1	TGGTTTTG	GATGGA	TATGGA	GCCGGA	ACGGGA
9	AGCCATAGCGCCTAC-1	TGCGATTT	CCGGGA	GCCGGA		
10	AGCCATAGCGCCTAC-1	ACCTCGTG	GCTGGA	GCTGGA	TGTGGA	CATGGA
11	AGCCATAGCGCCTAC-1	TTTCGTGA	CCGGGA	TATGGA		
12	AGCCATAGCGCCTAC-1	GTAAAGAT	GCCGGA	GCCGGA	CCGGGA	
13	AGCCATAGCGCCTAC-1	TTGAGGTG	CATGGA	GCCGGA	TATGGA	CCGGGA
14	AGCCATAGCGCCTAC-1	ATGGTAAG	CATGGA	CCGGGA		
15	AGCCATAGCGCCTAC-1	GCAGGGTG	GATGGA	TATGGA	CATGGA	ATAGGA
16	AGCCATAGCGCCTAC-1	ATTAGTCA	CCGGGA	GCCGGA	GCCGGA	
17	AGCCATAGCGCCTAC-1	ATCACGCT	TAAGGA	TATGGA	CATGGA	GCCGGA

DNA Typewriter

Step 3 (Parallel Execution): Visualization Tools

3a. PEbigram_plotting

Objective: Visualize bigram frequencies as a heatmap.

Input: Bigram matrix and unigram frequencies from TAPE_text_sorting.

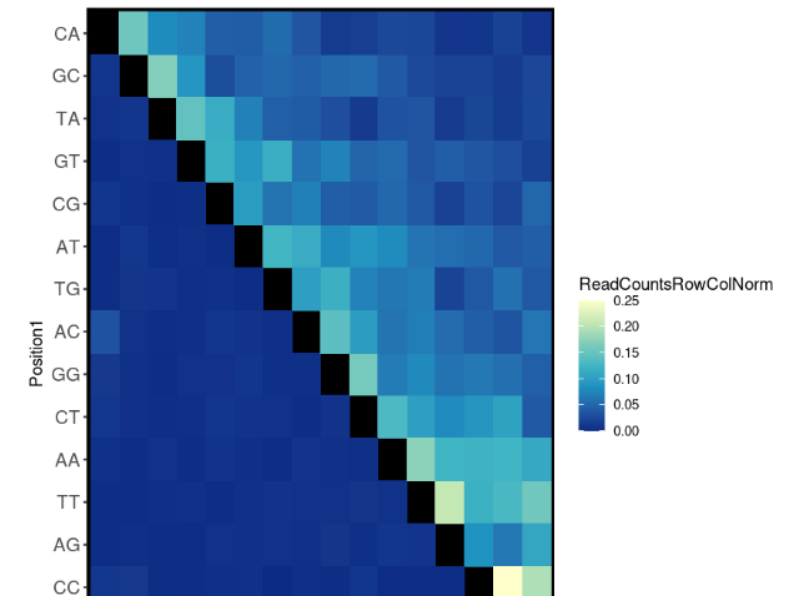
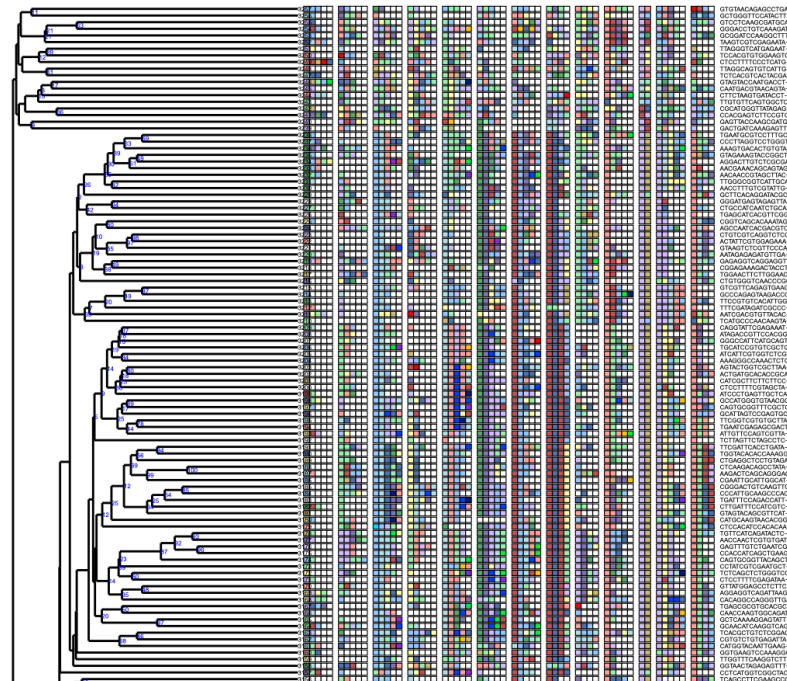
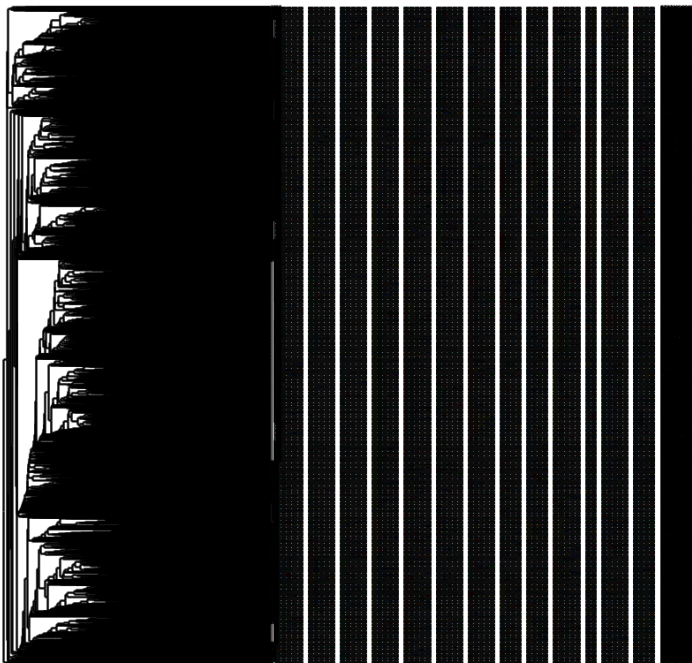
Output: Heatmap plot of bigram frequencies

3b. DNATypewriter_SingleCellLineage_Rscript

Objective: Generate lineage-specific plots for single-cell editing patterns.

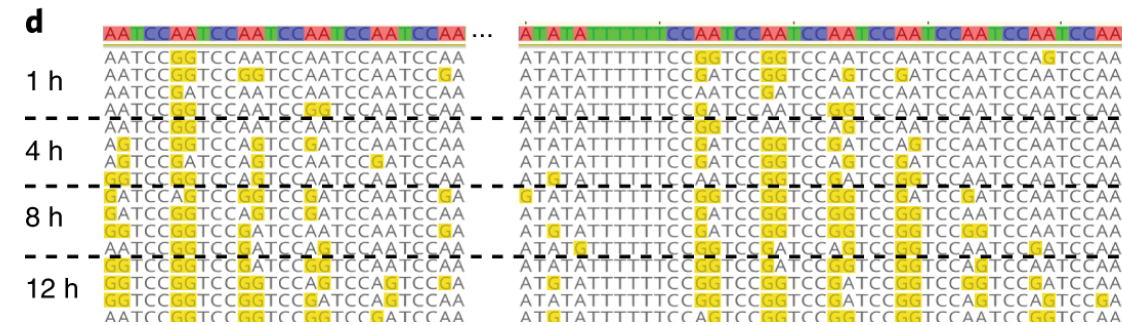
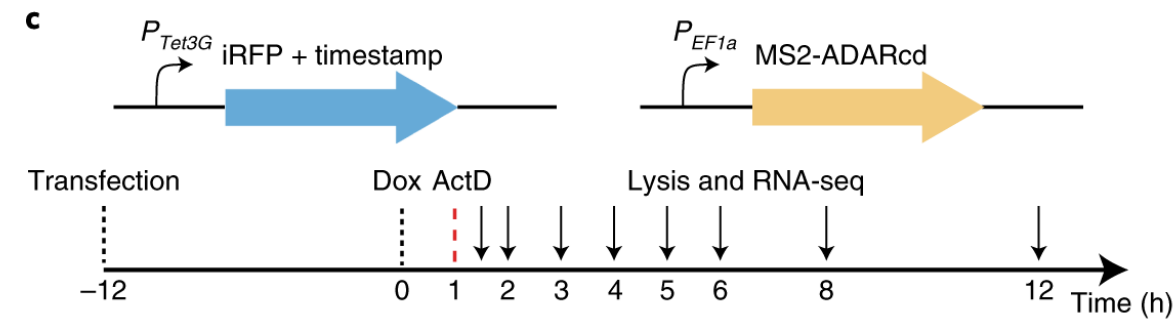
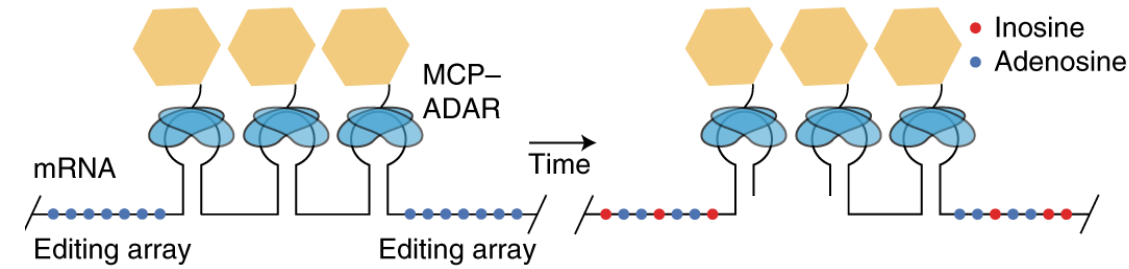
Input: Bigram matrix and unigram frequencies from TAPE_text_sorting.

Output: Lineage-specific visualization

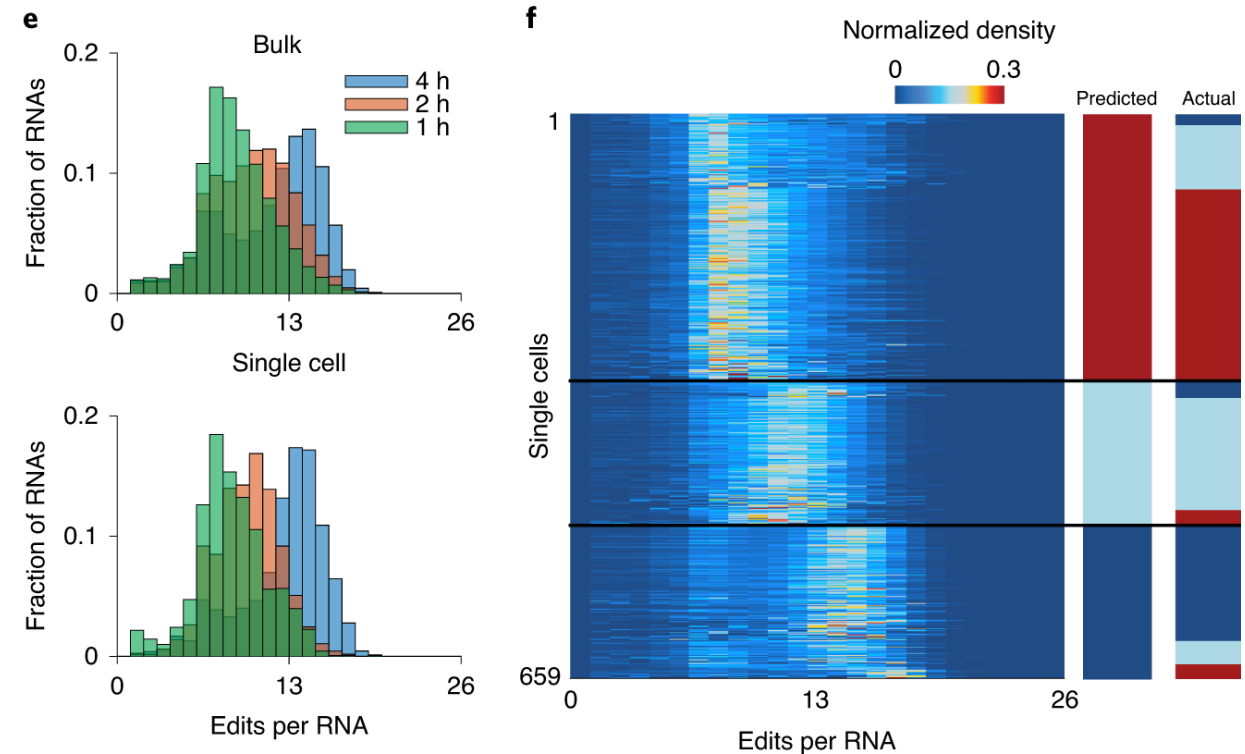


RNA seq + Pseudotime Inference

timestamps approach

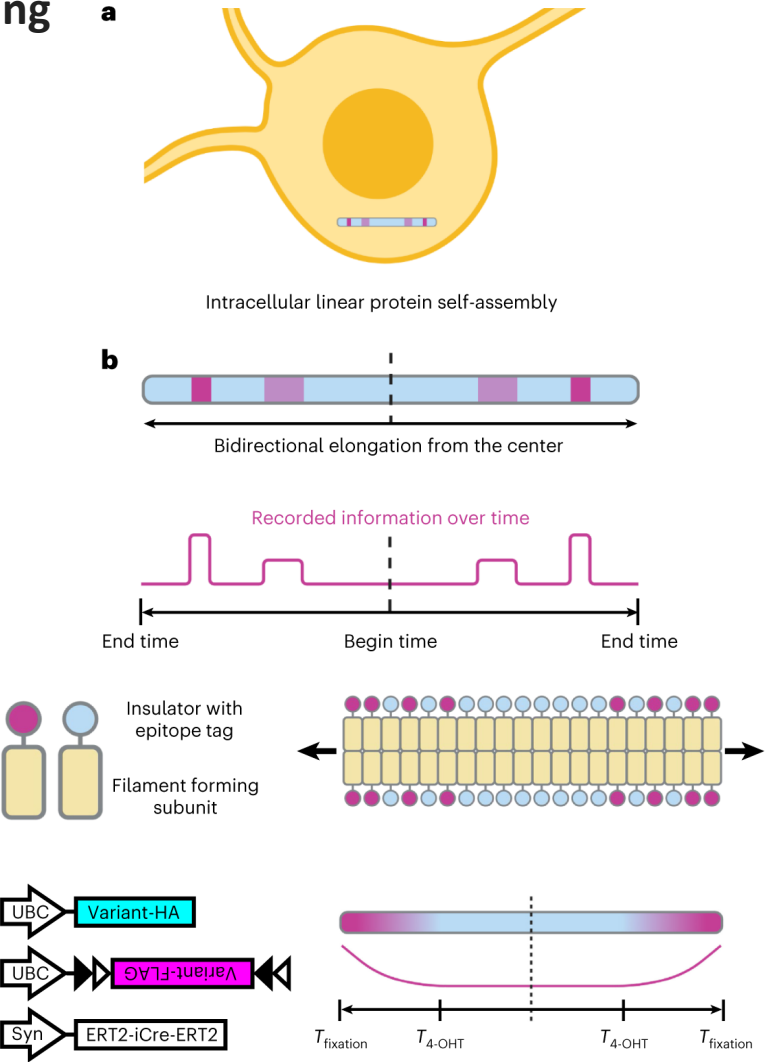


Pseudotime Inference

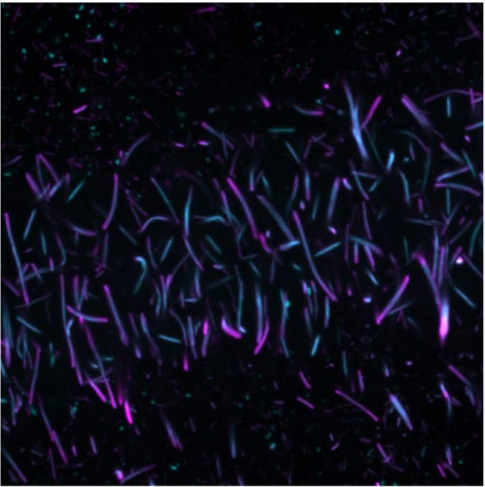
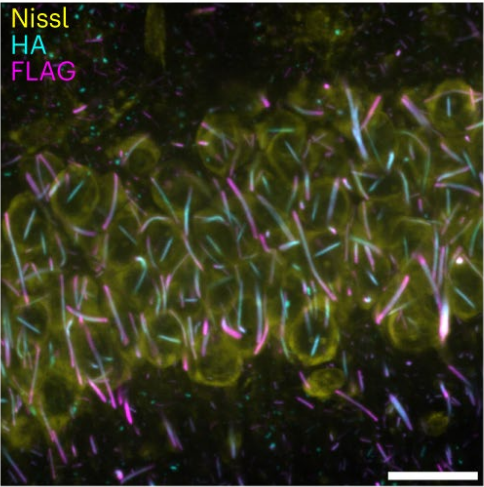
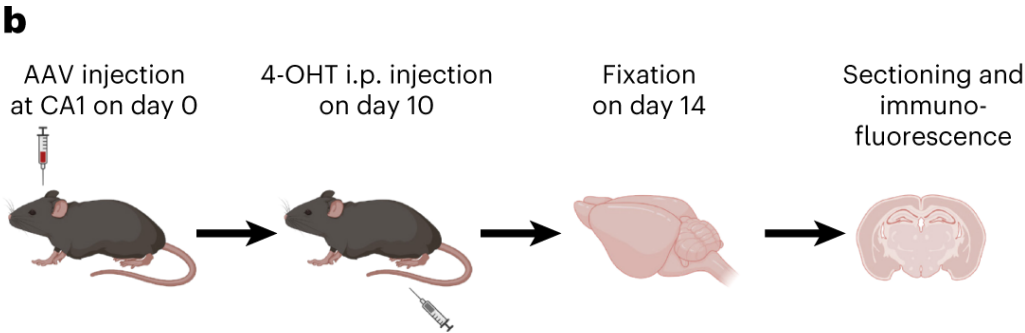
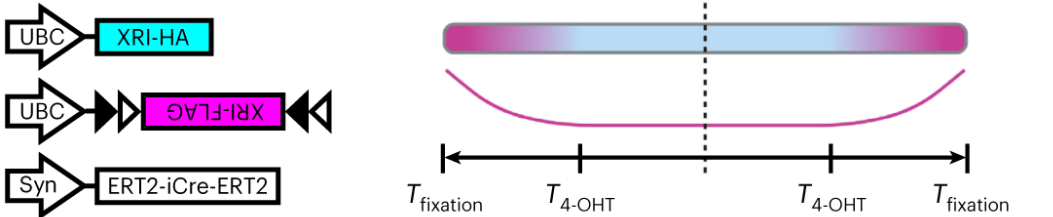


protein chain + absolute timing

linear protein self-assembly-based cellular physiology recording



In vivo XRI self-assembly in mouse brain

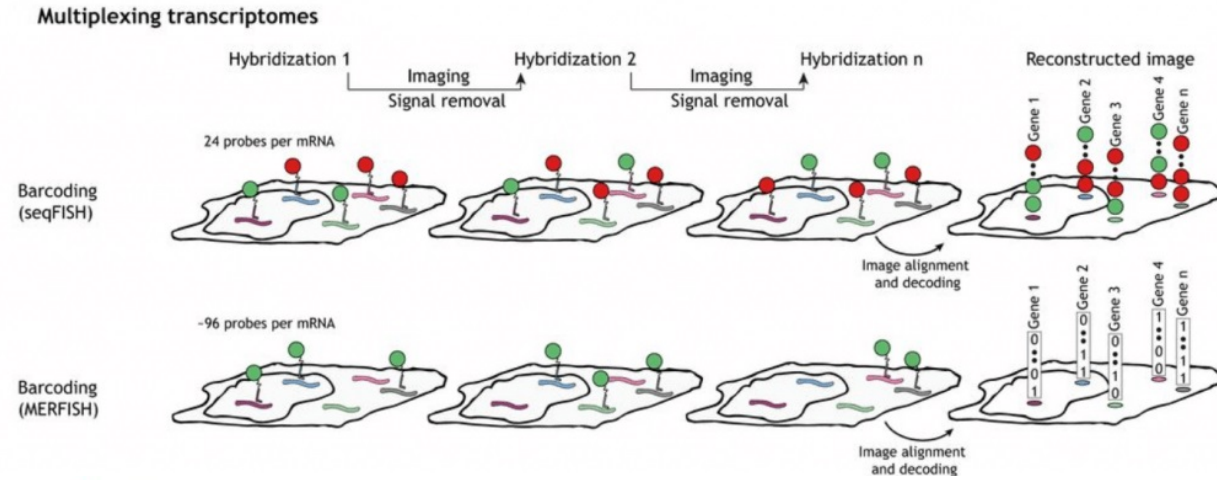


Lin, D. *Nat Biotechnol* (2023).

Linghu, C. *Nat Biotechnol* (2023)

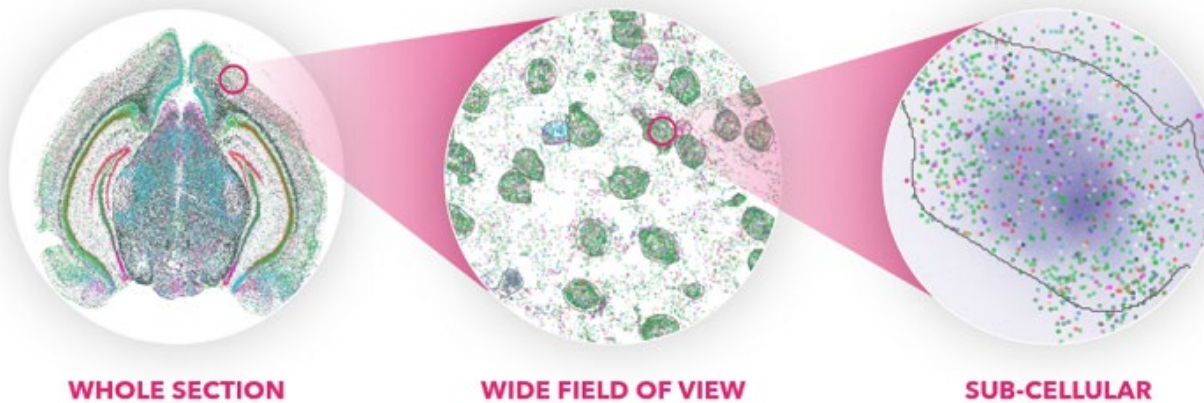
Orchestrate spatiotemporal signals in single-cell sequencing

MERFISH (Multiplexed Error-Robust Fluorescence In Situ Hybridization)



vizgen

MERFISH • Multiplexed Error Robust Fluorescence In-Situ Hybridization
Profiling 483 genes with subcellular resolution across a full mouse coronal slice



MAPseq for Brain Connectome

