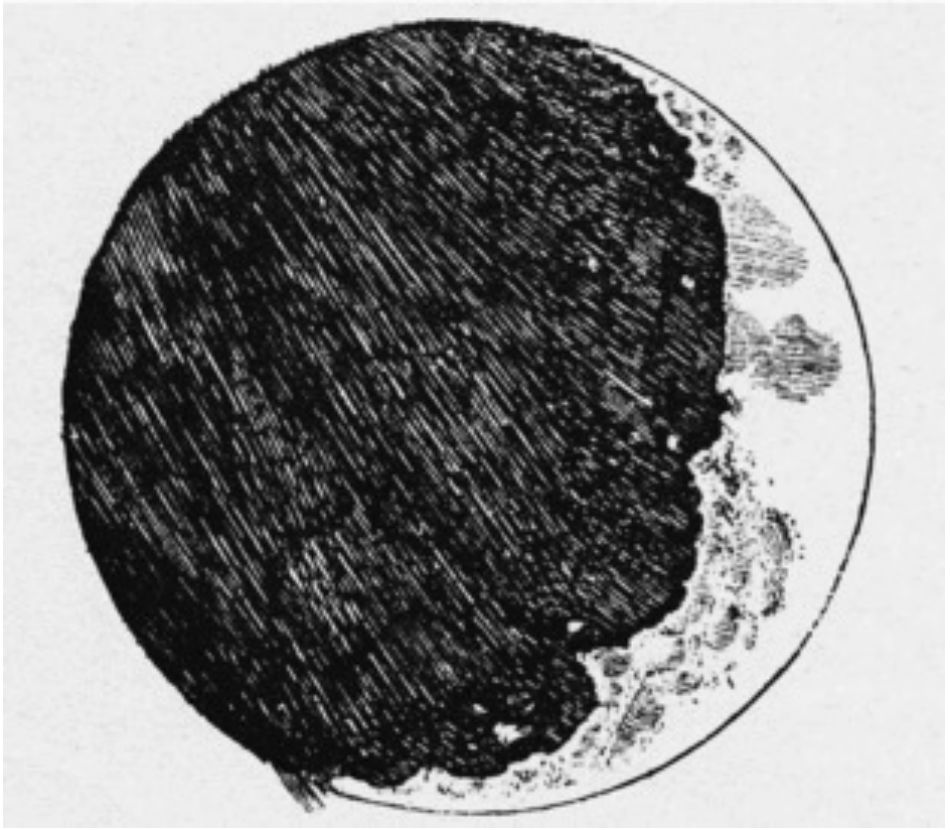
How to compose figures

At least one…

Katie Galloway

Purpose of figure?

1. Provide context, scale, frame the story
2. Convey complex ideas
3. Summarize data

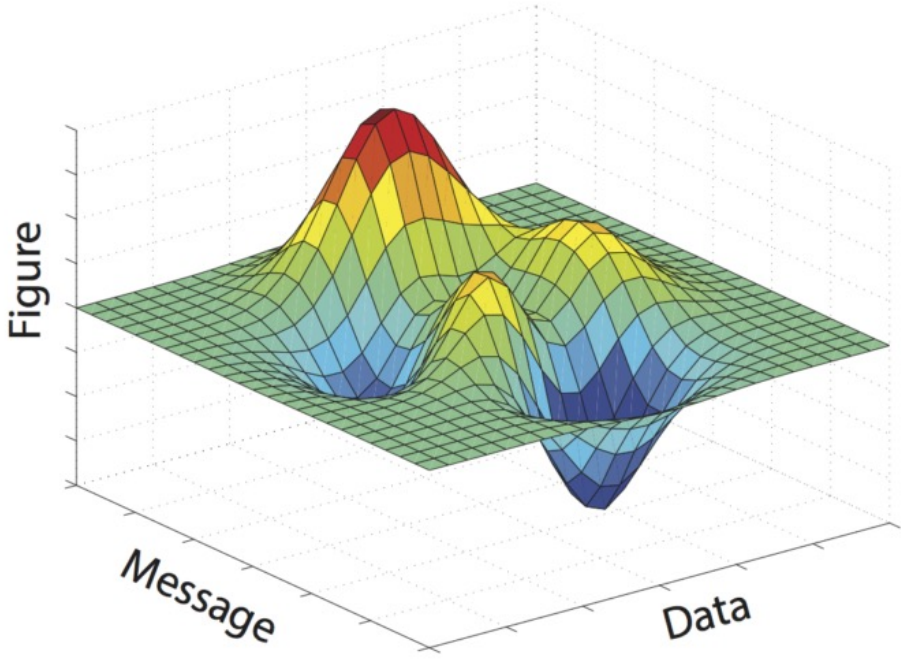


Engraving of the moon from Galileo's *Sidereus Nuncius* (III, 53-96) 1610.

Purpose of figure?

Figure 2 Identify the message and title of your figure first

One figure needs one (and only one) message



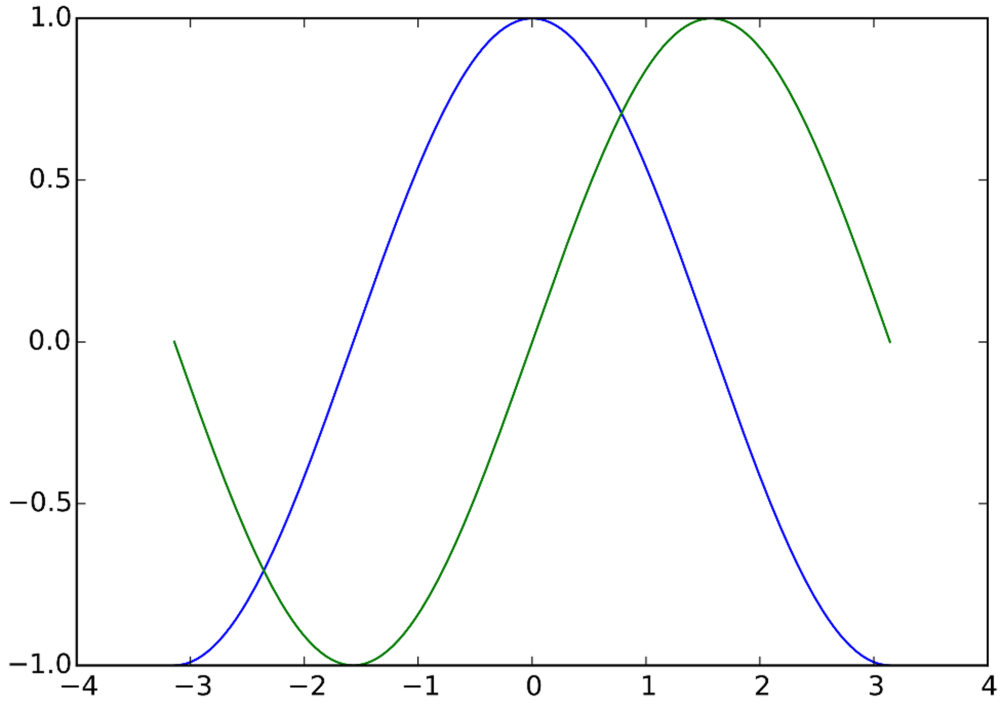
1. Obtain data
2. Analyze data
3. Decide on the most important message
4. Build a figure that supports your message

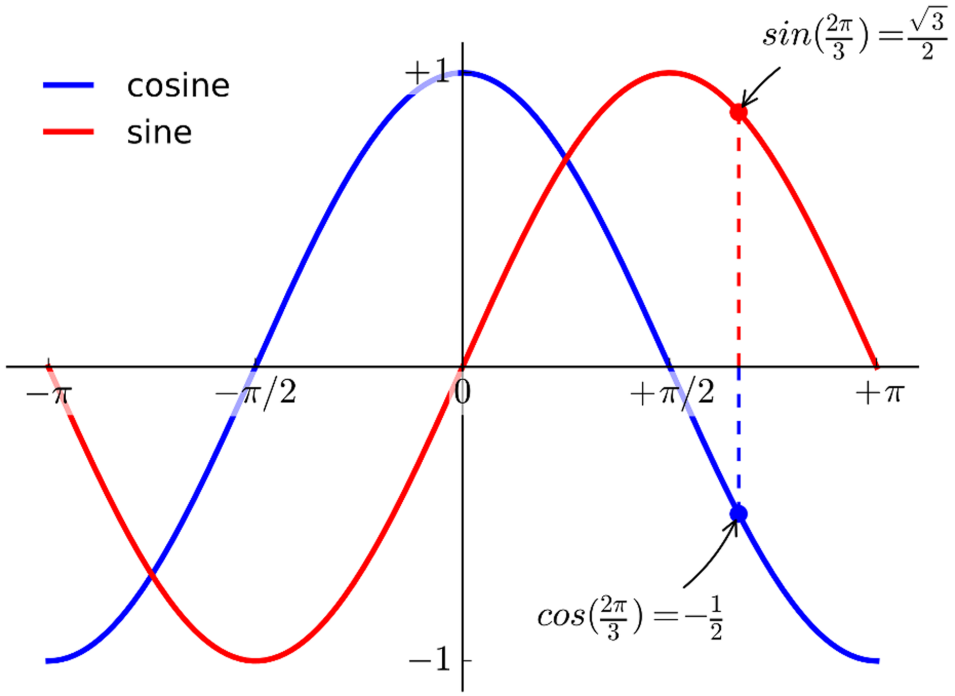
Source: Broad Comm Kit Figure Design

What are the key elements of a figure?

1. Graphics (See Tutorial 4)
2. Data
3. Legend

Do not trust the defaults!

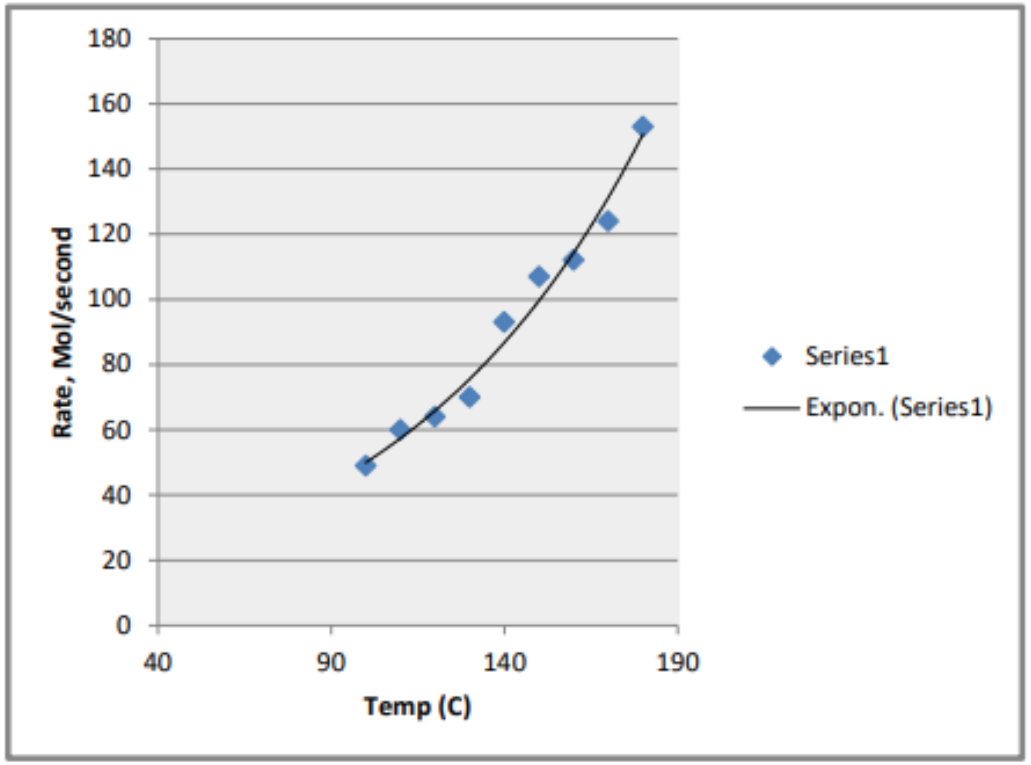




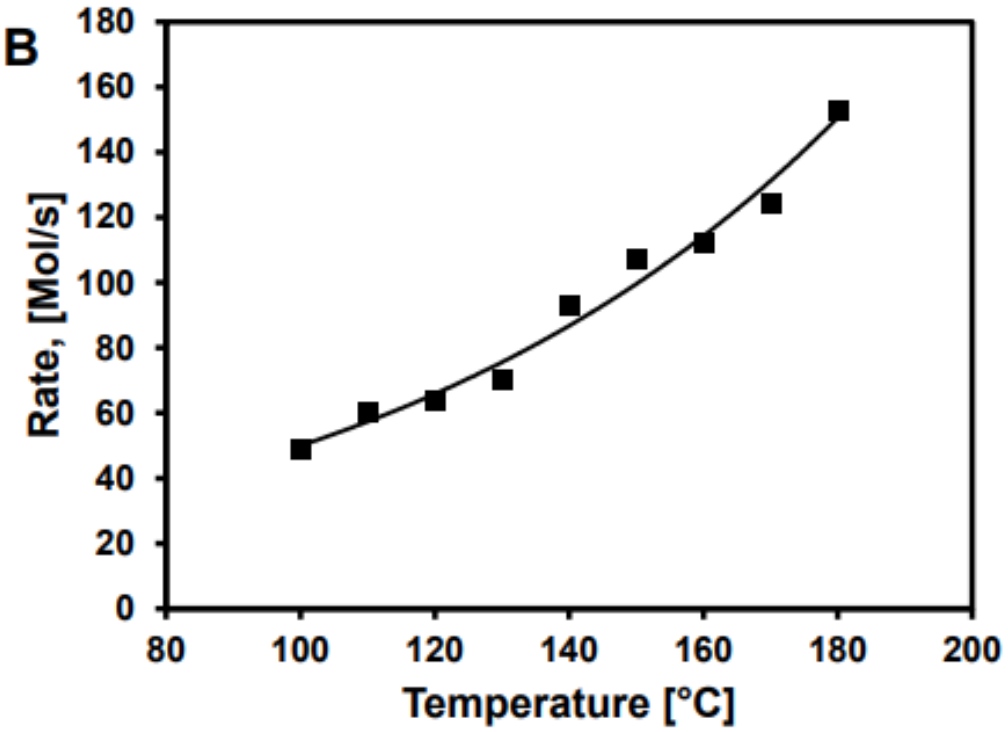
[Ten Simple Rules for Better Figures]

Do not trust the defaults!

A



B



Courtesy of Paul Dauenhauer from Spring_2020_ChEn_3401W_ReportGuidelines_ver_02

Do not trust the defaults!

Table 2.1 | Bad Example of a Data Table.

Trial	Temp (F)	Flow (gpm)
1	60	2.00
2	90	27.3421
3	80	18.118
4	70	10.20
5	80	18.11
6	60	2.009
7	90	26.99
8	70	10.30

Table 2.2 | Good Example of a Data Table.

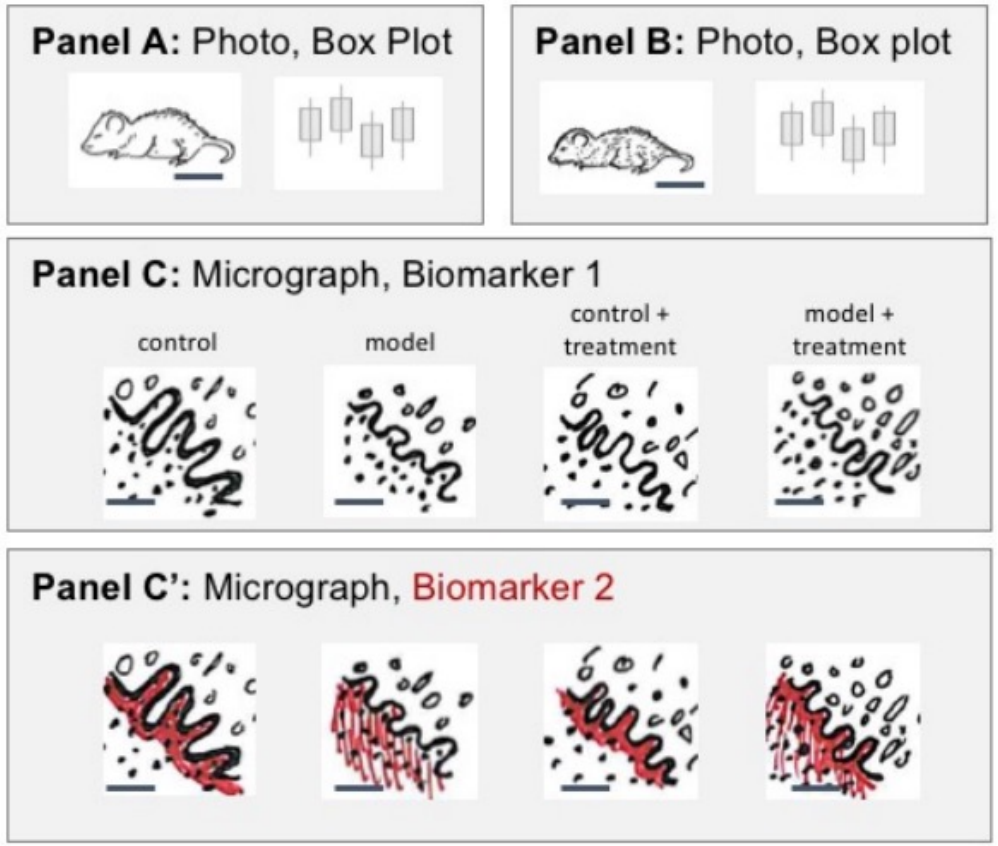
Temperature (°F)	Flow (gpm)	Trial
60	2.00	1
60	2.01	6
70	10.20	4
70	10.30	8
80	18.12	3
80	18.11	5
90	27.34	2
90	26.99	7

Courtesy of Paul Dauenhauer from Spring_2020_ChEn_3401W_ReportGuidelines_ver_02

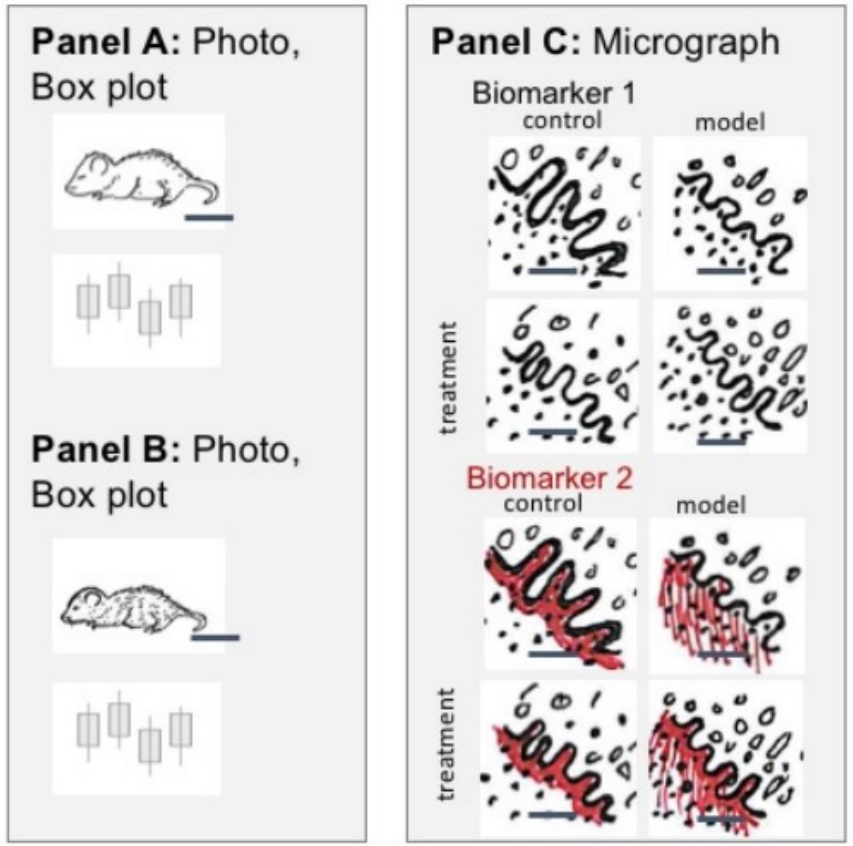
Layout…

Organize panels into **“Figure layout sketch”**, exemplary for Figure planning table in A

Layout in rows

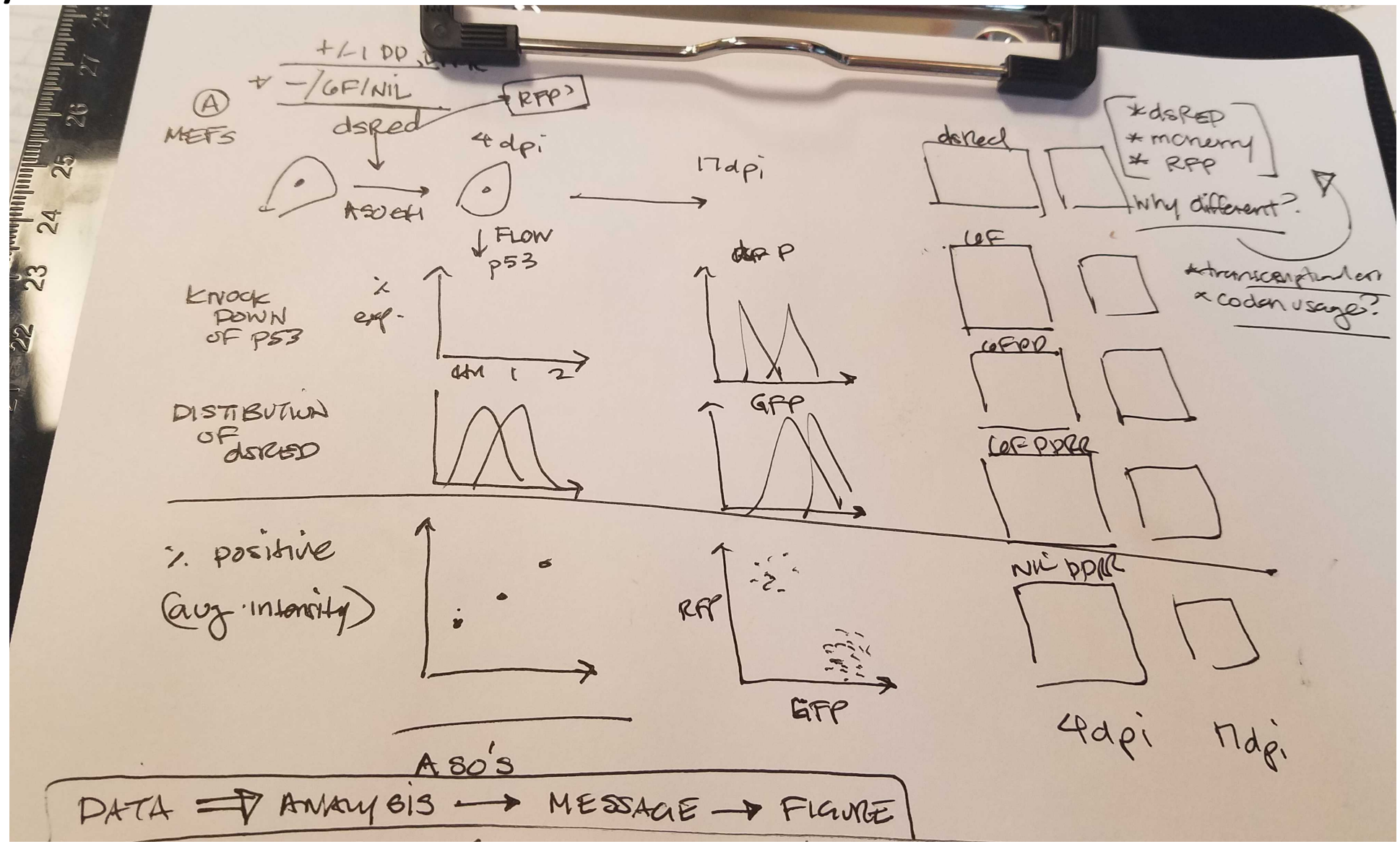


Layout in columns

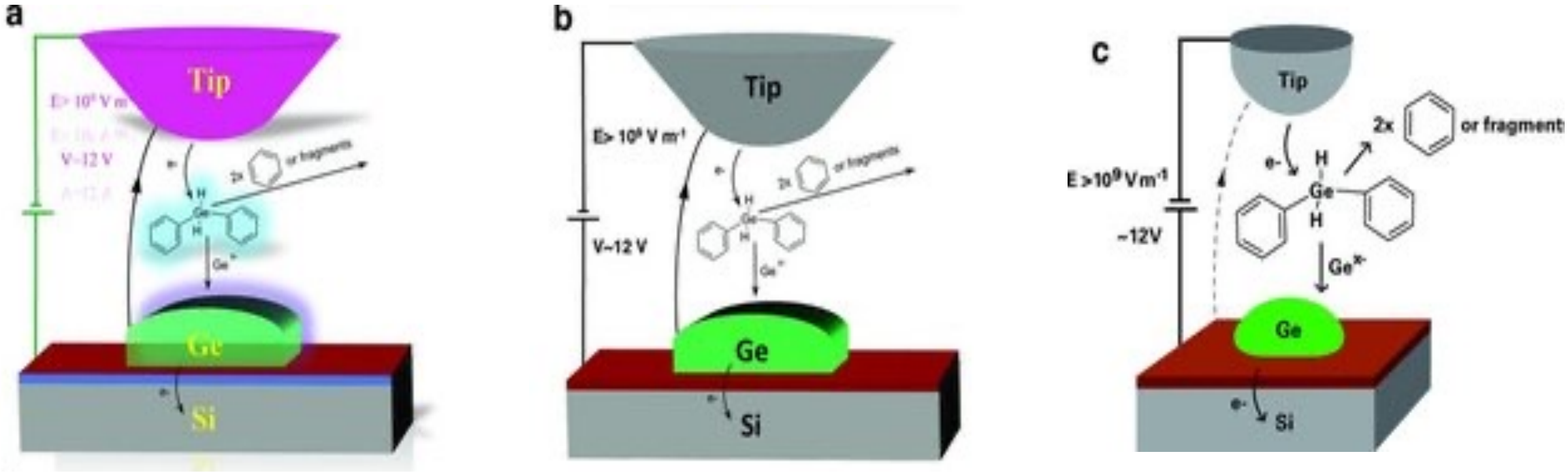


[Source: Creating Clear and Informative Image-based Figures for Scientific Publications]

Layout…

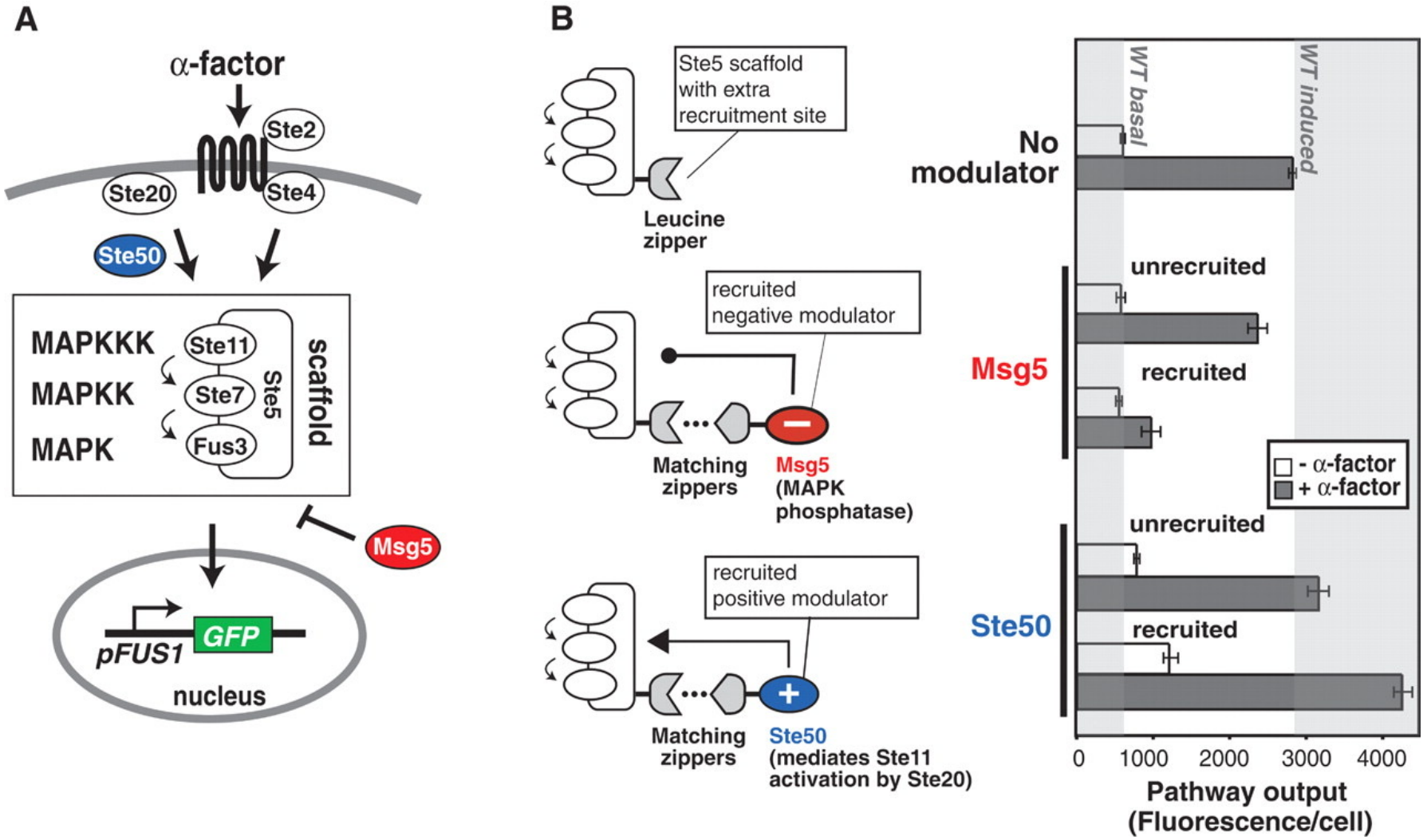

+/- DD
+/- GFP/NIL
RFP?
dsRed
MEFs
4 dpi
17 dpi
ASO ctrl
FLOW p53
KNOCK DOWN OF P53
exp.
DISTIBUTION OF dsRED
dsRed
*dsRED
*monomy
* RFP
Why different?
GFP
*transcriptomics
*codon usage?
% positive
(avg. intensity)
RFP
GFP
NIL DP
4dpi
17dpi
ASO's
DATA ⇒ ANALYSIS → MESSAGE → FIGURE

Scale and color…



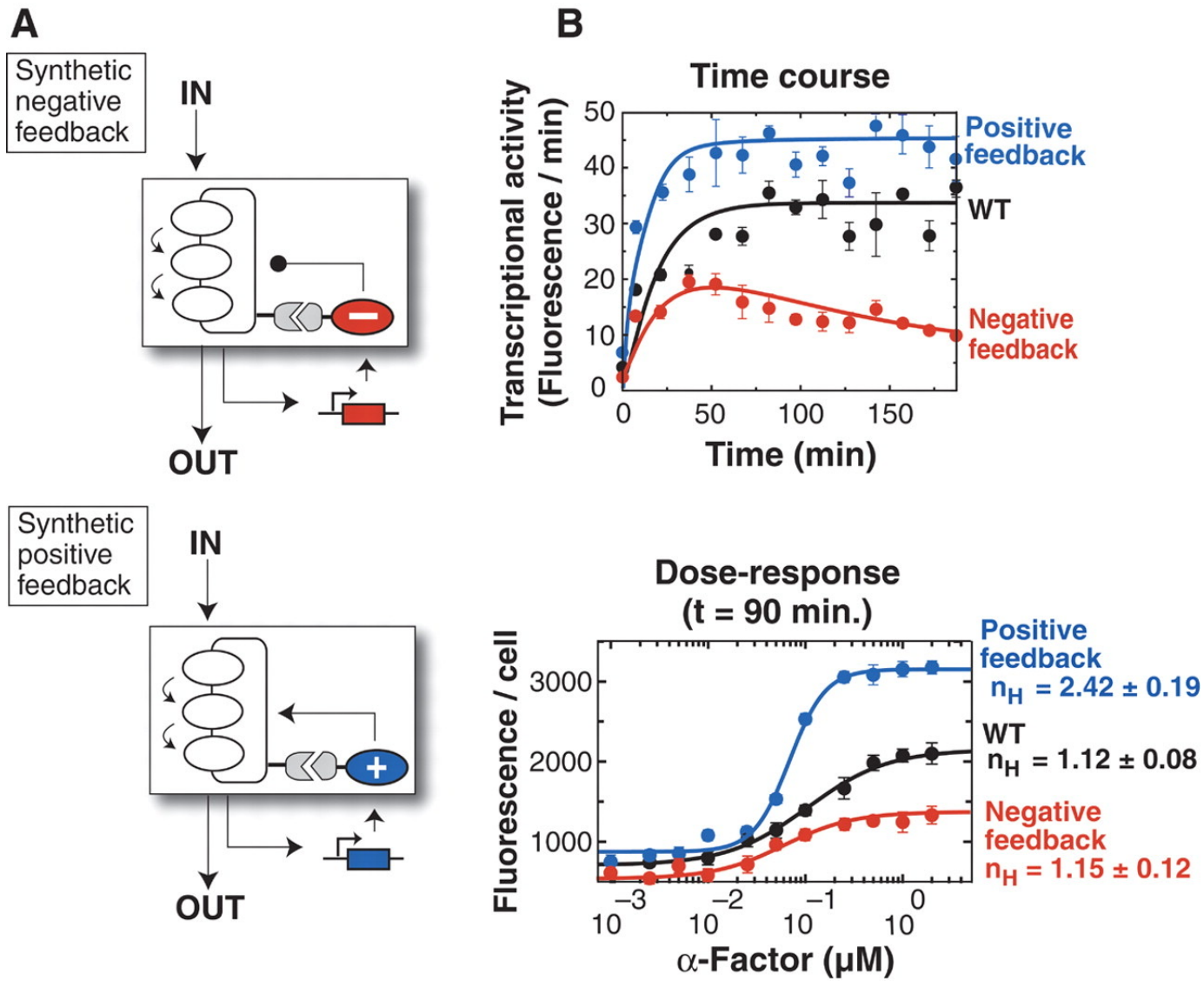
[Source: A Brief Guide to Designing Effective Figures for the Scientific Paper]

You can tell just by looking at figures…



[Using Engineered Scaffold Interactions to Reshape MAP Kinase Pathway Signaling Dynamics]

Consistent use of color

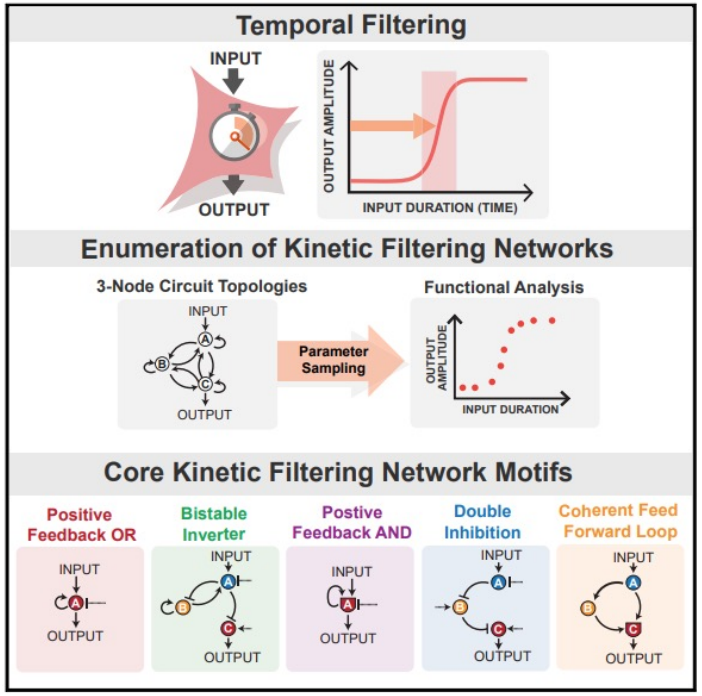


Using Engineered Scaffold Interactions to Reshape MAP Kinase Pathway Signaling Dynamics

Cell Systems

The Design Principles of Biochemical Timers: Circuits that Discriminate between Transient and Sustained Stimulation

Graphical Abstract



Authors

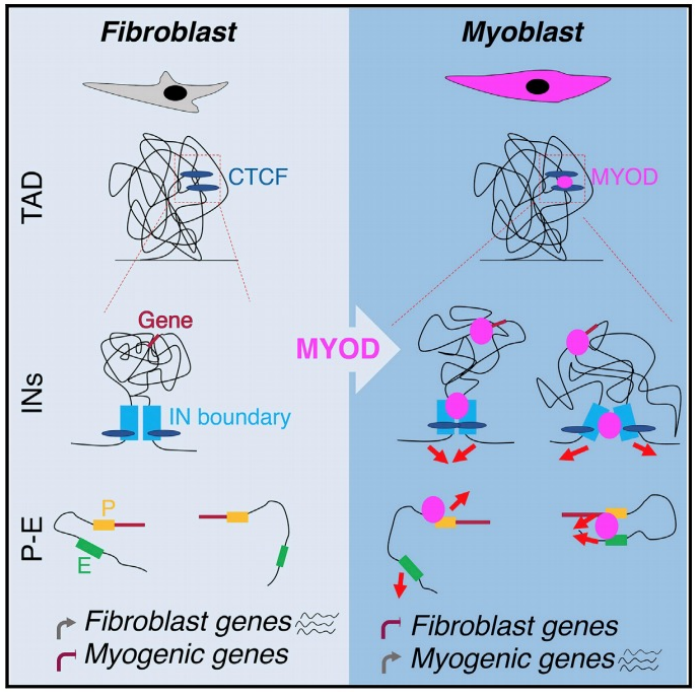
Jaline Gerardin, Nishith R. Reddy, Wendell A. Lim

Correspondence

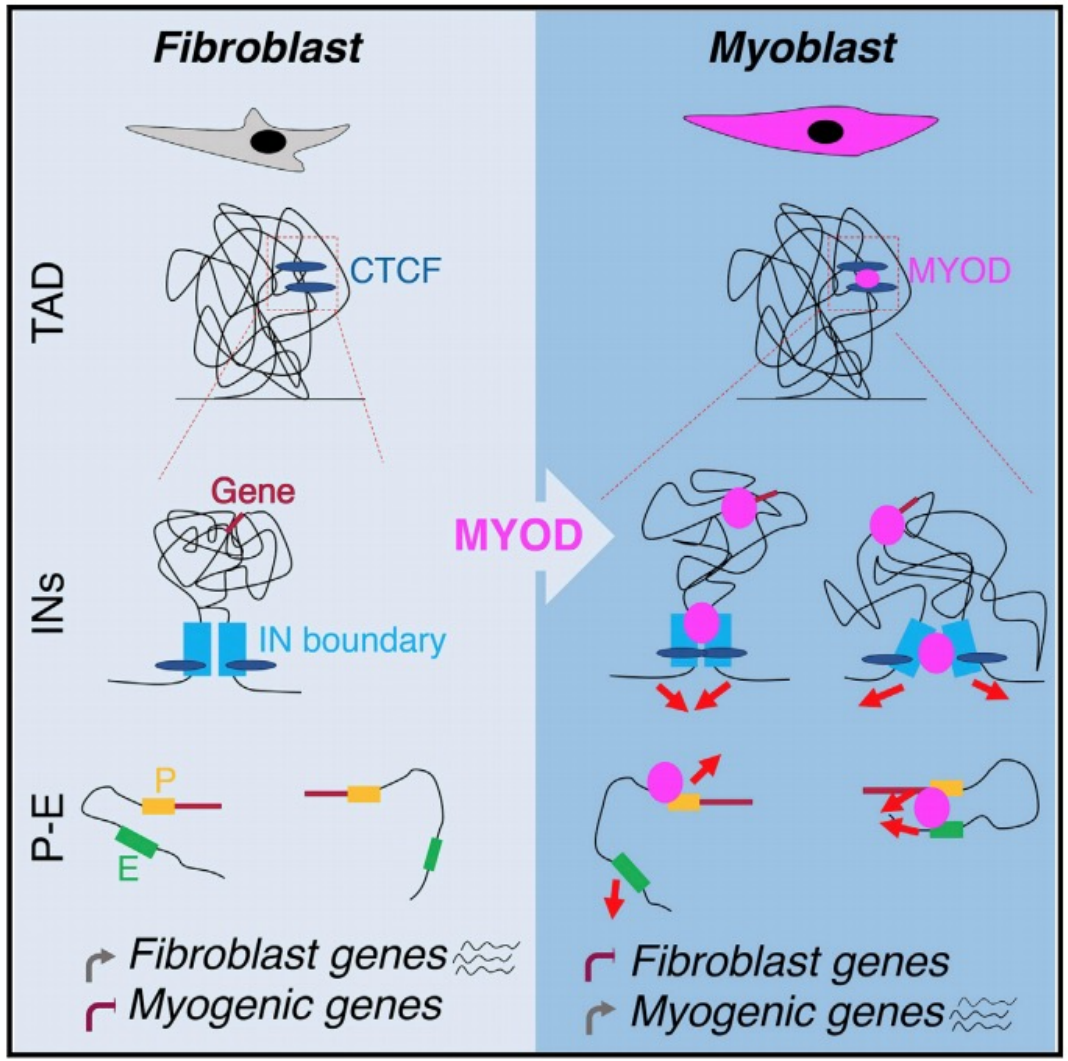
wendell.lim@ucsf.edu

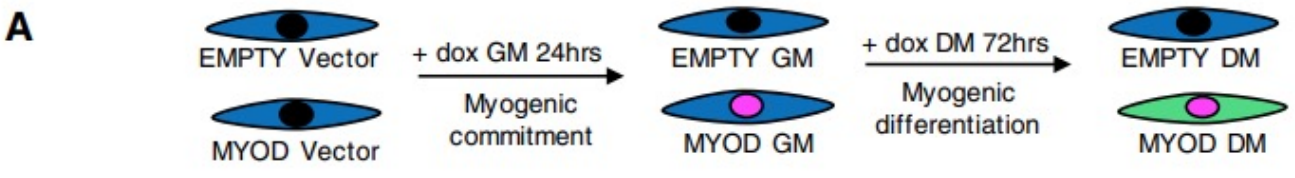
In Brief

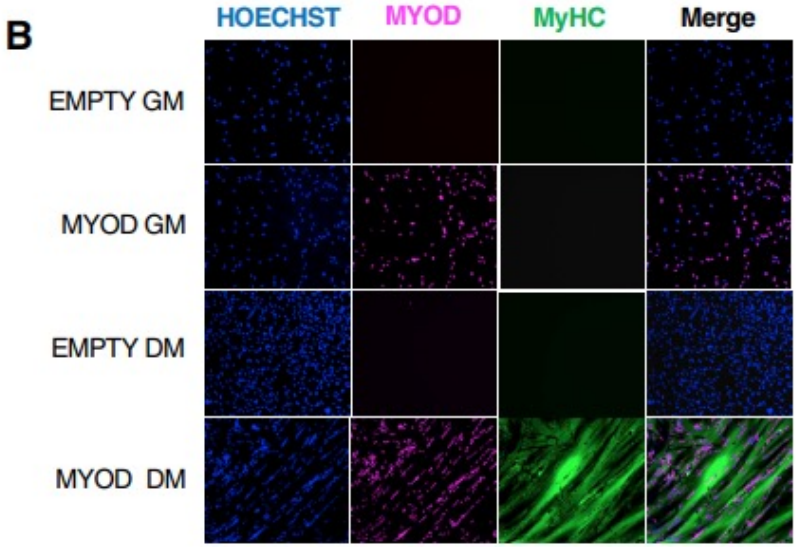
Timing is critical in biological regulation. In many cellular processes, biochemical networks can measure the duration of signaling inputs to coordinate the relative timing of cellular responses. To define biochemical circuits capable of this temporal filtering, we comprehensively searched the space of three-node biochemical networks. We identified five classes of core network motifs capable of temporal filtering with distinct functional properties and mechanisms. These core network motifs provide insight into how cells can interpret dynamic information

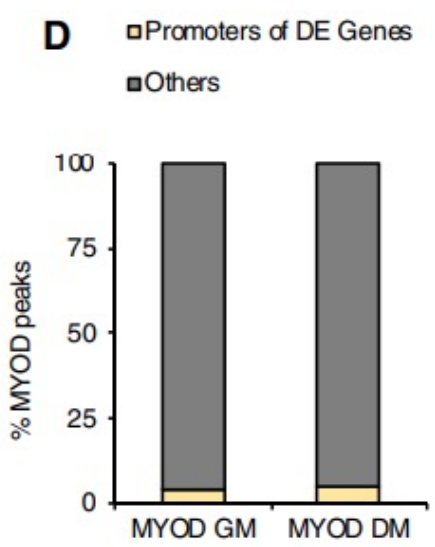


You can tell just by looking at figures…





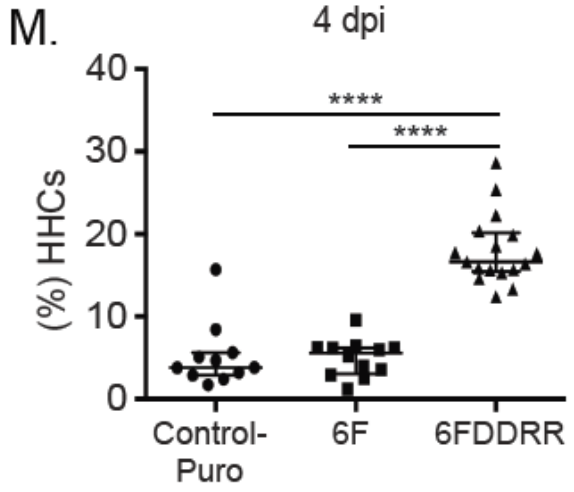




This is extremely important data…

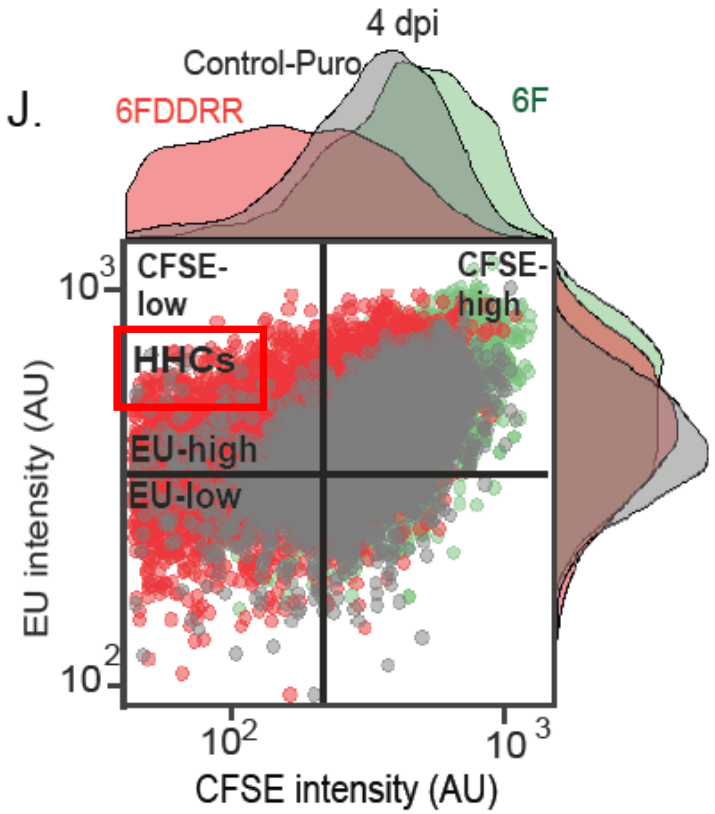
But alone this graph raises a lot of questions:

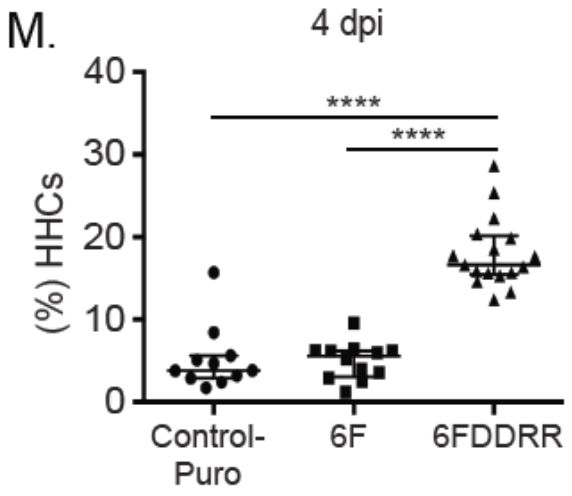
Why do I care what an HHC is?
Is it a rare type of pokémon?
Where can I find this creature?



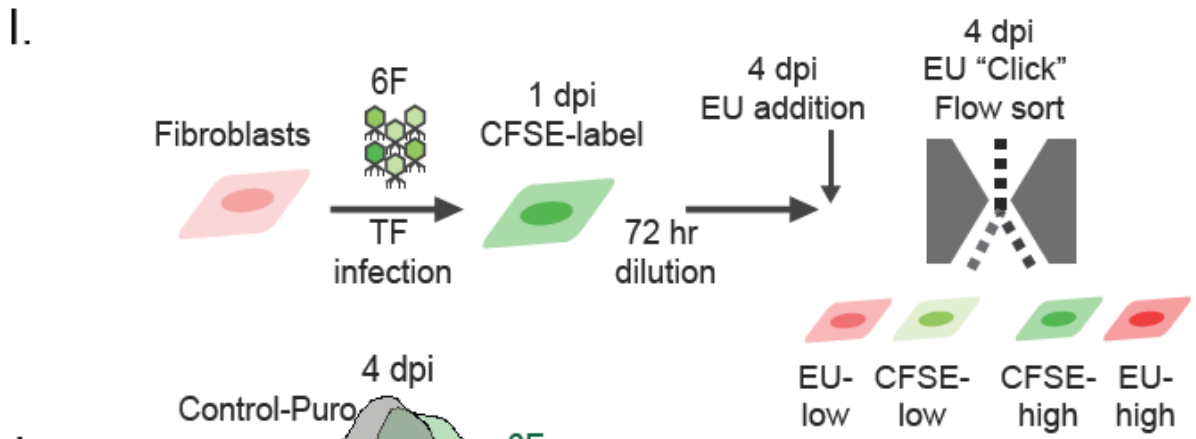
Best practice: Graphic of primary data (J) to match with processed data (M)

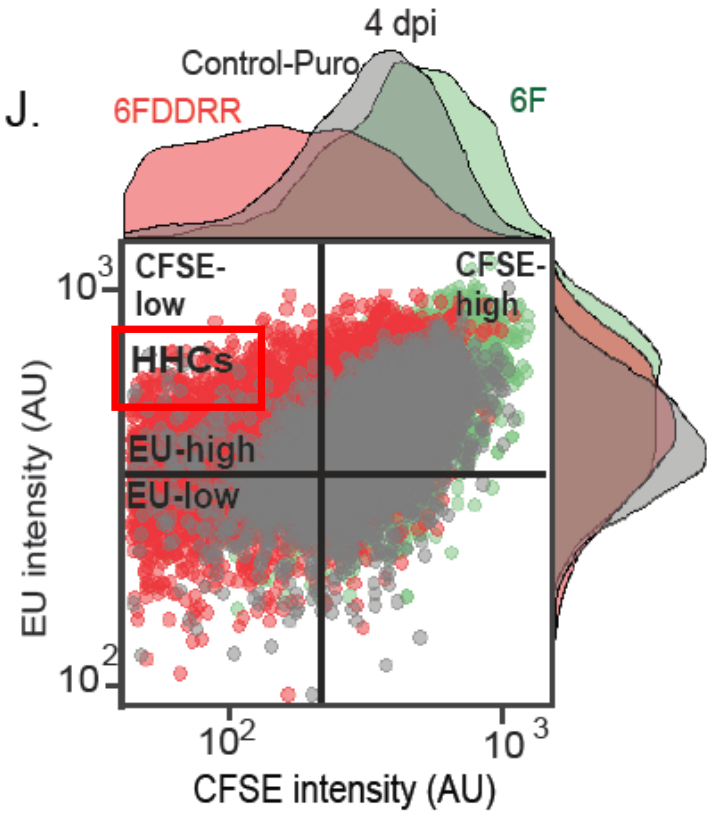
Oh there it is, it's that population in the top-left quad!
But, still how do we identify it? Is that a cell?

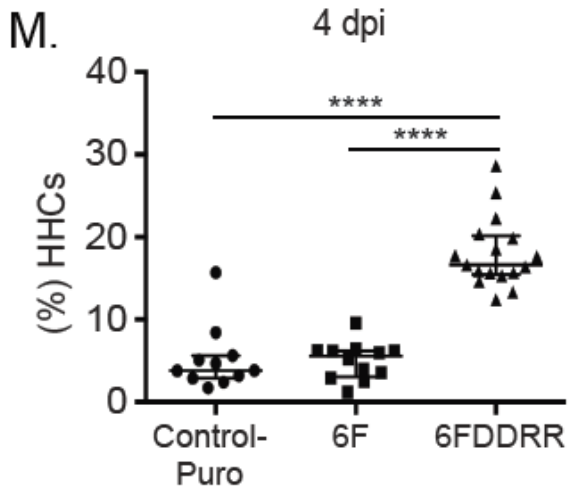




Diagraming complex processes is essential







By adding in a graphic the reader can understand what process was used to identify the cells as HHCs.

Without a descriptive legend, this is still nonsense to most readers

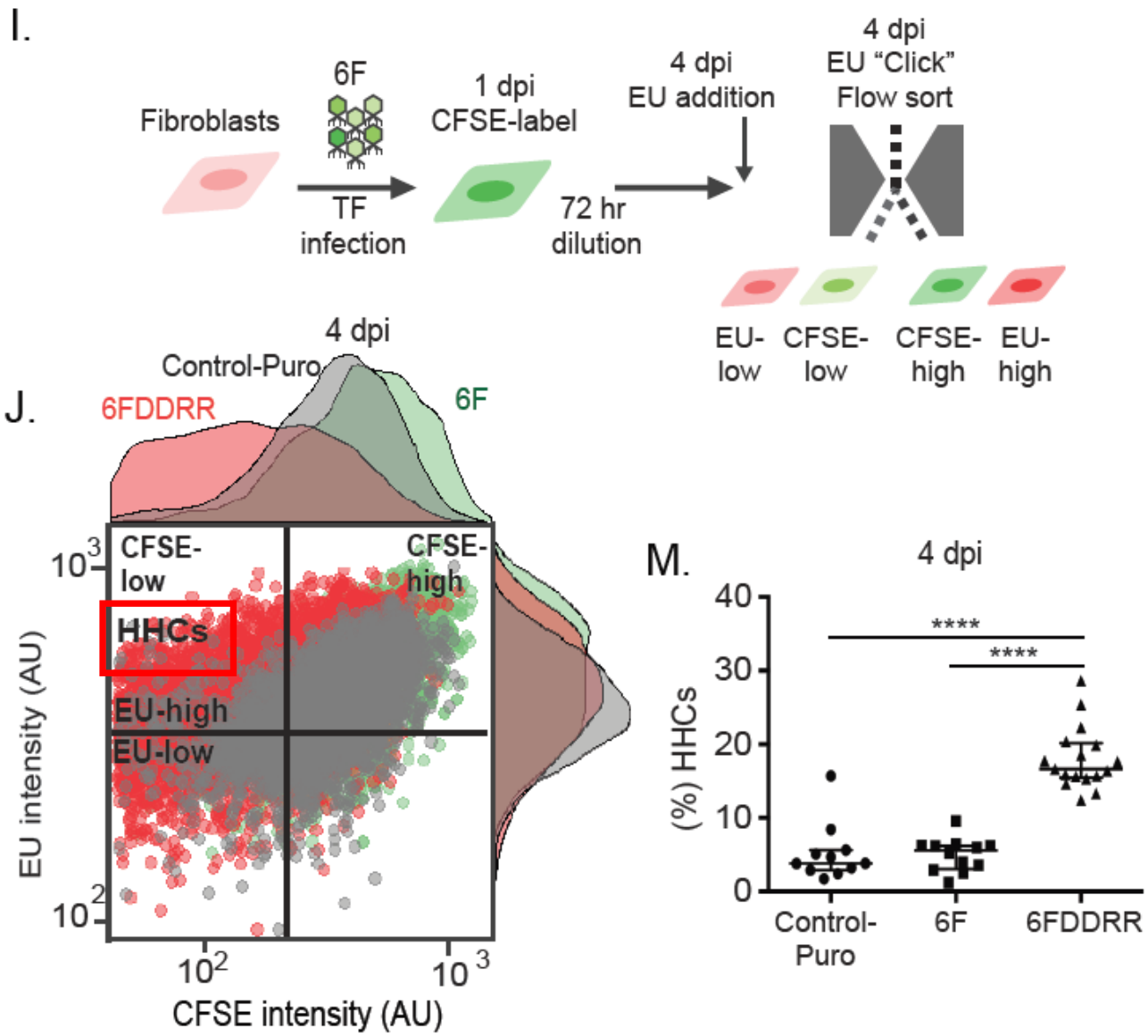
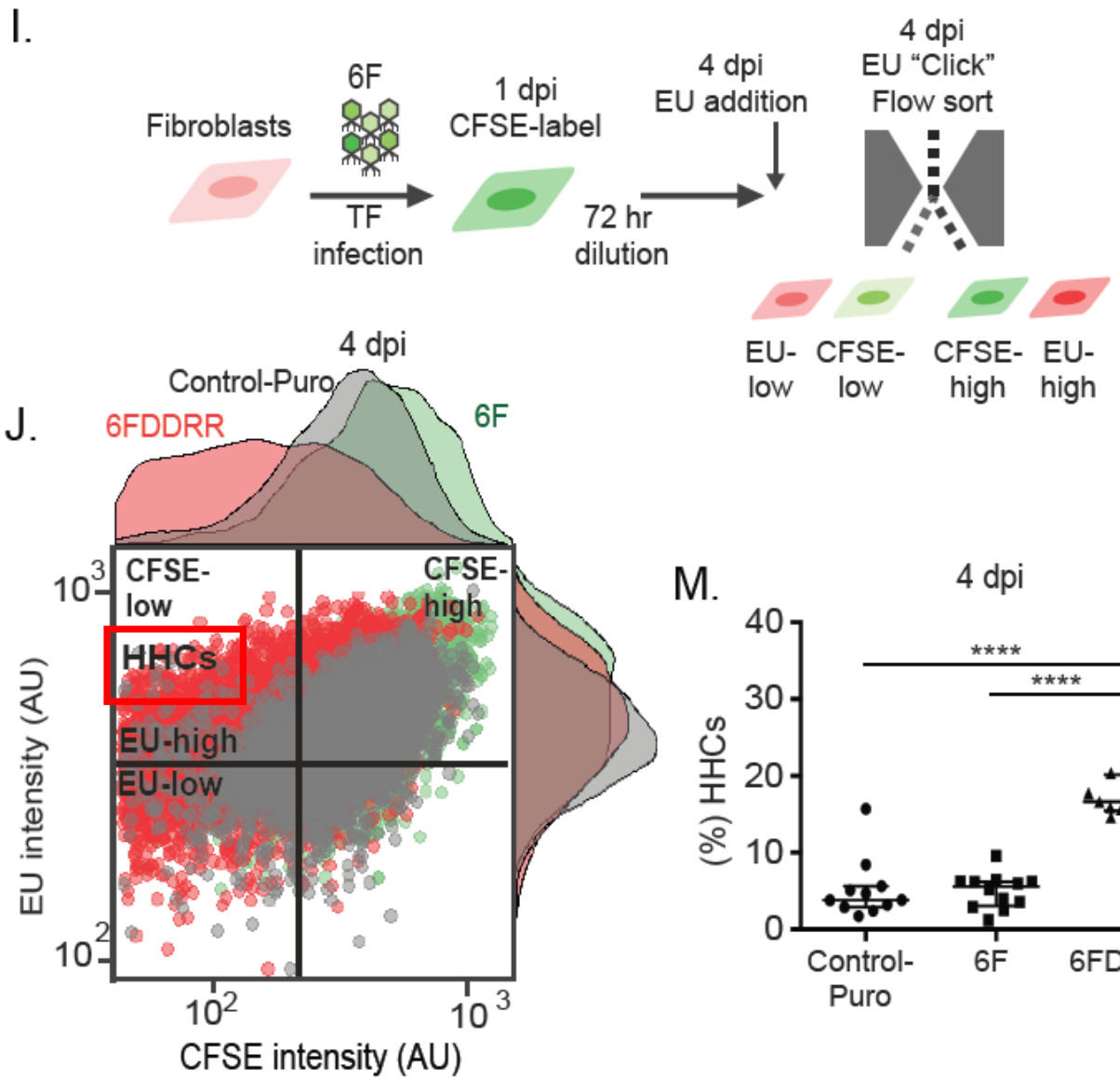


Figure legend. (I) Schematic of CFSE-EU assay for measuring transcription and proliferation rates in converting cells via flow cytometry at 4 dpi. Transcription rates measured through 5-ethynyl uridine (EU) incorporation during 1 hr incubation with 1mM EU followed by "click" reaction with fluorescent dye to visualize EU incorporation. CFSE assay performed as described previously.

(J) Representative dot plot of CFSE intensity and fluorescently labeled-EU for Control-Puro (grey), 6F (green), and 6FDDRR (red). Histograms of CFSE and EU intensity adjacent to dot plot. Quadrant to demark hypertranscribing, hyperproliferating cells (HHCs) set by reference to 6F condition. Hyperproliferating and slow cycling cells set by selecting CFSE value in 6F condition to allow the dimmest 15%. High EU values set by top half of 6F condition, resulting in ~7% HHCs in 6F.

(M) Percentage of HHCs for Control-Puro, 6F, and 6FDDRR condition as assayed. n = 11-16 independent conversion per condition. Median +/- interquartile range. Kruskal-Wallis Test.

Summary of what a figure needs



1. **A single message:**
   HHCs increase with DDRR

2. **Diagrams of experimental context**
   In cells? Flow? Time scale?

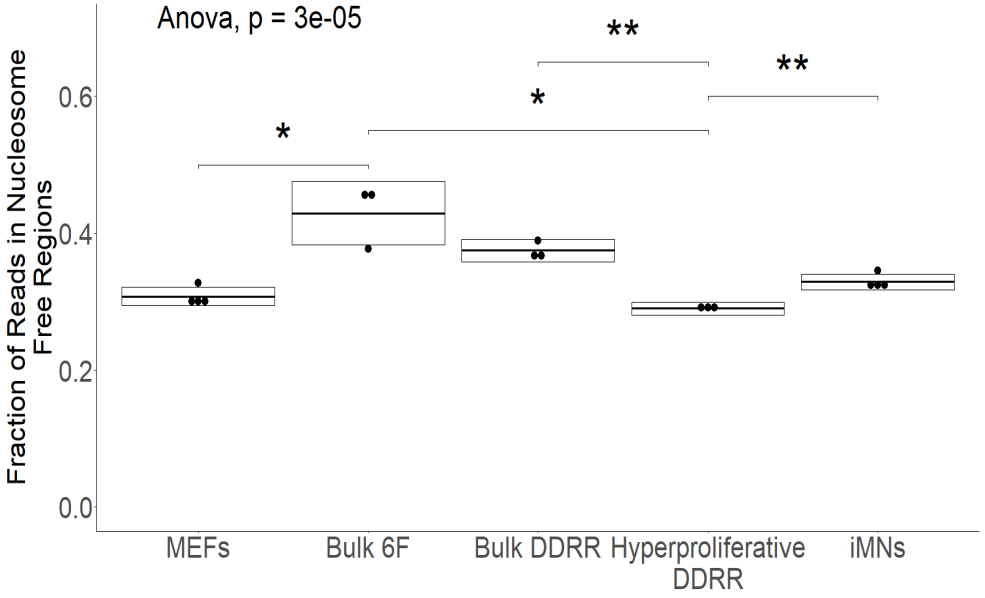
3. **Most primary form of data**
   Flow plots

4. **Processed data**
   %HHCs, statistical analyses

5. **A legend describing WHAT is in the figure.**
   Just the what (not results themselves! That is for the main text).

Example: From data to figure

Data for analysis → Message → Figure



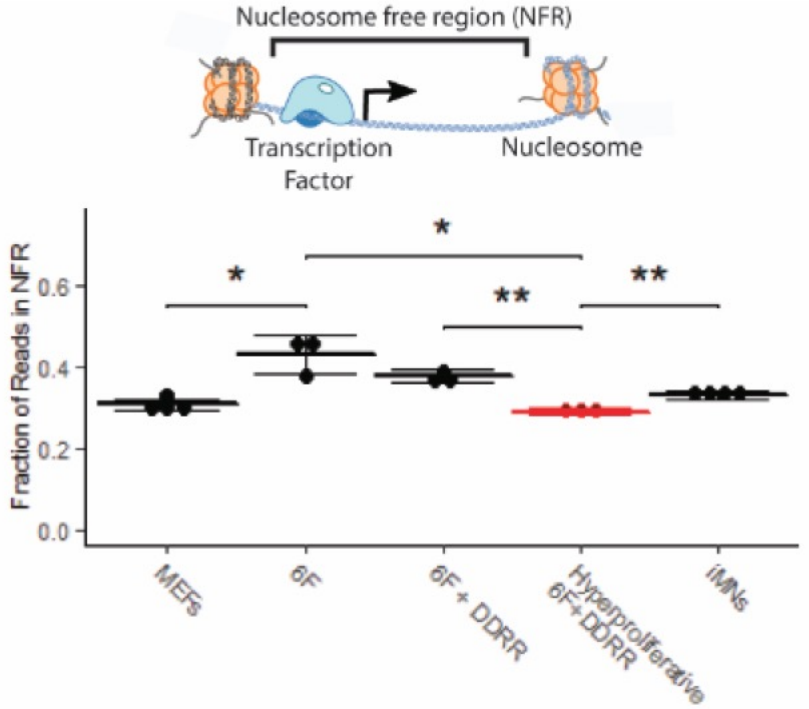
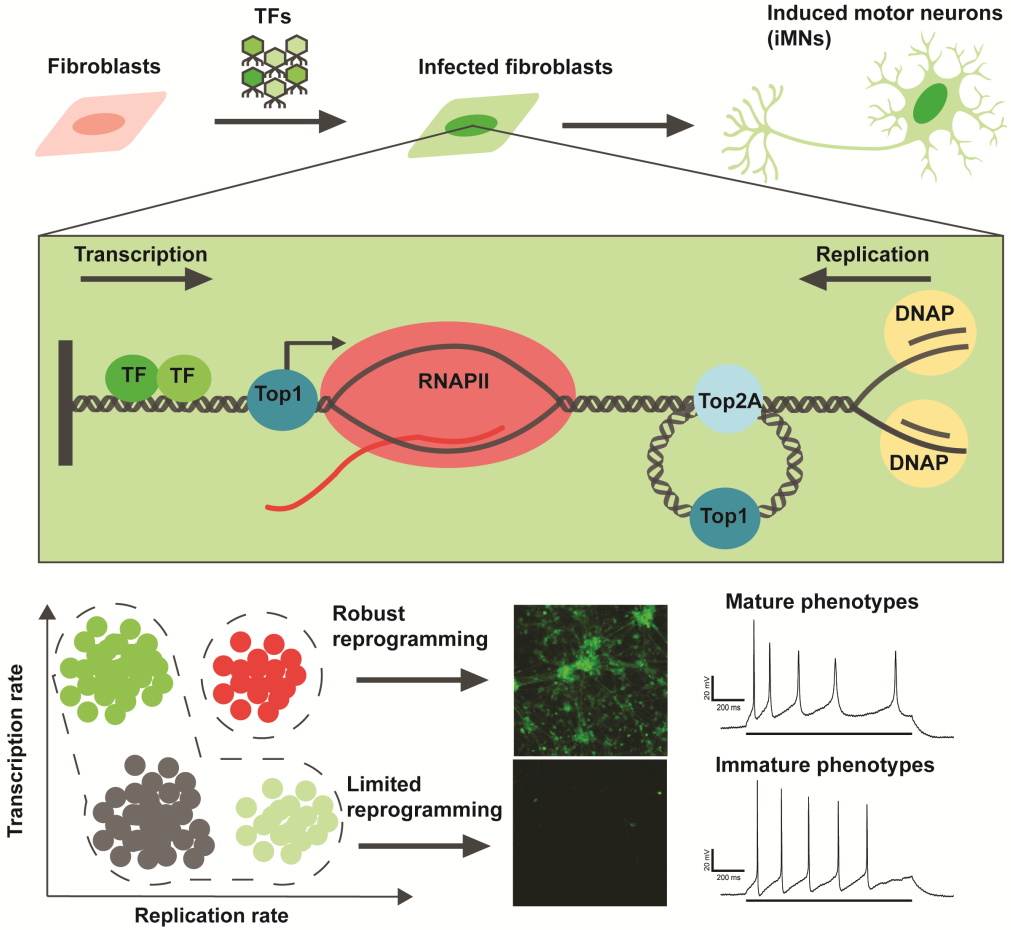
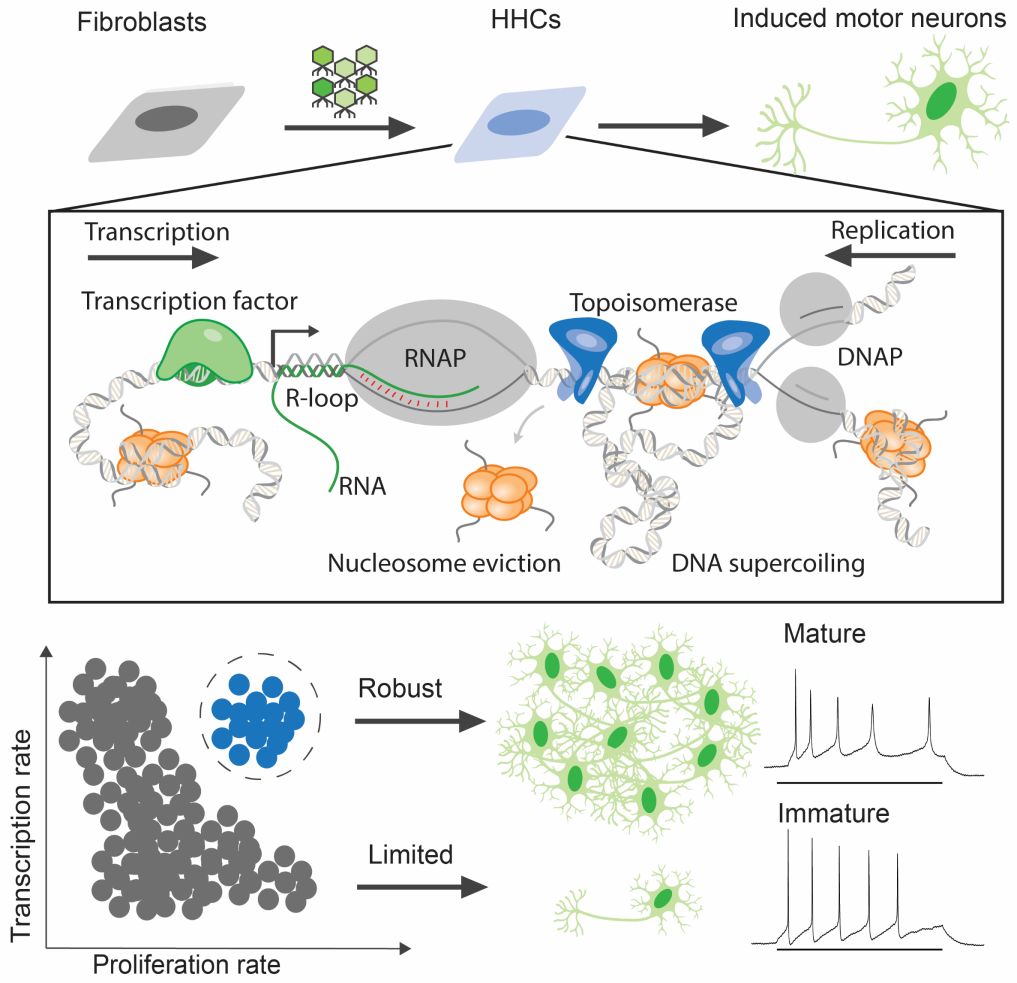
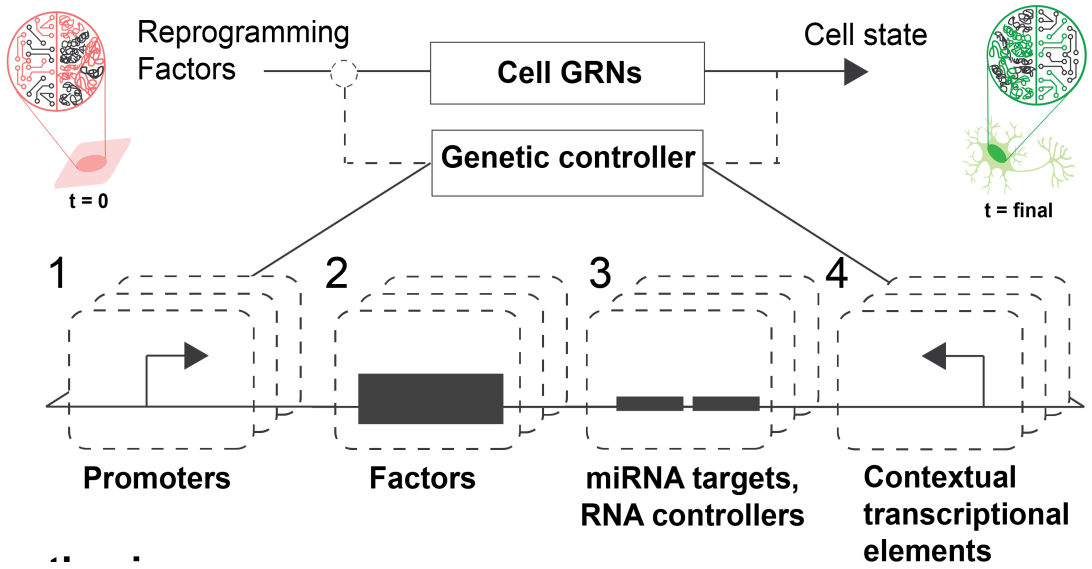
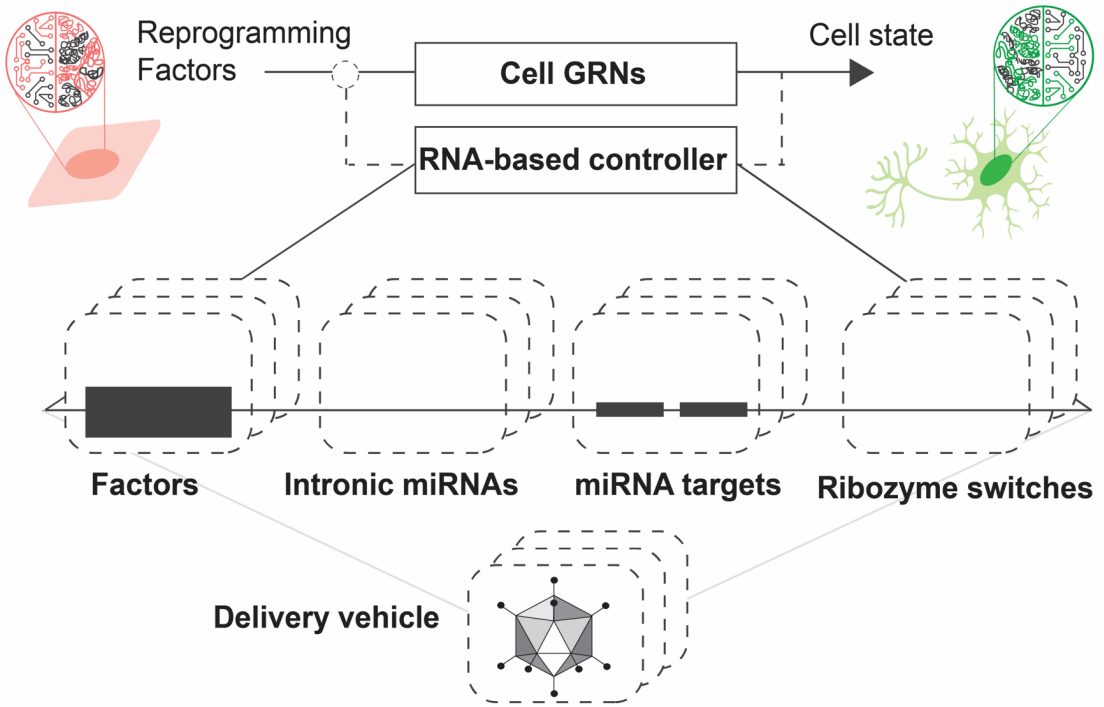


Figure 3: Nucleosome free regions vary by cell populations


Fibroblasts
TFs
Infected fibroblasts
Induced motor neurons
(iMNs)
Transcription
Replication
TF
TF
Top1
RNAPII
Top2A
Top1
DNAP
DNAP
Transcription rate
Replication rate
Robust
reprogramming
Limited
reprogramming
Mature phenotypes
Immature phenotypes


Fibroblasts
HHCs
Induced motor neurons
Transcription
Replication
Transcription factor
RNAP
R-loop
RNA
Topoisomerase
DNAP
Nucleosome eviction
DNA supercoiling
Transcription rate
Proliferation rate
Robust
Limited
Mature
Immature


Reprogramming
Factors
Cell GRNs
Cell state
Genetic controller
t = 0
t = final
1
2
3
4
Promoters
Factors
miRNA targets,
RNA controllers
Contextual
transcriptional
elements


Reprogramming
Factors
Cell GRNs
Cell state
RNA-based controller
Factors
Intronic miRNAs
miRNA targets
Ribozyme switches
Delivery vehicle

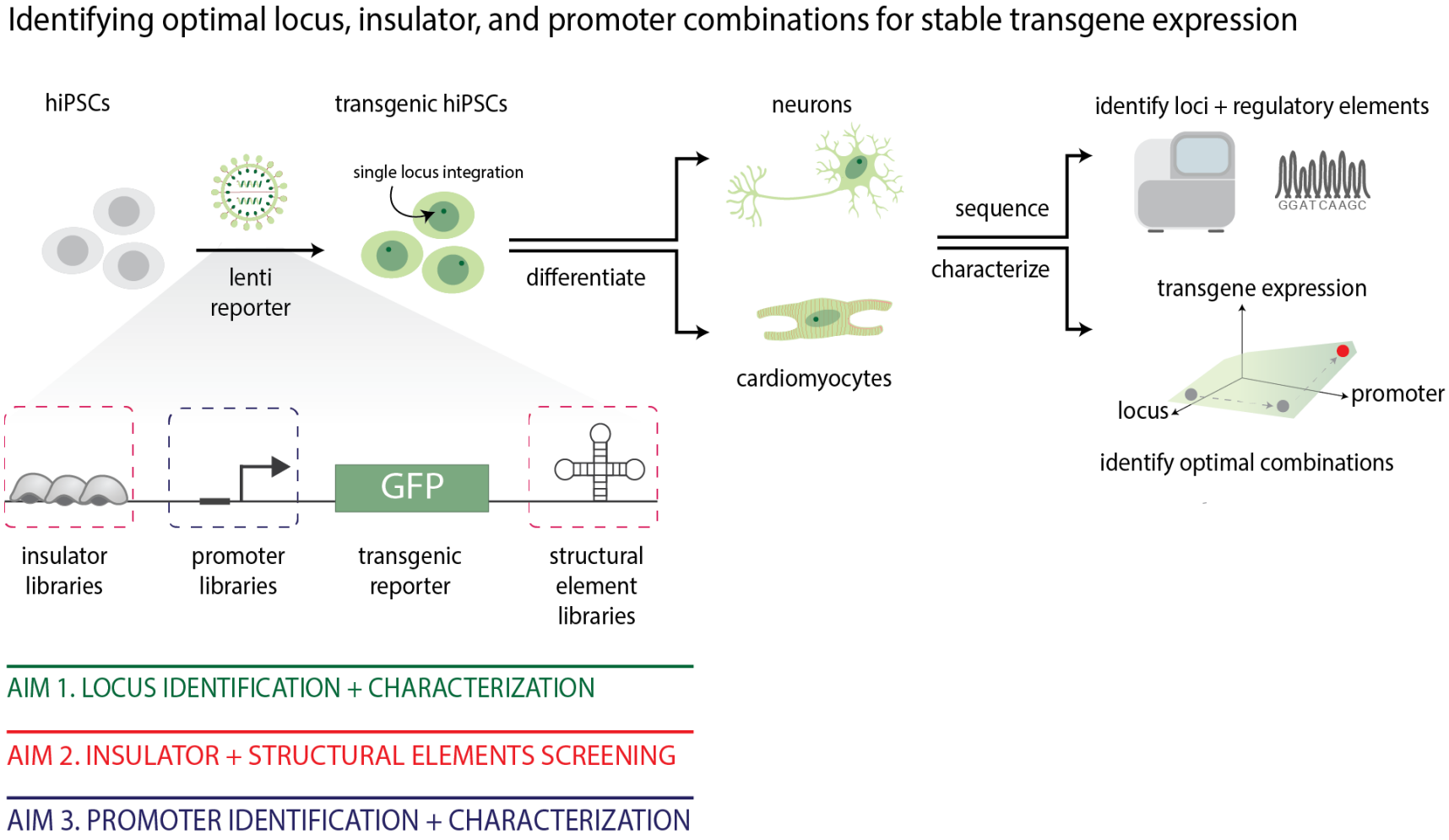


Figure 1. Developing a robust framework for genomic engineering in models of mammalian development. Human induced pluripotential stem cells (hiPSCs) are infected using libraries of lentiviruses bearing the green fluorescent protein (GFP) to generate transgenic hiPSCs. Libraries of integrating lentiviruses are composed with varying promoters, insulators, and structural elements. To ensure single locus integration, cells are infected at low multiplicity of infection and sorted by reporter activity (e.g. GFP-positive). GFP-positive transgenic hiPSCs are differentiated into neurons and cardiomyocytes. Loci and regulatory elements including promoters, insulators, and structural elements that maintain transgene activity are identified by sequencing GFP-positive differentiated cells. Following identification of loci and regulatory elements, optimal combinations for each cell-type will be verified through differentiation of iPSCs engineered via targeted CRISPR-based integration of candidate regulatory element combinations at identified loci.

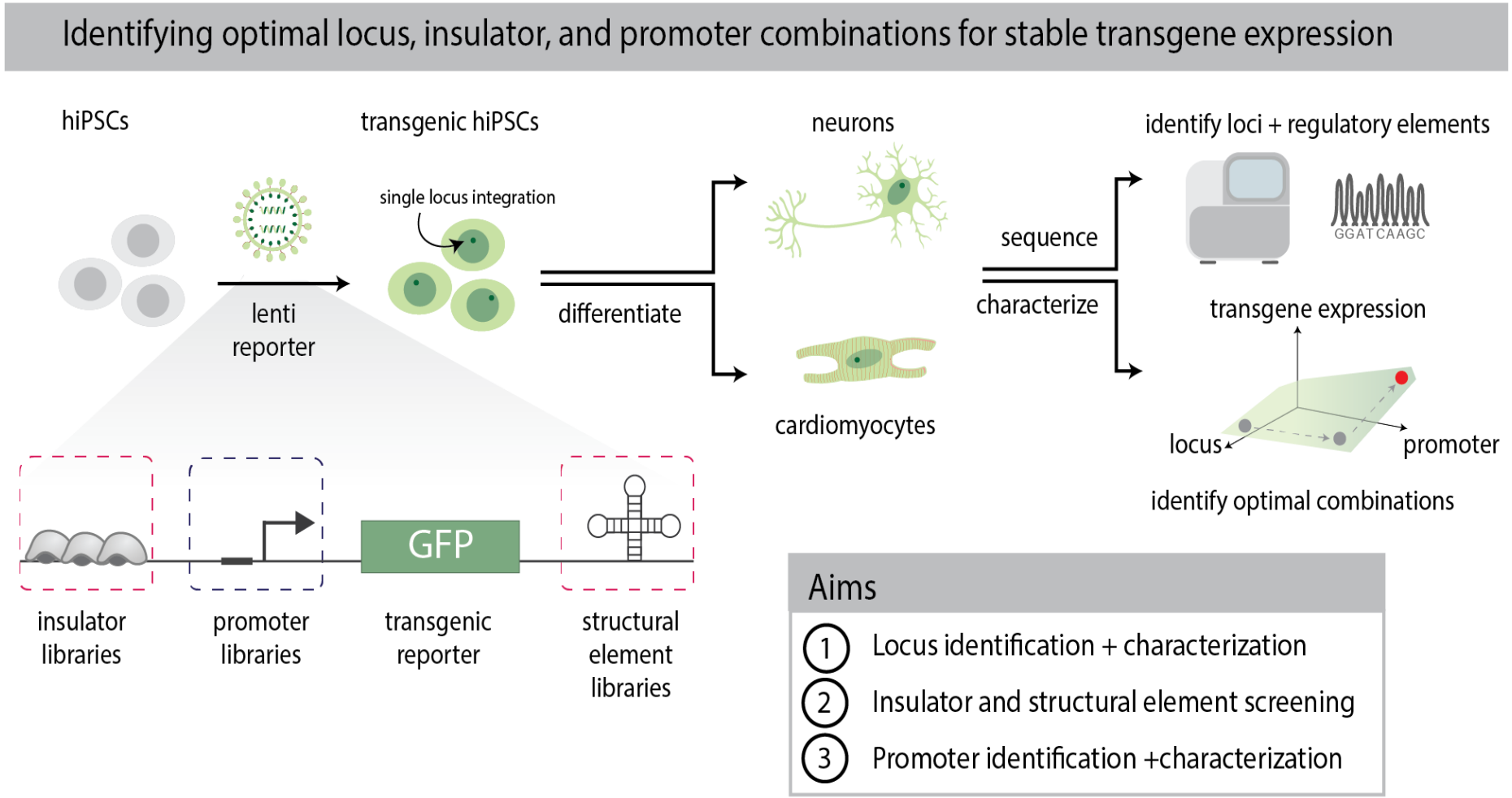


Figure 1. Developing a robust framework for genomic engineering in models of mammalian development. Human induced pluripotential stem cells (hiPSCs) are infected using libraries of lentiviruses bearing the green fluorescent protein (GFP) to generate transgenic hiPSCs. Libraries of integrating lentiviruses are composed with varying promoters, insulators, and structural elements. To ensure single locus integration, cells are infected at low multiplicity of infection and sorted by reporter activity (e.g. GFP-positive). GFP-positive transgenic hiPSCs are differentiated into neurons and cardiomyocytes. Loci and regulatory elements including promoters, insulators, and structural elements that maintain transgene activity are identified by sequencing GFP-positive differentiated cells. Following identification of loci and regulatory elements, optimal combinations for each cell-type will be verified through differentiation of iPSCs engineered via targeted CRISPR-based integration of candidate regulatory element combinations at identified loci.

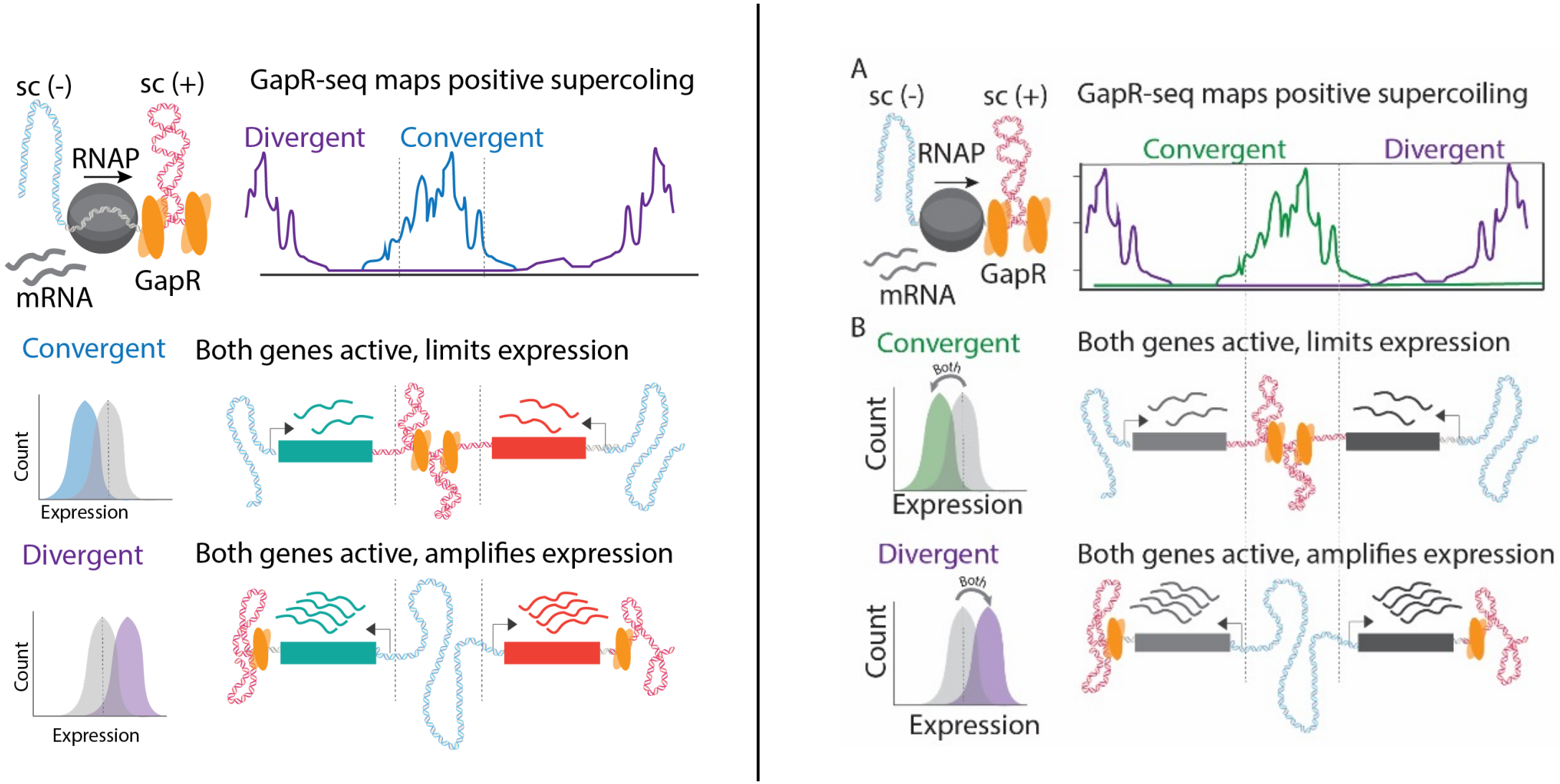


Figure 1. A. RNAP induces DNA supercoils as mRNA is transcribed. GapR protein binds to positively supercoiled DNA (sc(+)) enabling GapR-seq to map supercoiling across genes. B. As predicted by models of supercoiling-mediated gene regulation, convergently oriented genes display reduced expression when both genes are active. Divergently oriented genes show amplification.

References and further reading

Ten Simple Rules for Better Figures

Broad Comm Kit Figure Design

The misuse of colour in science communication

Creating Clear and Informative Image-based Figures for Scientific Publications

A Brief Guide to Designing Effective Figures for the Scientific Paper

Nature Data Viz Guide

The Visual Display of Quantitative Information

Trees, Maps, and Theorems Jean-Luc Doumont