CSC2541H / (PCL3107H, PCL3108H): AI for Drug Discovery

Phenomics

Chris J. Maddison
University of Toronto

Learning Objectives

In this lecture, we will:

1. Understand the challenge of causal inference in cell biology and why interventional experiments are critical
2. Know what cell painting is as an assay, its role as a cheap high content assay for drug discovery, and the importance of high content screen, and the major cell painting datasets
3. Introduce neural networks and self-supervised methods
4. Bring it all together with phenomics models that are trained and applied to predict intervention effects

Motivating Question

Motivating question: "Will my experimental agent resolve a pathology and result in a healthy phenotype in a certain population?"

Experimental agents:

- Small molecules
- Biologics
- Oligonucleotides

Pathologies:

- Oncogenic transformation
- Metabolic dysfunction
- Neurodegeneration

Populations:

- Cell lines
- Primary cells
- Patient-derived xenografts

Key distinction we will explore in the context of this question: correlation vs. causation

Running Example

To be more concrete: "How does intervening on the expression of gene B change the expression of gene A in a certain cell line?"

For example:

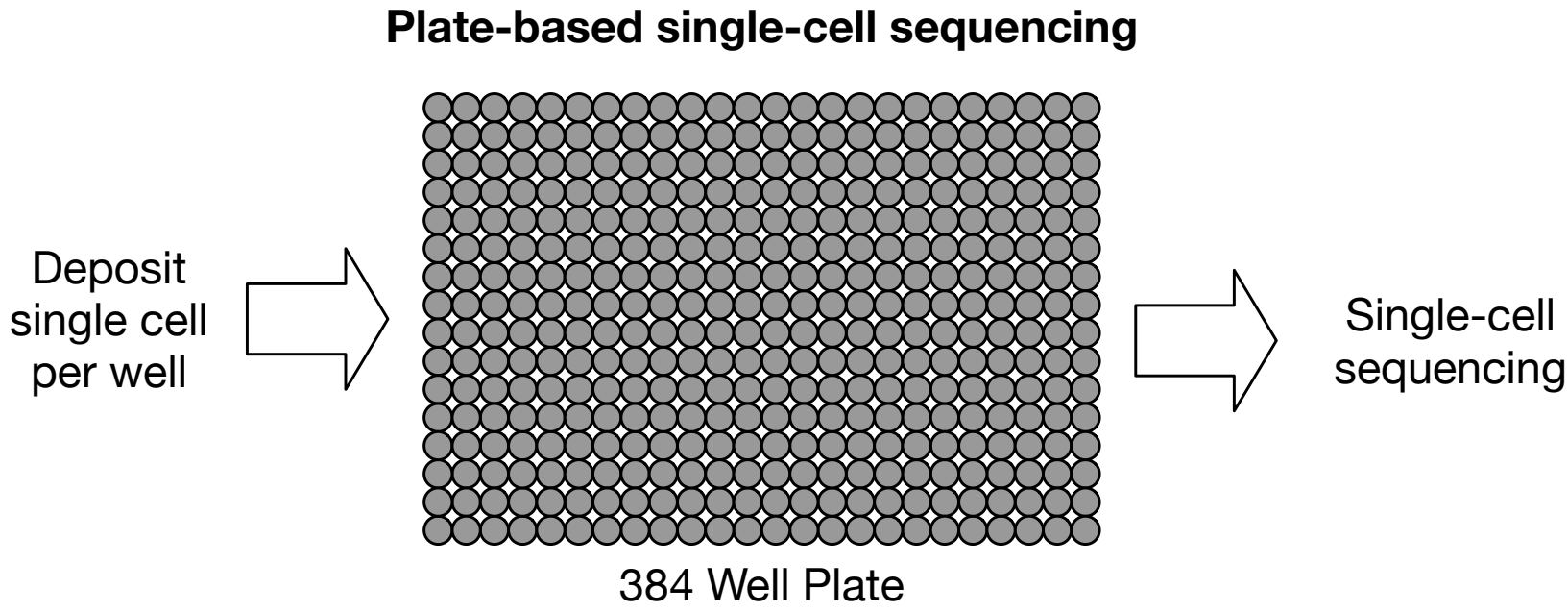
- **Gene A:** MYC (oncogene driving uncontrolled cell proliferation)
- **Gene B:** KRAS (upstream signaling gene, mutated in many cancers)
- **Experimental agent:** siRNA knockdown of KRAS
- **Pathology:** Oncogenic transformation in lung cancer
- **Cell line:** A549 cells

We will focus on cell biology in this lecture, but keep this running example abstract.

Question

What sort of experimental approaches can we take to answer our question?

RNA-Seq



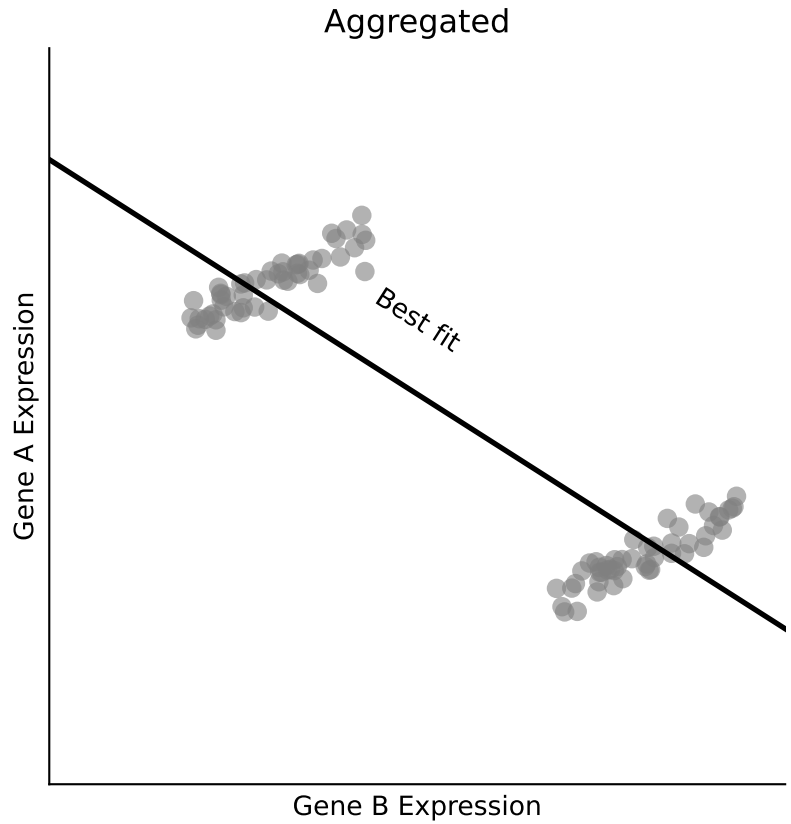
The scenario:

- We perform single-cell RNA-seq on cells in culture
- We analyze the correlation between Gene A and Gene B across all cells and find a negative correlation

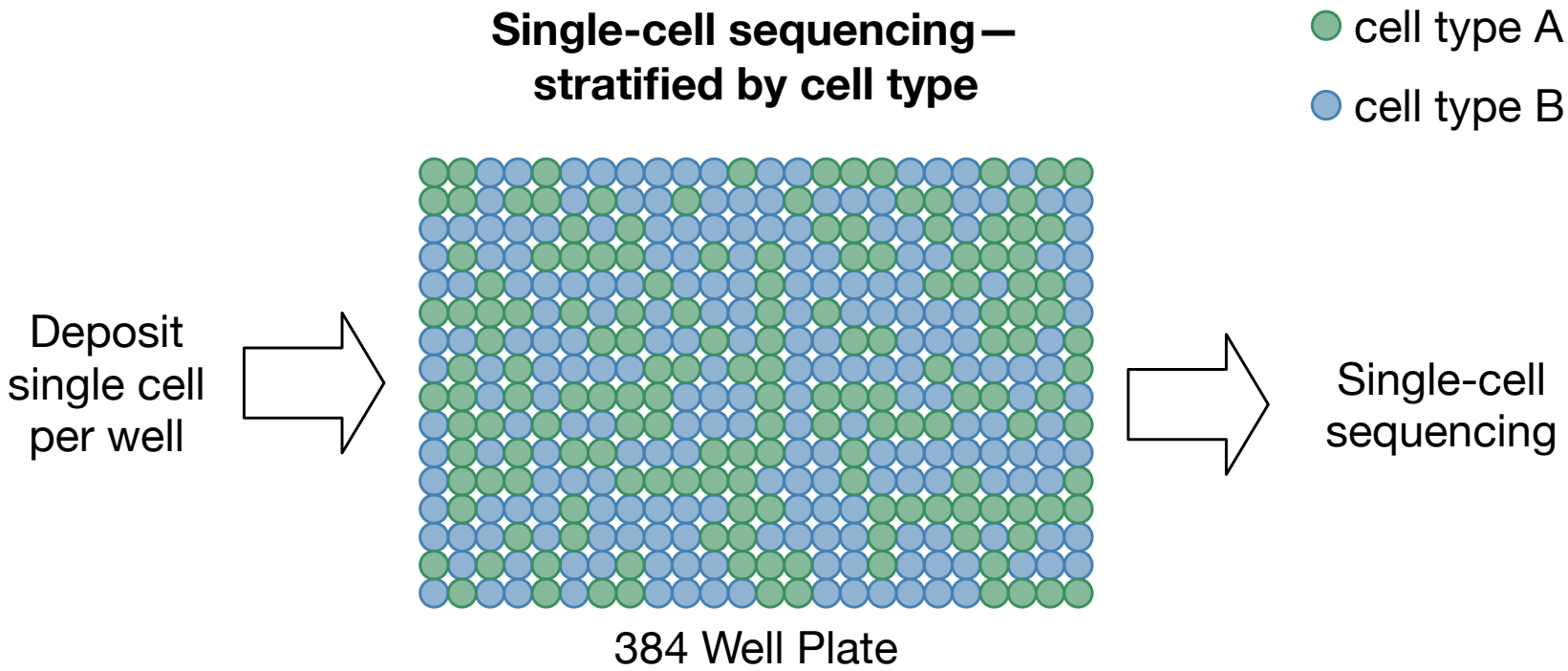
Question: If we intervened to reduce Gene B expression, would Gene A increase?

Aggregated Analysis

- Each point is a single cell
- Gene A and Gene B appear negatively correlated
- **Conclusion:** Genes are mutually exclusive—knocking down B should increase A



Stratified by Cell Type

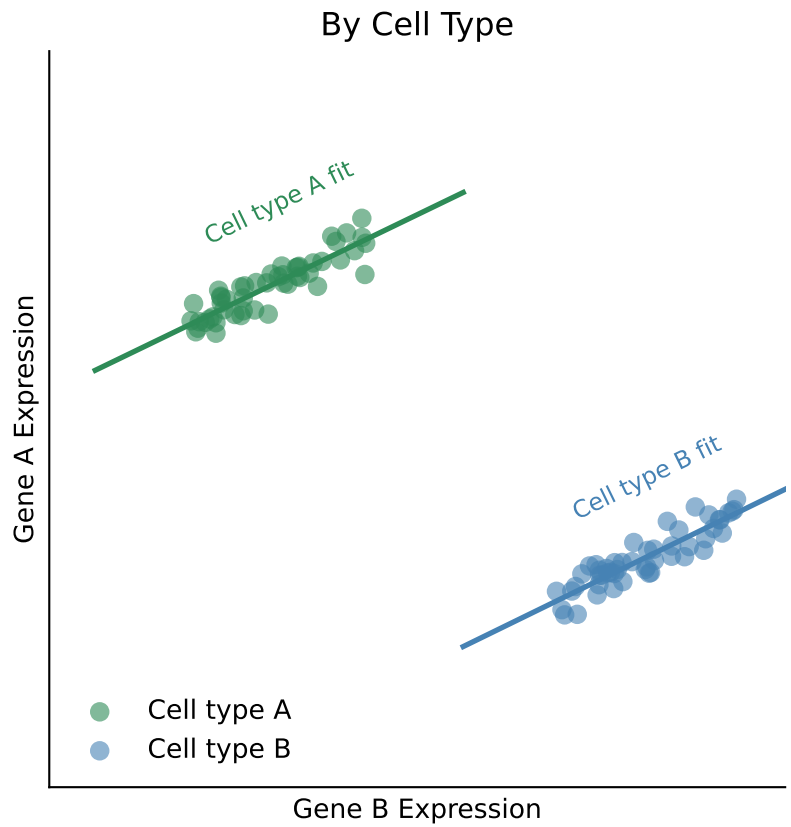


- Suppose that our culture actually has two cell types (unbeknownst to us) that we can detect with the expression levels of a third gene (eg Gene C)
- What if we stratified our data by cell type?

Stratified by Cell Type

Grouping by cell type reveals a different story:

- Same data, but now you stratify by another gene that determines cell type—two cell types
- Within each cell type, genes are *positively* correlated
- **New conclusion:** Gene B drives Gene A—knocking down B should *decrease* A

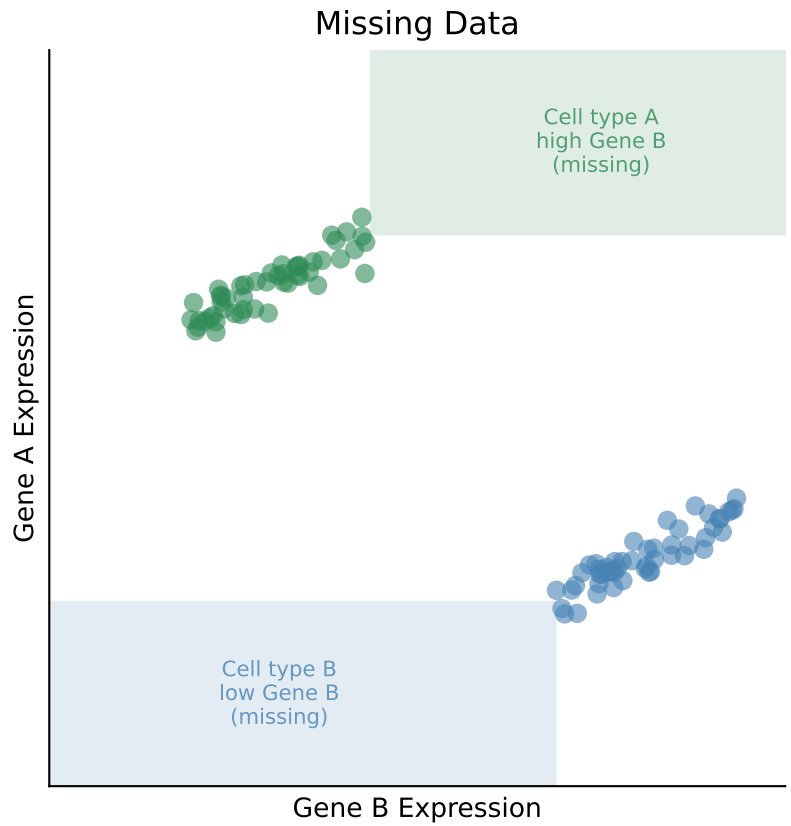


What is going on? Observational Data

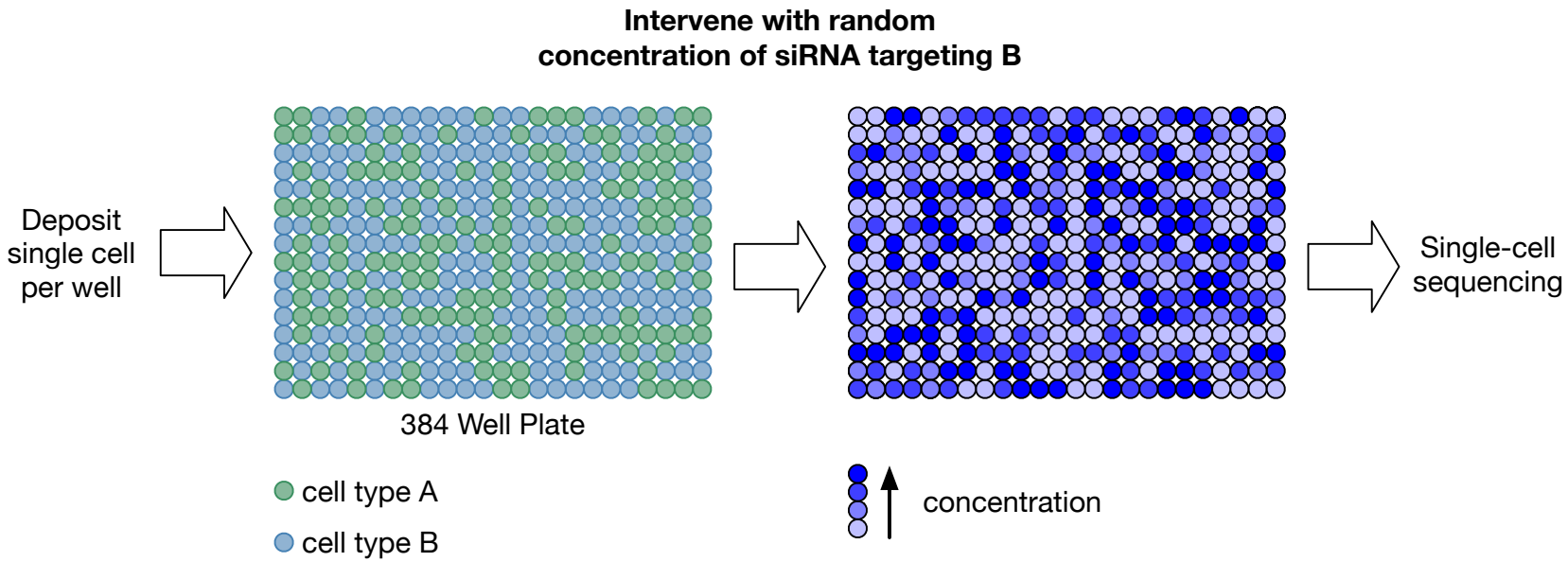
- We used **observational data**—data of the system in its natural state— and it can have unobserved subpopulations
- If those unobserved subpopulations have different baseline levels, it can **confound** our estimate of cause-effect—known as **Simpson's Paradox**

What is going on? Observational Data

- Baseline differences occur in nature—varying by cell type
- Some combinations don't occur (shaded regions)
- $\rightarrow$ conclusions from aggregated analysis can be **confounded by the differences in unobserved cell type** as opposed to the variable of interest (Gene B)



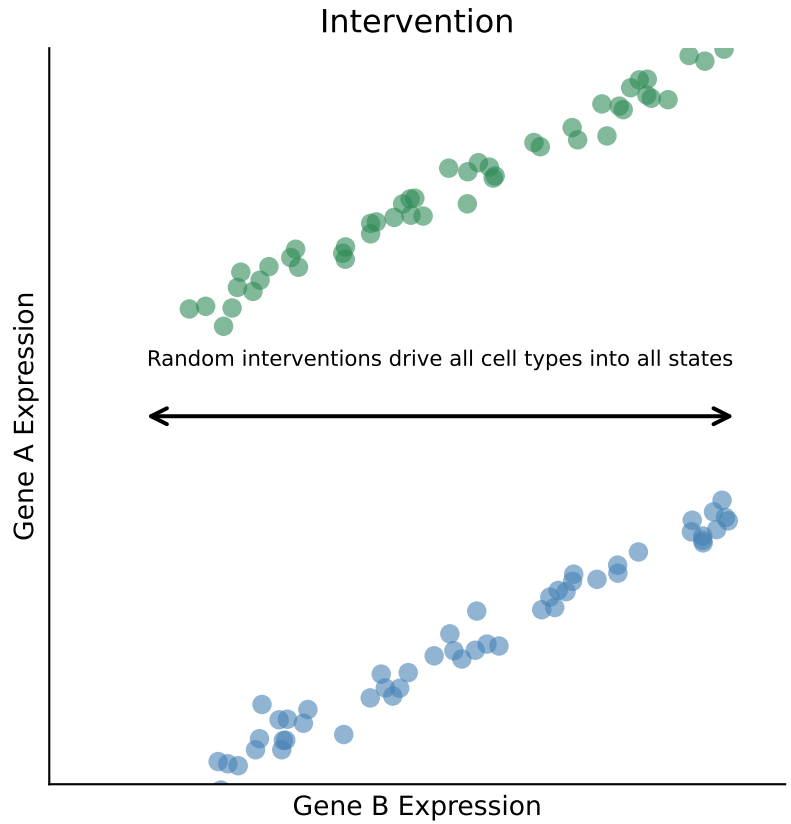
Inteventional Data



- To overcome this, we can **intervene** on the system to drive cell types into unnatural states
- By **randomly** assigning perturbations, we break confounding of unobserved variables in nature

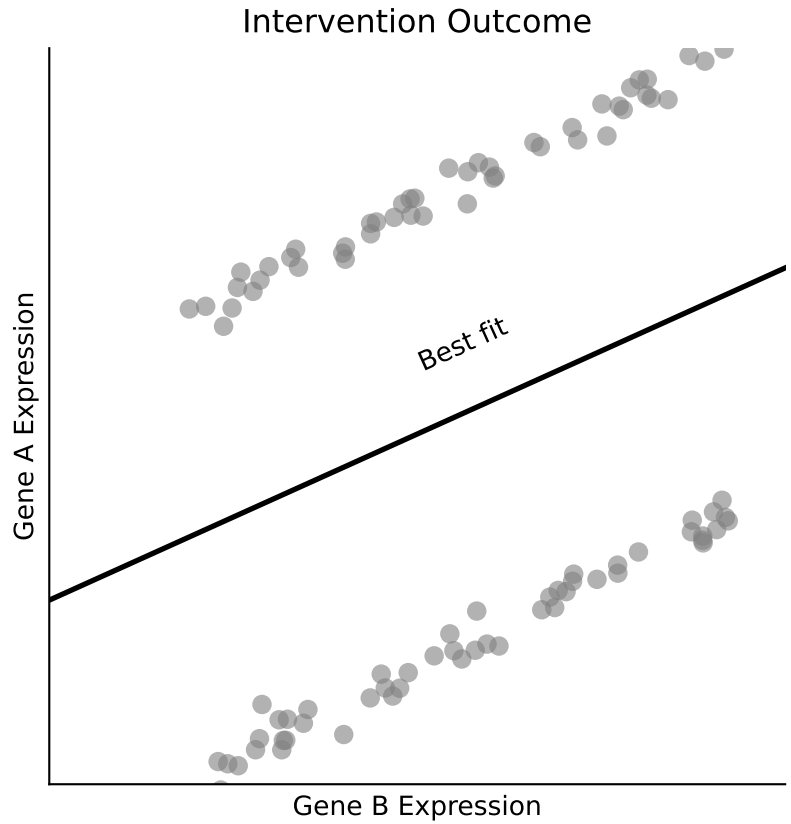
Inteventional Data

- Random interventions drive all cell types into all states—even if unnatural
- This gives us insight into mechanism



Interventions Resolve Simpson's Paradox

- Now there is a positive correlation—the stratified RNA-Seq data gave the right impression!
- **Key point:** Interventional experiments are necessary to understand cause-effect
    - This is a deep area of research in computer science, pioneered by Judea Pearl

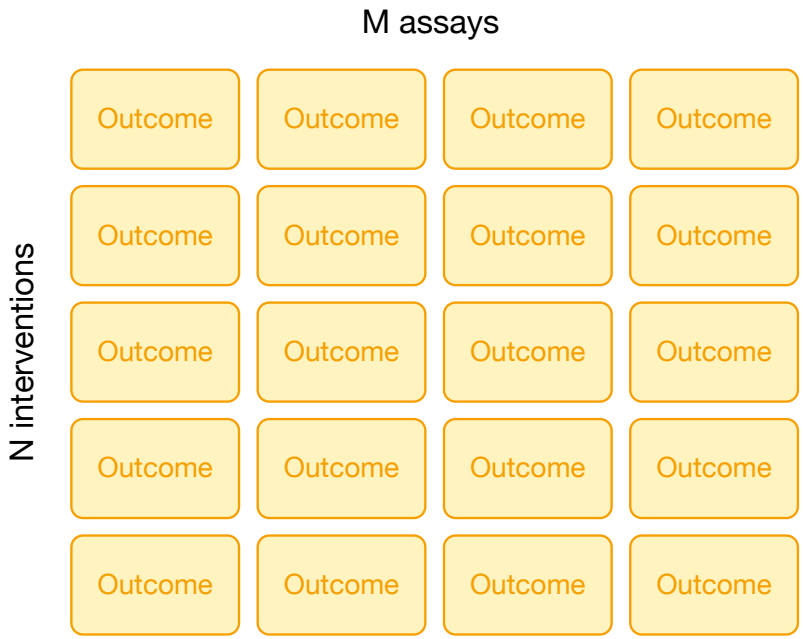


The Cost of Interventional Experiments

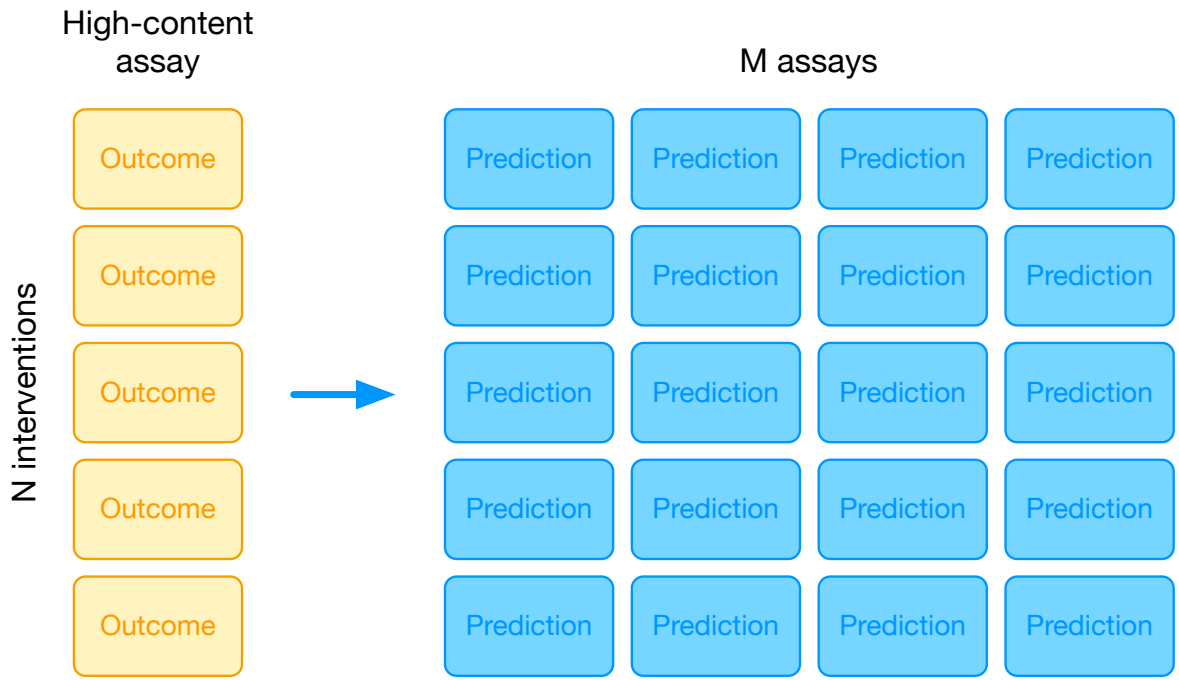
BUT: They are expensive

- Each intervention requires a separate experiment for each outcome of interest
- N interventions $\times$ M outcomes $=$ N$\times$M experiments

Question: Can we be more efficient?



A Better Idea: A High-Content Assay



- Run ONE assay per intervention that has rich outcomes
- **Use the rich outcome to predict many downstream assay outcomes**

A High-Content Assay

Desiderata:

- **Rich:** Contains information about many downstream outcomes
- **Scalable:** Cheap to measure at scale
- **Universality:** Applies broadly across cell types
- **Single-cell:** captures single-cell outcomes, not just bulk aggregates

Question: what do you think we should pick for a high-content assay?

Candidate High-Content Assays—Omics

Transcriptomics:
Measures mRNA levels across genes to capture the cell's transcriptional state.

- Pros: Interpretable at gene/pathway level
- Cons: Full RNA-seq is expensive; reduced assays (e.g., L1000: 1000 landmark genes) are cheaper but still dollars per sample

Proteomics:
Measures protein abundance via mass spectrometry.

- Pros: Closer to functional biology than mRNA
- Cons: Technically challenging, less scalable

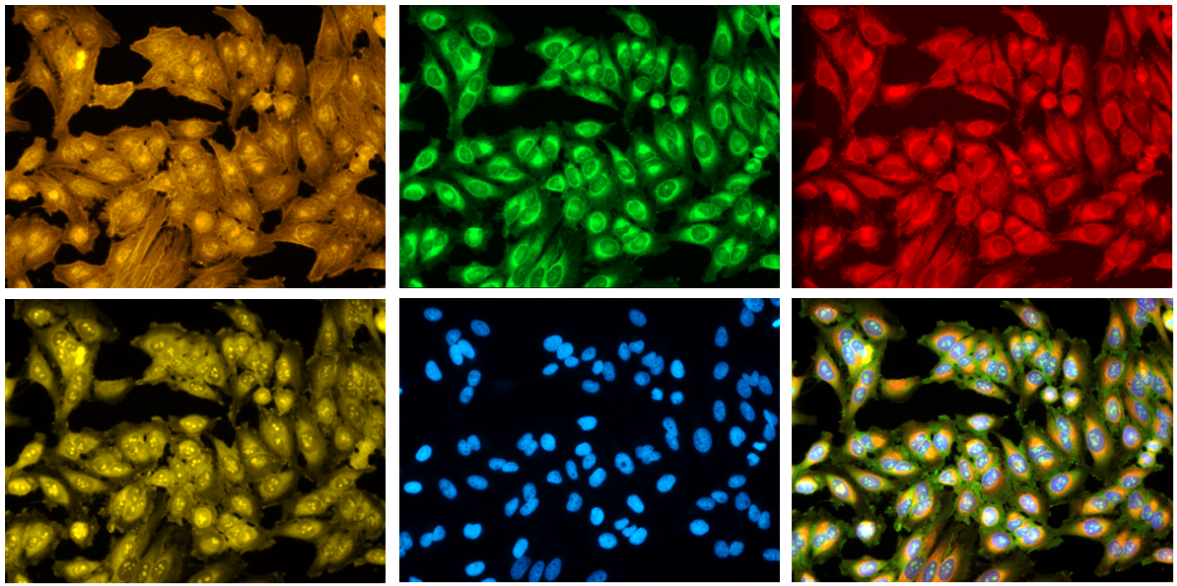
Metabolomics:
Measures small molecule metabolites in the cell.

- Pros: Captures metabolic state directly
- Cons: Complex sample prep, annotation challenges

Challenges with -omics

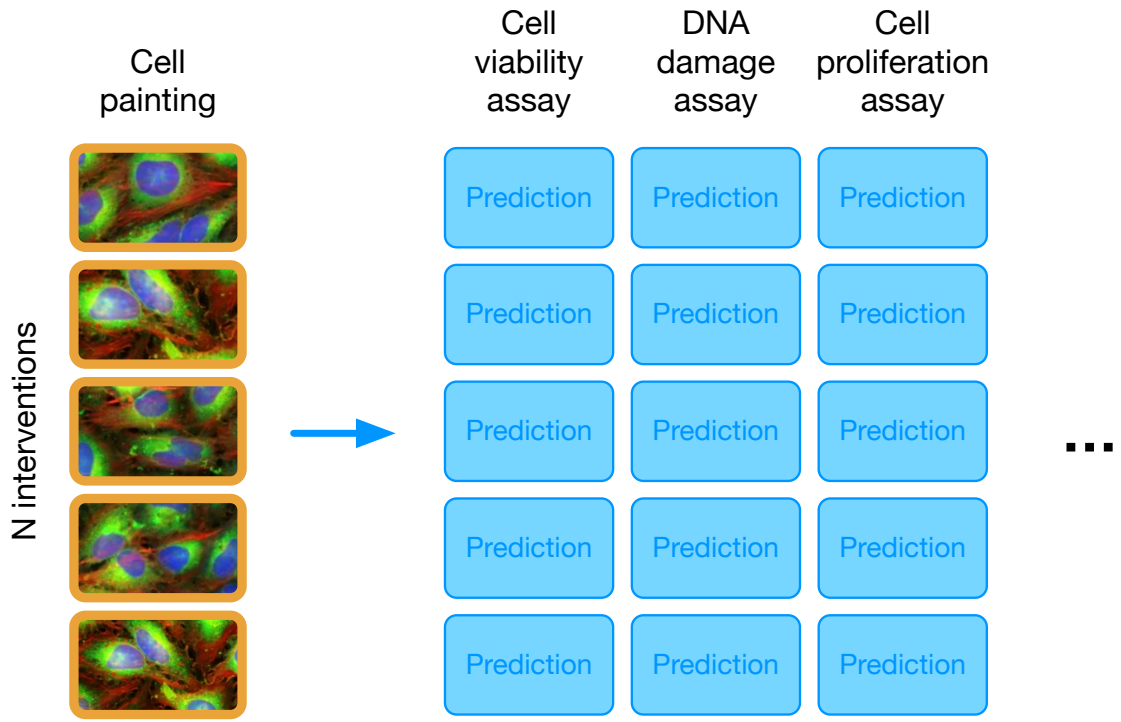
- Rich? If you measure lots of targets and lots of omics, but it's not clearly holistic unless you get
- Scalable? It's still quite expensive
- Universal? transcriptomics can give you a signal, but it may still be a very non-specific signal of the functional and phenotypic properties of cell types
- Single-cell? there's a tradeoff between single-cell and data depth (how many reads per target)

What about microscopy images of cell cultures?



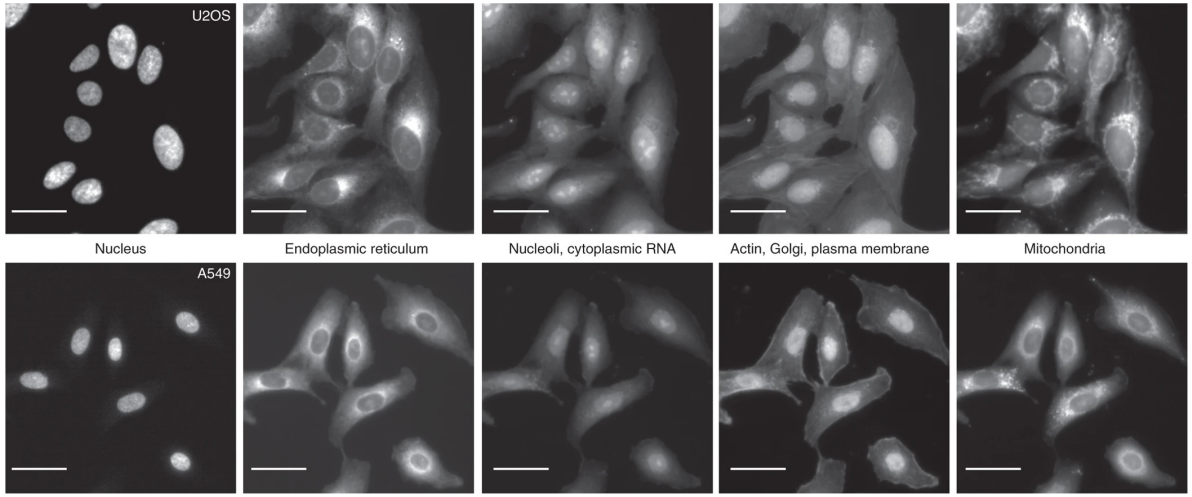
- Rich: Captures morphology of entire cell
- Scalable: reasonably cheap once set up
- Universality: All cells have the stained structure but in different conformations and abundance
- Single-cell: Captures single-cell outcomes

Cell Painting—The Proposal



- Instead of asking "what does this compound do to a culture as measured by a specific assay?" we ask "what does this compound do to cell morphology as measured by dyes?"
- **Run once, use to predict many downstream assay outcomes**

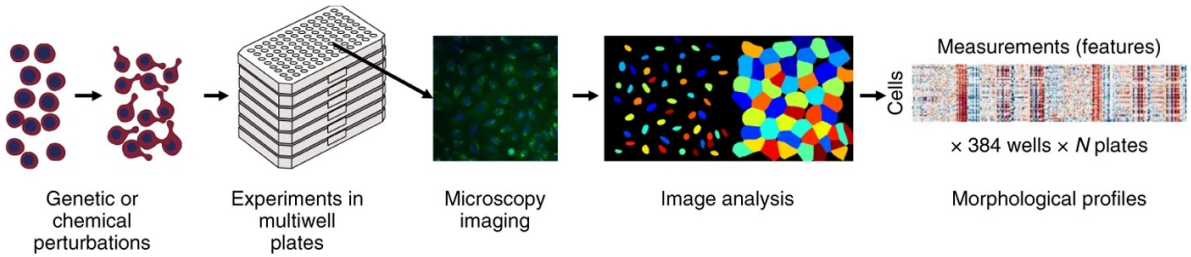
What is Cell Painting?



- 6 fluorescent dyes in 5 channels reveal different organelles and structures
  1. Hoechst 33342—DNA
  2. Concanavalin A/Alexa 488—Endoplasmic reticulum
  3. SYTO 14—RNA
  4. Phalloidin + WGA—Actin
  5. MitoTracker Deep Red—Mitochondria
- Combined: A comprehensive "photograph" of cellular state

Bray et al., "Cell Painting, a high-content image-based assay", Nature Protocols, 2016

What is Cell Painting?



Protocol details:

- Perturbations applied to cells in 9 sites per well in 384-well plates
- Cells stained and imaged (20x magnification)

Scale:

- Data volume: $\sim$ 40 GB per 384-well plate (17,280 images)
- Throughput: $\sim$ 1 week time from culture to images for one plate
- Cost: significantly less per sample than transcriptomics

Bray et al., "Cell Painting, a high-content image-based assay", Nature Protocols, 2016

Downstream Applications of Cell Painting

- **Mechanism of action identification:** Match unknown compounds to known drugs by profile similarity
- **Hit identification:** Screen for compounds with phenotypic effects
- **Polypharmacology detection:** Reveal off-target effects through morphological changes
- **Drug repurposing:** Identify new indications based on phenotypic rescue

"Profiling casts a much wider net, and avoids the intensive customization that is usually necessary for problem-specific assay development."

Bray et al., "Cell Painting, a high-content image-based assay", 2016

Recap

- We want to predict the effect of an intervention on many assay outcomes
- Proposed cell painting as a high-content screen (HCS) to address the combinatorial blow-up of intervention $\times$ assay combinations

Now we'll cover the two major cell painting datasets.

The JUMP Consortium

- **Analogy:** "Just as is the case for genomics and transcriptomics, a public reference database is critical for image-based profiling."
- **JUMP** = Joint Undertaking in Morphological Profiling
- **Goal:** Create the largest public Cell Painting reference dataset
- **Partners:** 10 pharma companies, 2 non-profits, 6 technology partners

Chandrasekaran et al., "JUMP Cell Painting dataset", 2023

JUMP Dataset Scale

Scale of JUMP-CP dataset:

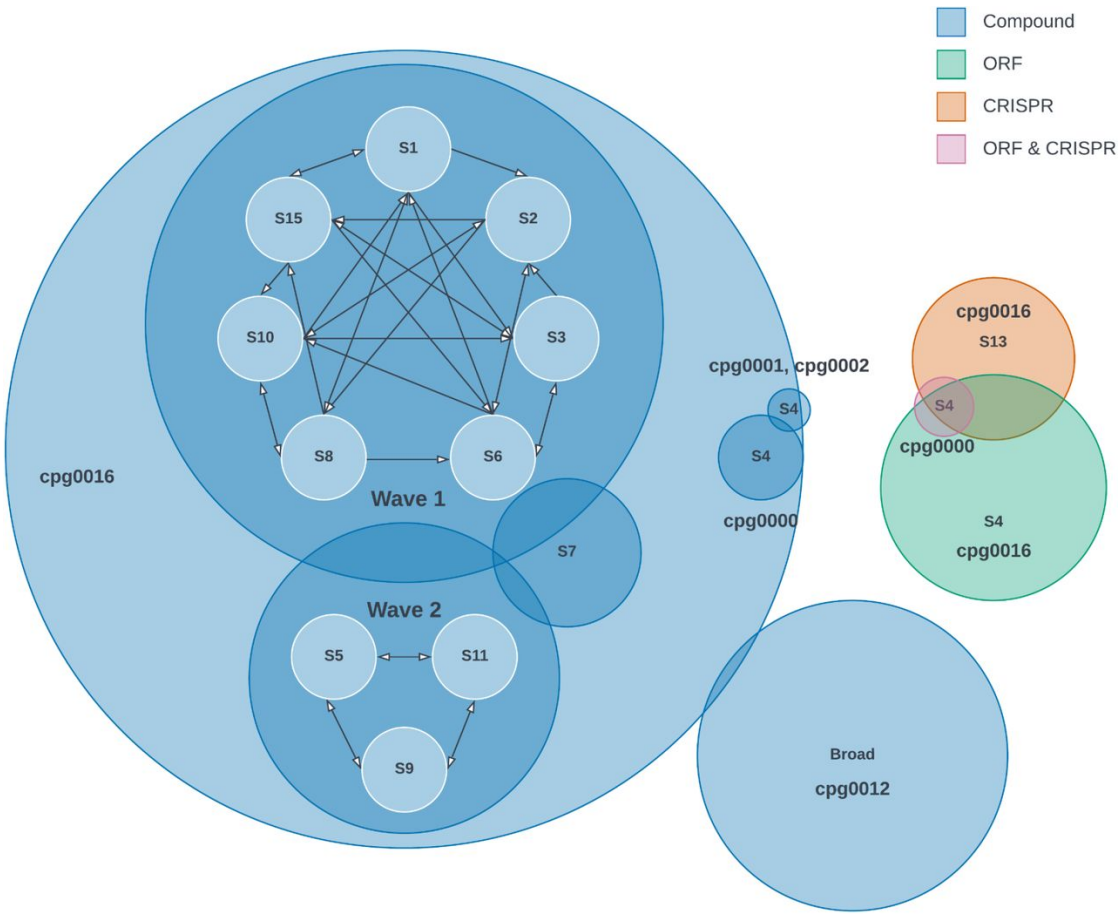
- 116,750 unique compound interventions
- 15,136 overexpression interventions (lentiviral ORF) encompassing 12,602 unique genes
- CRISPR knockout interventions of 7,975 genes
- 115 TB of data
- 1.6 billion cells captured

Experimental details:

- Cell line: U2OS (human osteosarcoma)
- Incubation hours before staining:
  - Compounds: 48h
  - ORF: 48h
  - CRISPR: 96h
- Replicates: 5 per perturbation across 2–5 sites
- Data generated across 12 different sites

Chandrasekaran et al., "JUMP Cell Painting dataset", Nature Methods, 2023

JUMP Dataset Structure

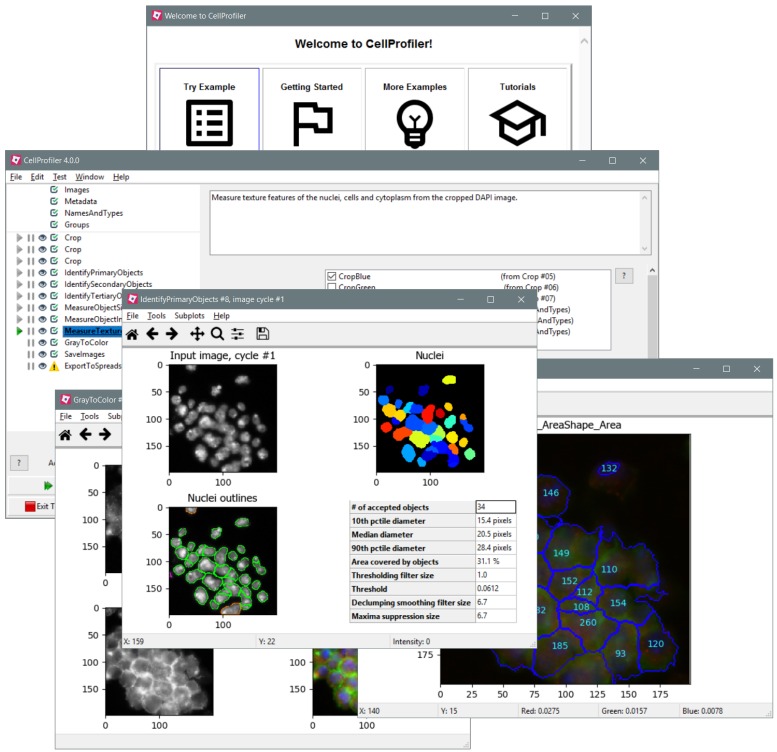


Chandrasekaran et al., Nature Methods, 2023

- Four dataset components (cpg0000, cpg0001, cpg0002, cpg0016)
- Multi-site approach (SN) allows us to understand site-specific confounders
- Negative (no intervention) positive controls (known effect interventions) both across and within plates

Processing The Data—CellProfiler

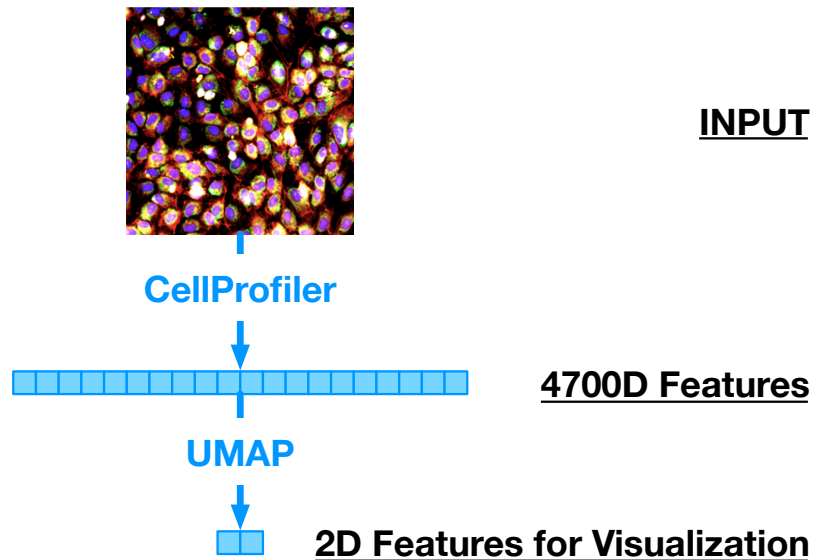
- Images are hard to interpret for humans and ML algorithms
- JUMP uses **CellProfiler features** to represent images—a bioimage analysis program that uses classical interpretable and "hand-crafted" methods to measure cellular features
    - E.g., intensity, granularity, texture, and density for each cell
- Features are averaged per well and normalized



CellProfiler

Visualizing The Data—UMAP

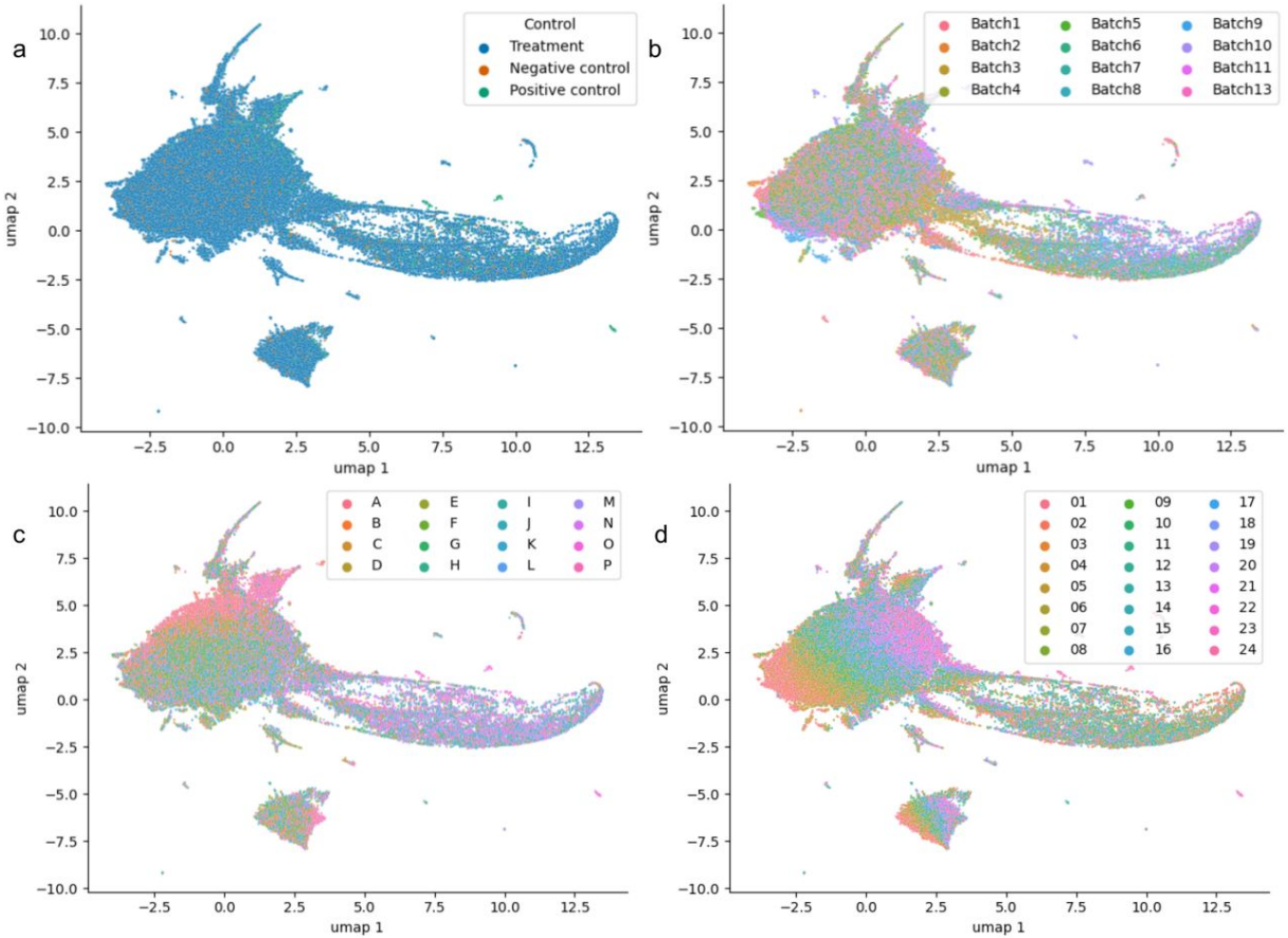
- CellProfiler features are ∼4,700 dimensional—can't visualize
- JUMP used **UMAP dimensionality reduction** visualize in 2D
- UMAP tries to **preserves distances** between points



Key idea: Similar images → nearby points in UMAP

Visualizing The Data with UMAP

The ORF Overexpression subset—each dot represents a well



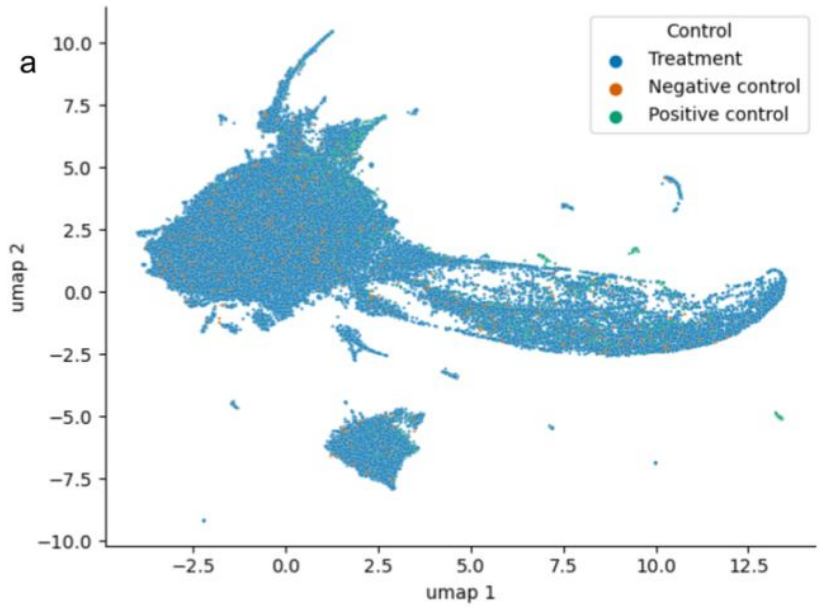
Chandrasekaran et al., "JUMP Cell Painting dataset", 2023

Visualizing The Data—Controls vs. Treatments

Control wells not clearly distinguishable from treatment well on UMAP

Possible explanations:

1. ORF treatments produce weak morphological effects
2. Processing effects dominate biological signal

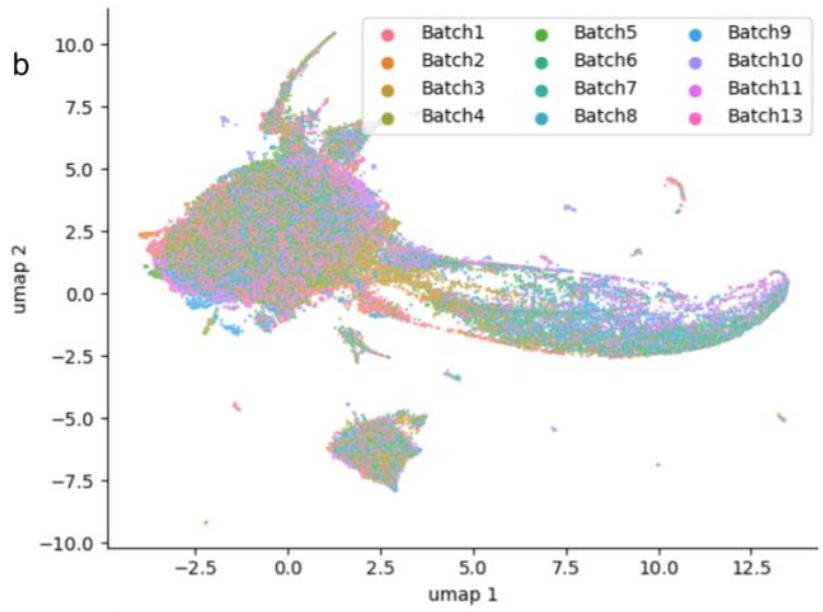


Chandrasekaran et al., 2023

Question: Does this imply there's no ORF overexpression effect?

Visualizing The Data—Batch Effects

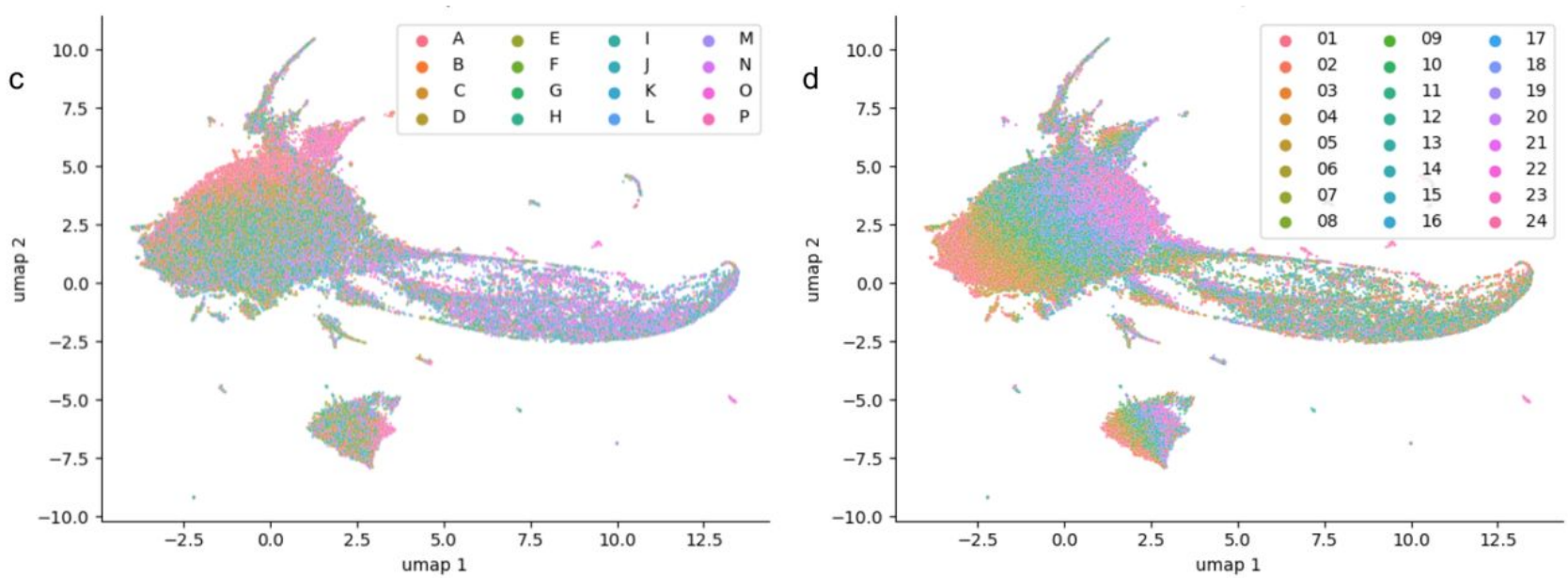
No clear batch effect on UMAP



Chandrasekaran et al., 2023

Question: Is this good?

Data Quality Challenges—Well Position Effects



Chandrasekaran et al., 2023

- Wells from same row or column cluster together
- Effect persists across perturbation types and batches
- Cannot be explained by placement of interventions in wells

Implication: Well-position effects can **confound** biological signal

RxRx3—Single-Lab Scale

Alternative approach:

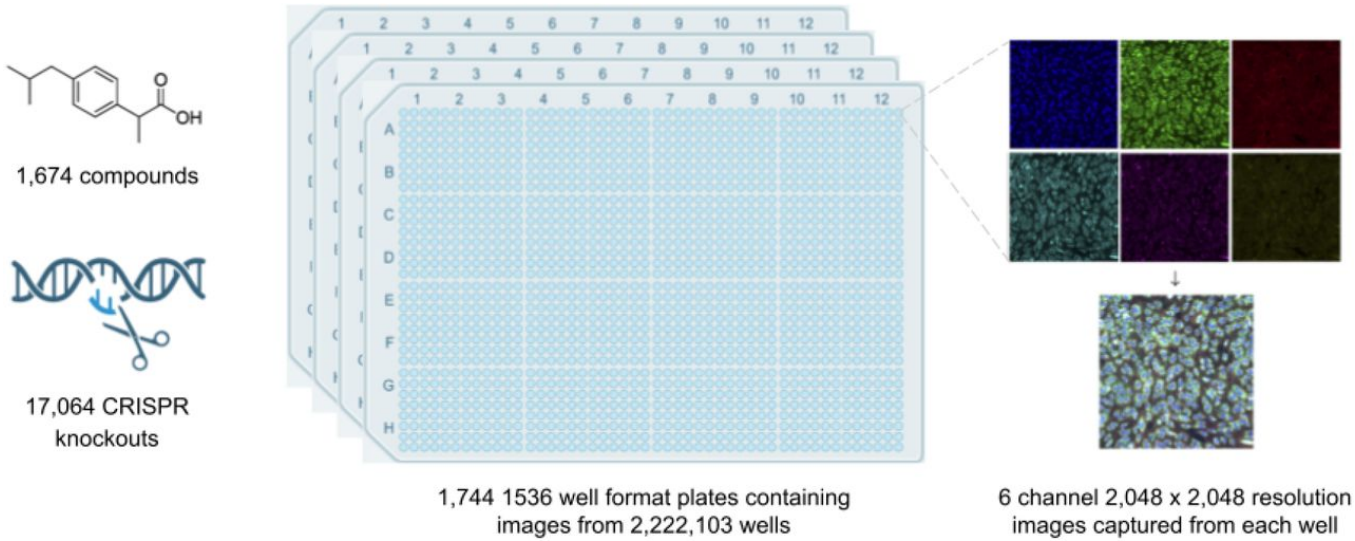
- Generate all data in Recursion's automated laboratory
- Claim: multi-center data can suffer from "confounding by cross-laboratory batch effects"

RxRx3 specifications:

- Cell type: HUVEC (human endothelial cells)
- 17,063 CRISPR knockouts (most with 6 guide RNAs per gene)
- 1,674 compounds at 8 concentrations (0.003–10 μM)
- 1,536-well plate format

Fay et al., "RxRx3: Phenomics Map of Biology", 2023

RxRx3 Overview



Fay et al., "RxRx3: Phenomics Map of Biology", 2023

- 5 standard cell painting stains + Alexa Fluor 555 Agglutinin (plasma membrane) for 6 total
- Imaging at 2,048 $\times$ 2,048 pixel resolution

JUMP vs. RxRx3 Comparison

Aspect	**JUMP-CP**	**RxRx3**
Cell type	U2OS	HUVEC
Sites	12 centers	Single lab
Genetic interventions	22K	17K
Compounds	116K	1.6K (8 conc.)

Different design philosophies:

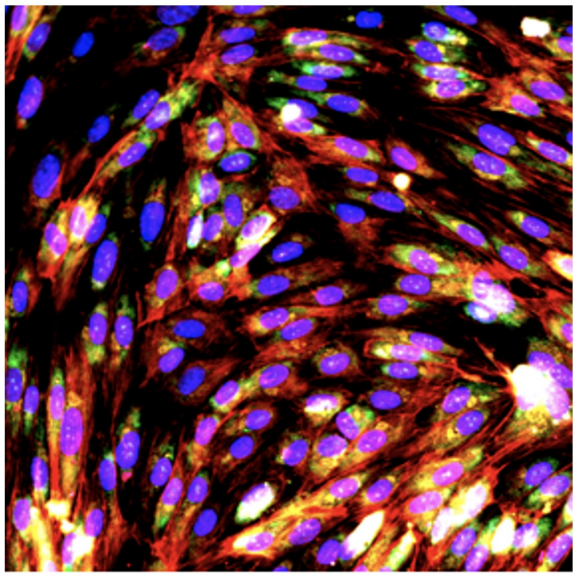
- **JUMP:** Diversity, breadth, multi-site
- **RxRx3:** Consistency, depth, single-lab

Recap

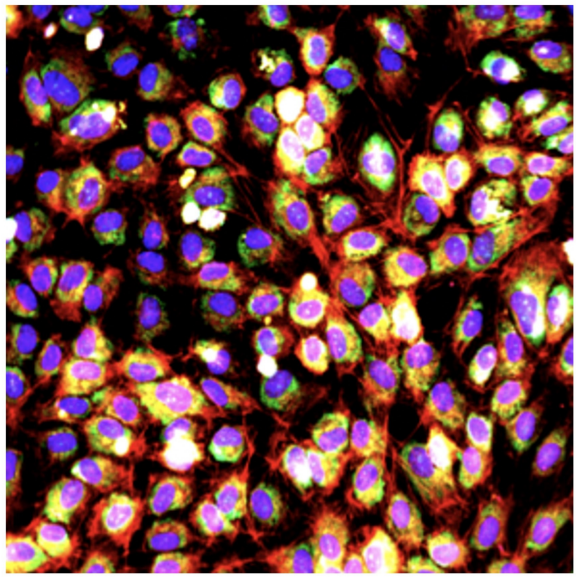
- We want to predict the effect of an intervention on many assay outcomes
- Proposed cell painting as a high-content screen (HCS) to address the combinatorial blow-up of intervention $\times$ assay combinations
- But the way JUMP represents HCS data doesn't obviously allow us to predict intervention effect—we can't even distinguish control vs treatment!
- **Questions:**
  - Do we think this means the genetic and compound interventions don't have an effect?
  - Is the information actually in the image?
  - What can we do to better represent effects when they exist?

The Information Is Clearly There

Which is control vs. treatment?

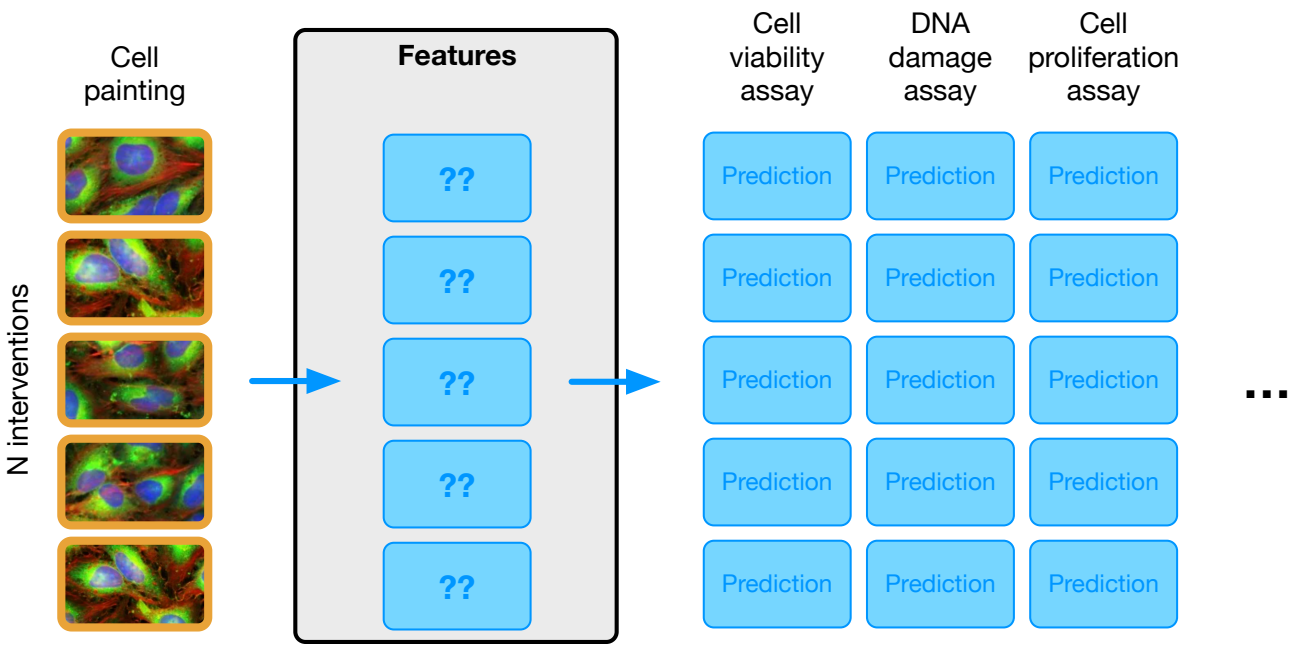


vs



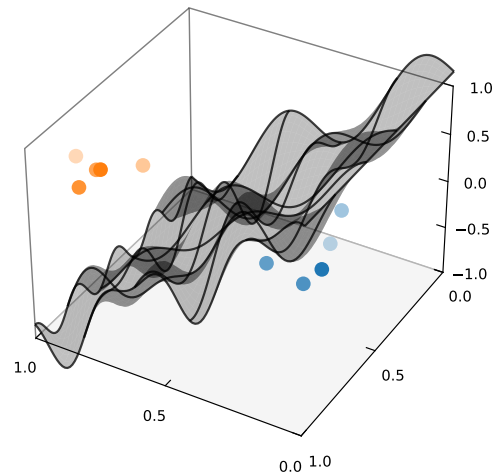
Question: How should we represent the HCS data (aka find features) to allow us to better distinguish these?

Learned Features—The Challenge



- **Images**: not interpretable and difficult for ML
- **CellProfiler**: unclear whether biologically relevant
- **UMAP**: unclear whether biologically relevant, dim. reduction may remove information
- **What we want**: biologically relevant, rich, and easy for ML
- **Approach**: what if we used ML to learn features?

Opening the Black Box—Neural Networks

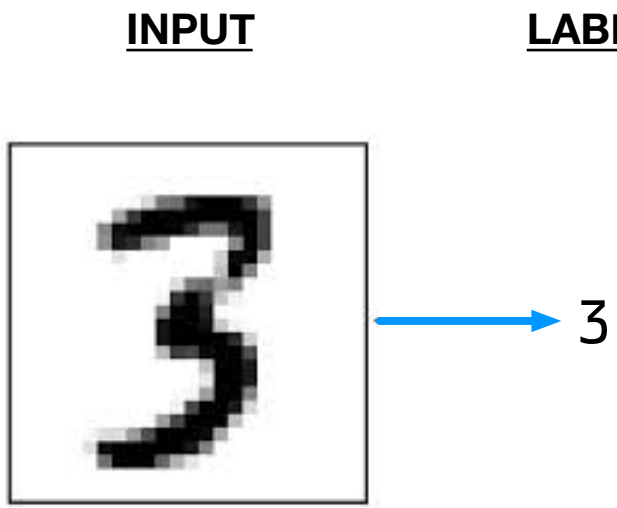


- Recall: The "flexible predictors" we discussed are typically **neural networks**
- Neural networks were invented in part to solve the feature learning challenge because **hand-crafted features** like CellProfiler had hit a wall
- What are neural networks? Let's find out!

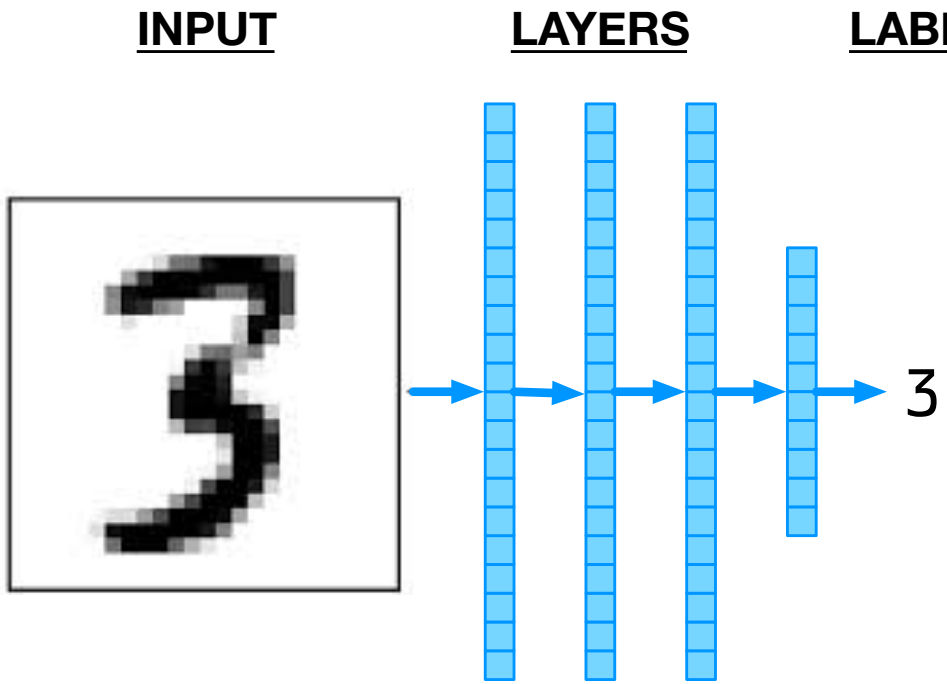
Opening the Black Box—Neural Networks

Watch this video:
`https://www.youtube.com/watch?v=aircAruvnKk`

- Digit classification: the video considers classifying handwritten digits (input) by predicting 0-9 (label)
- Introduces how neural networks solve this problem



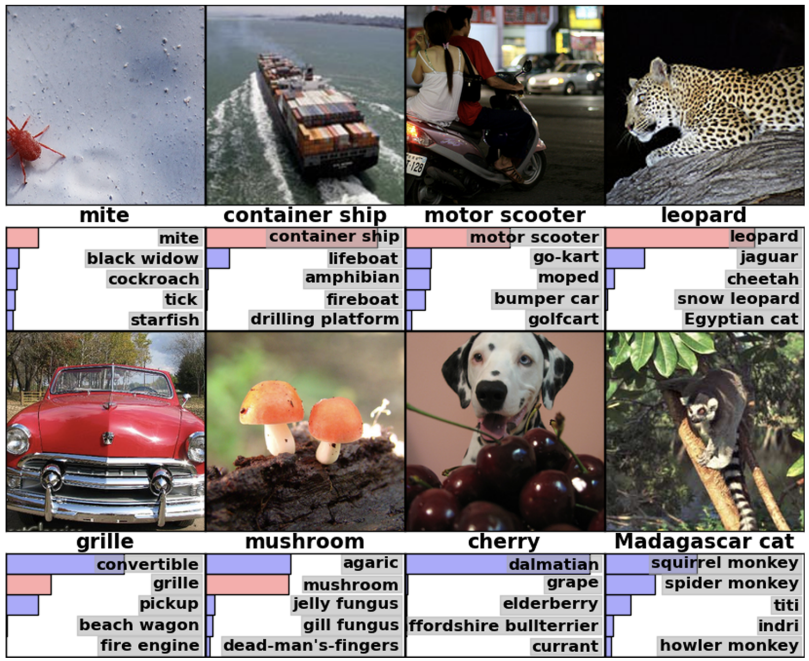
Opening the Black Box—Neural Networks



- These flexible predictors have intermediate values computed during prediction—called **layers**
- These intermediate values tend to learn to recognize composite properties of the input
- These intermediate values can be treated as **features**
- This information can potentially be used for **other tasks**

ImageNet

- Historical detour into a watershed moment in the field
- In early 2010s there was a big image classification dataset and challenge called ImageNet
- AlexNet was a neural network built here that made remarkable progress on ImageNet



Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks", NIPS, 2012

AlexNet

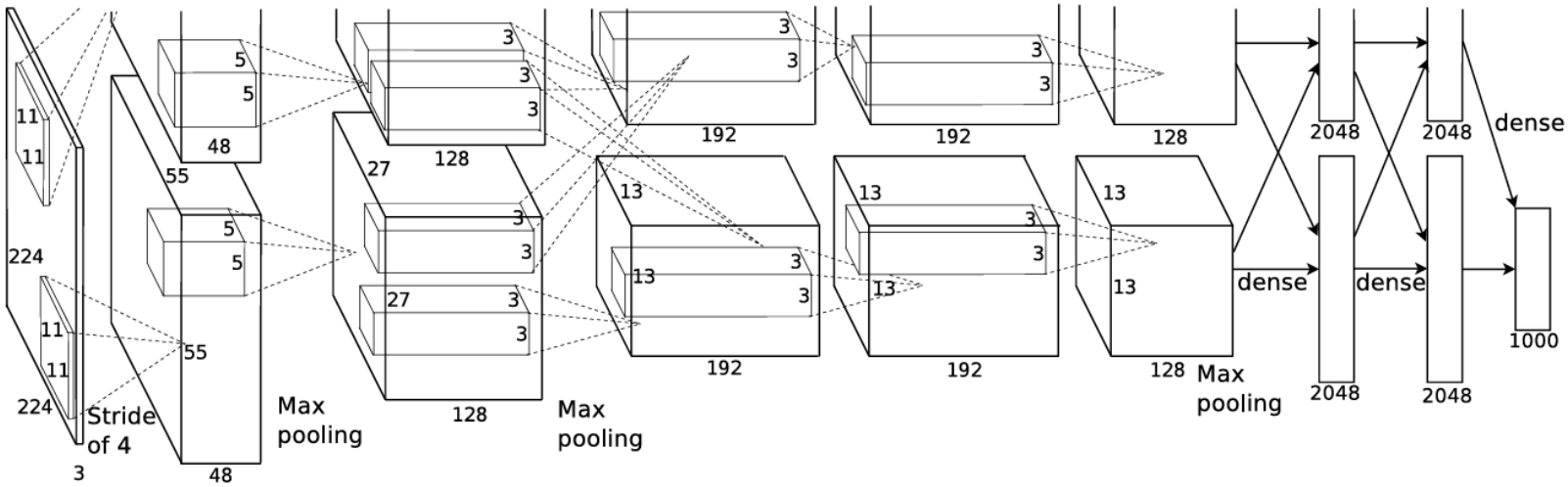
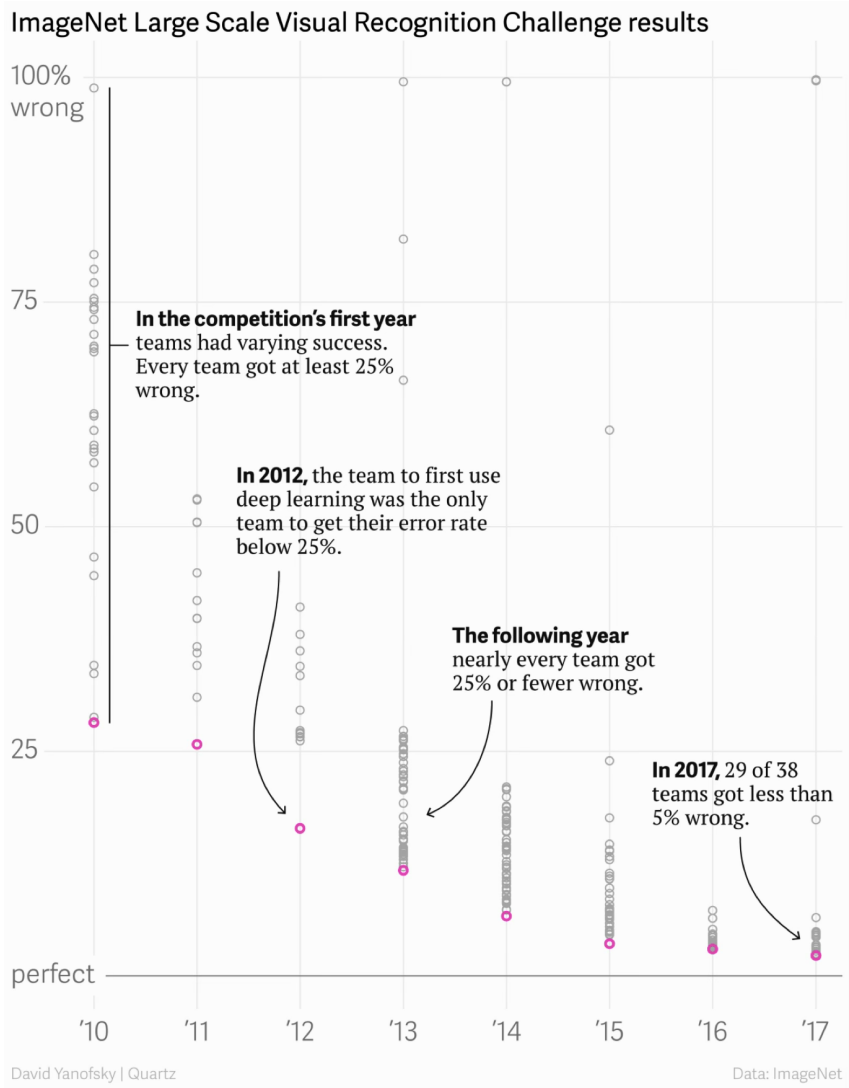


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

- AlexNet (2012) trained on ImageNet: 1.2 million images, 1000 classes

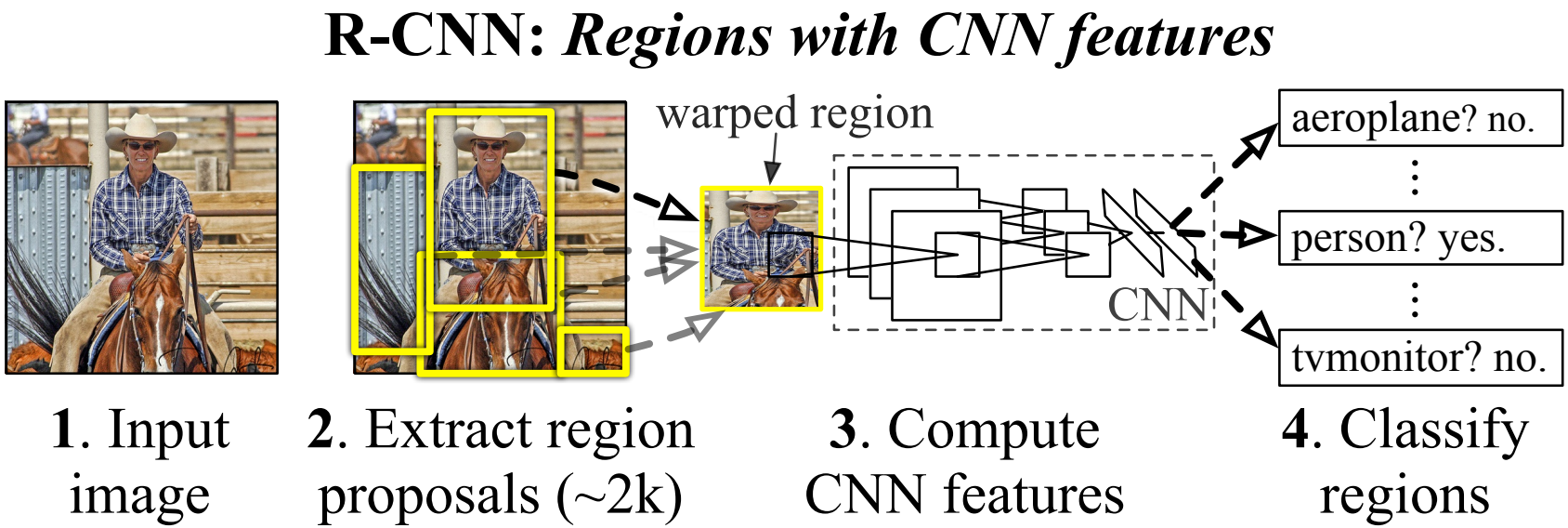
Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks", NIPS, 2012

AlexNet—Incredible Progress



AlexNet and Transfer Learning

Girschick et al. asked the following question: what if we use AlexNet layers as features instead of existing hand-crafted features for image processing?



Girshick et al., "Rich feature hierarchies for accurate object detection", CVPR, 2014

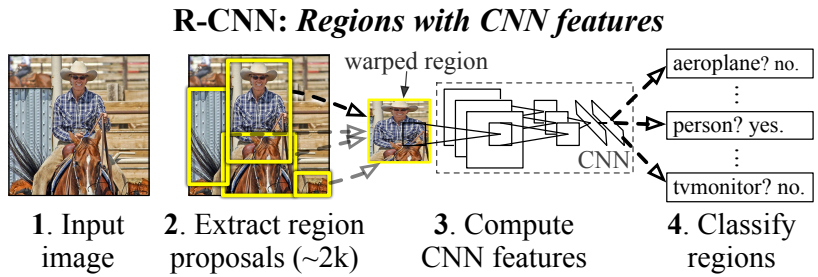
Focus: object detection—find and classify objects—distinct from image classification

AlexNet and Transfer Learning

Surprising discovery: Using intermediate layers of AlexNet as representations for the data improves simple ML classifiers on new tasks and new data sets

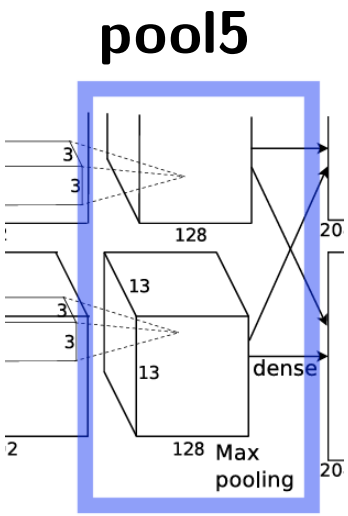
R-CNN results:

- Intermediate AlexNet layers: 44–46% mean average precision on object detection
- Hand-crafted features: 33–35% mean average precision



Girshick et al., "Rich feature hierarchies for accurate object detection", CVPR, 2014

What Do Networks Learn?



Find the images the most activate neurons in pool5 layer.

Bicycle neuron

pool5 feature: (3,3,6) (top 1 – 24)

Grill neuron

pool5 feature: (3,3,19) (top 1 – 24)

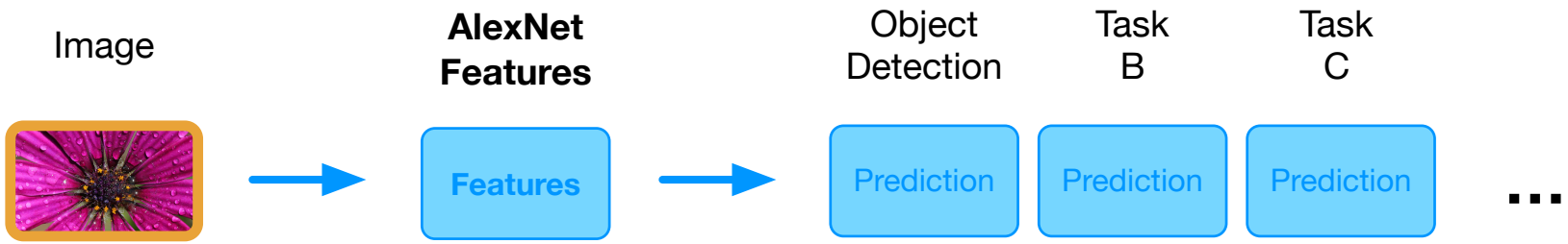
Window neuron

pool5 feature: (3,3,15) (top 1 – 24)

Girshick et al., "Rich feature hierarchies for accurate object detection", CVPR, 2014

Key insight: Networks learn interpretable, reusable concepts

AlexNet—Answering The Challenge



- AlexNet $\rightarrow$ feature learning and task transfer
- Discussion: Are AlexNet features good for cell painting images?
- Approach limited by the availability of ImageNet labels
- Are there methods that can benefit from more widely available data?

Proposed solution: self-supervised learning methods that "supervise" themselves from abundant data

Self-Supervised Learning Overview

Definition: Training a neural network with abundant data by creating a "pretext task"

Core idea: Use the structure of the data as-you-find-it to create a **supervision signal** to construct your loss function

Method	**Supervision Signal**	**Pretext Task**
SimCLR	Augmentation	Match similar images
CLIP	Text descriptions	Match text-image pairs
MAE	Image itself	Reconstruct masked patches

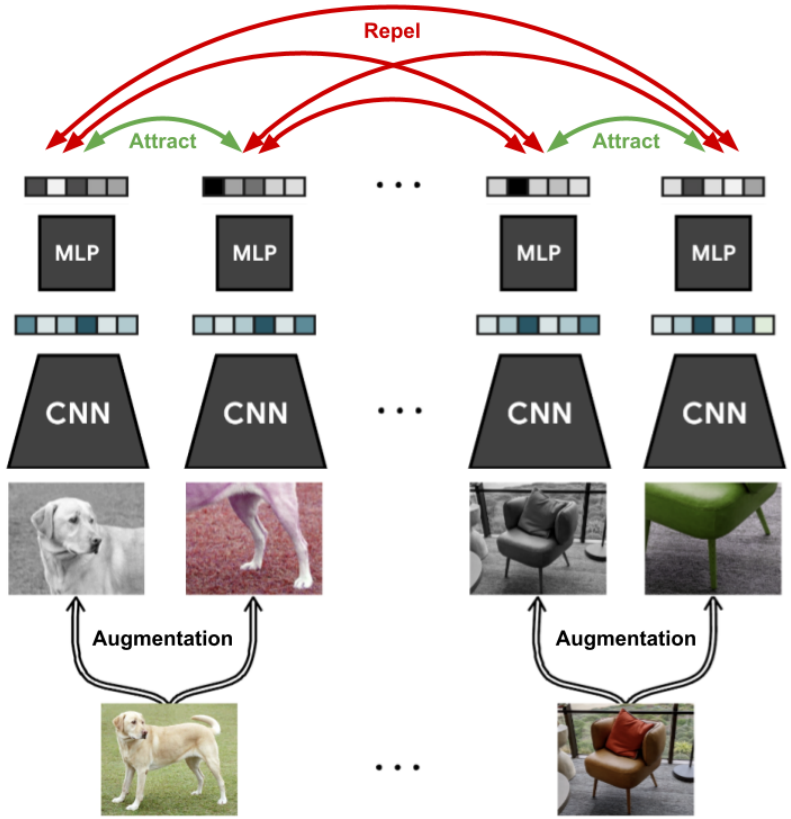
Key hypothesis: If the model can solve these pretext tasks, it must have learned useful representations for downstream tasks

Aside: For all of the methods I present, there's no guarantee they work on your data

SimCLR—Chen et al., 2020

Core idea:

1. Take one image, create two preturbed versions (crop, color distort, blur—sometimes called **augmentations**)
2. Train network to identify that these came from the same image
3. Other images in batch serve as "negatives"

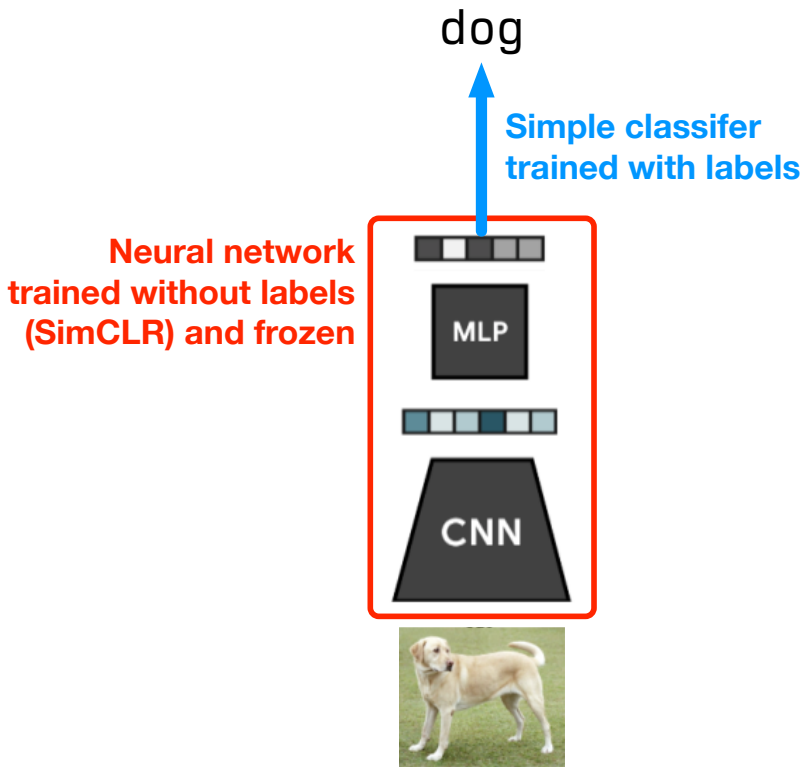


Chen et al., "A Simple Framework for Contrastive Learning", ICML, 2020

SimCLR—Chen et al., 2020

Evaluating representations:

1. "Freeze" the neural network trained via SimCLR (without labels)
2. Train a simple classifier with labels using the features of the pre-trained neural network—often called **linear probing**



Chen et al., "A Simple Framework for Contrastive Learning", ICML, 2020

SimCLR Key Findings

Critical findings:

- Linear probing on SimCLR embeddings achieves 76.5% top-1 accuracy—exceeding AlexNet's original reported performance!
- Augmentation composition is essential:
  - Crop alone = 33.1%
  - Crop + Color = 56.3%
- SimCLR is trying to streamline its representation by discarding the information in the augmentation

Chen et al., ICML, 2020

Why augmentation matters:

- Most classification doesn't depend on color
- So, it makes sense for SimCLR to use color augmentation as the network discards that information

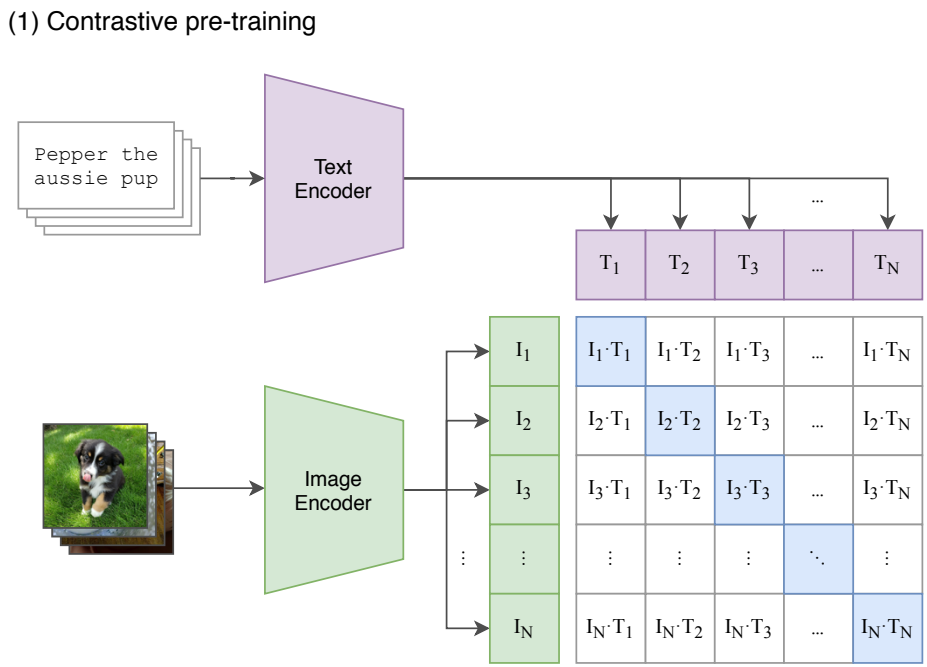
Discussion: For cell images, what augmentations make sense? What might destroy biological signal?

CLIP—Multimodal Contrastive Learning

Training data: 400 million (image, text) pairs from the internet

Method:

1. Train image and text encoder jointly
2. Maximize similarity between matching pairs
3. Minimize similarity between mismatched pairs

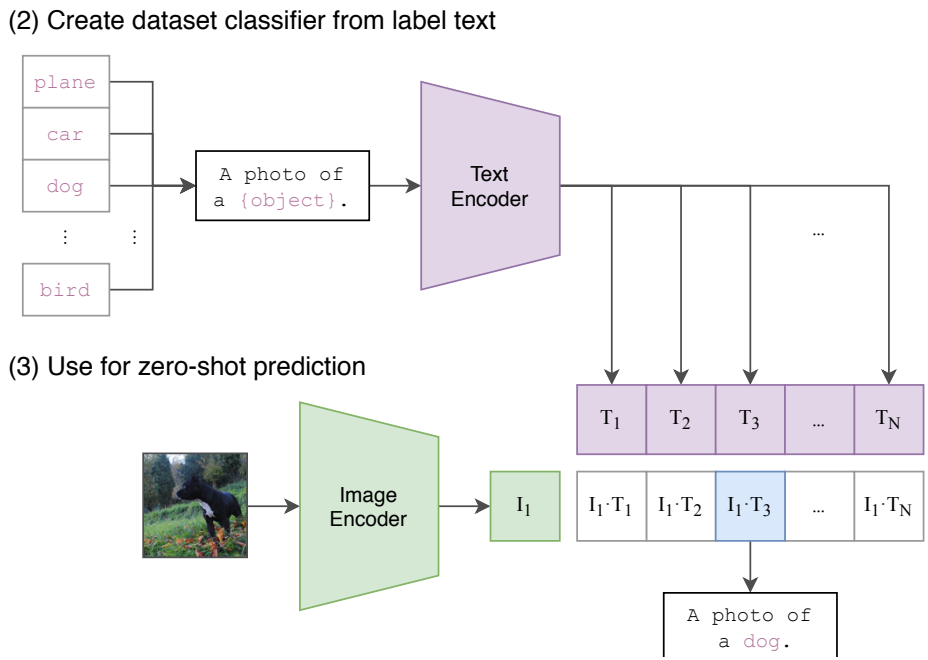


Radford et al., "Learning Transferable Visual Models From Natural Language Supervision", ICML, 2021

CLIP—Zero-Shot Performance

Zero-shot classification:

- Describe classes in text, classify images by matching image and text features
- Evaluate this predictor on ImageNet
- No expensive ImageNet labels used for training—just captions found on the internet



Radford et al., "Learning Transferable Visual Models From Natural Language Supervision", ICML, 2021

CLIP Zero-Shot Results

Key result: CLIP achieves 76.2% top-1 accuracy on ImageNet

WITHOUT using ANY of ImageNet's 1.28 million labeled training examples

This matches the original AlexNet!

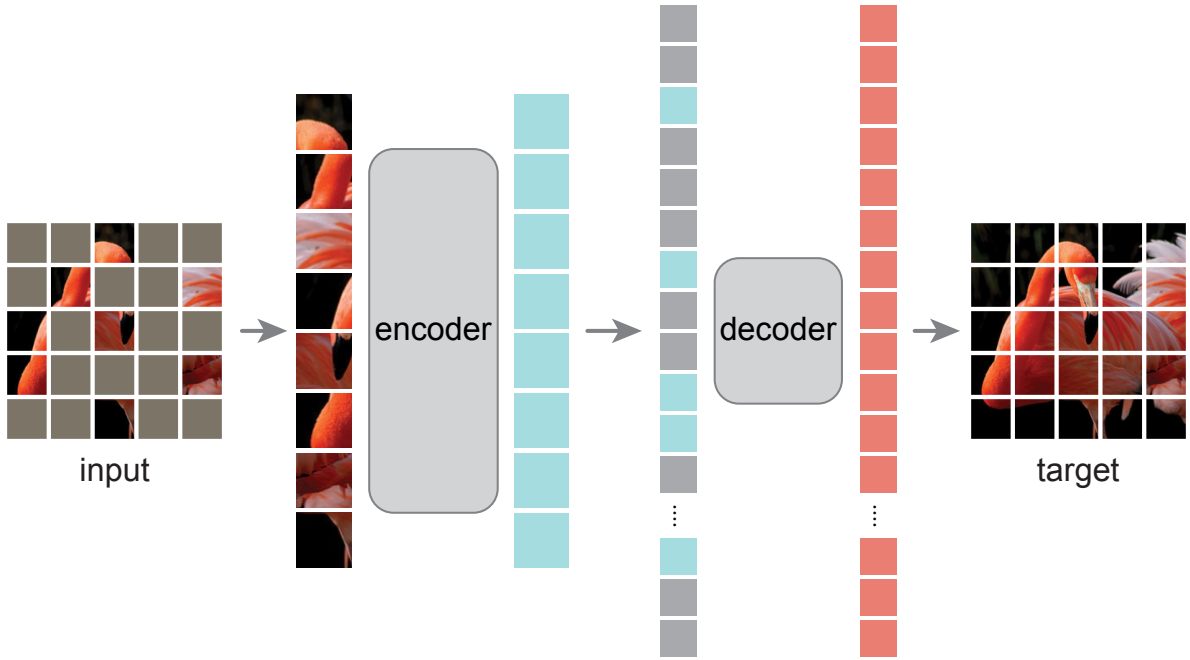
Additional benefits:

- Much more robust to distribution shift
- Open vocabulary—can recognize anything describable in text
- "No task-specific training needed"—the model recognizes concepts from descriptions alone

Relevance: Can we describe biological concepts in text and classify cells?

Radford et al., ICML, 2021

MAE—Masked Autoencoders

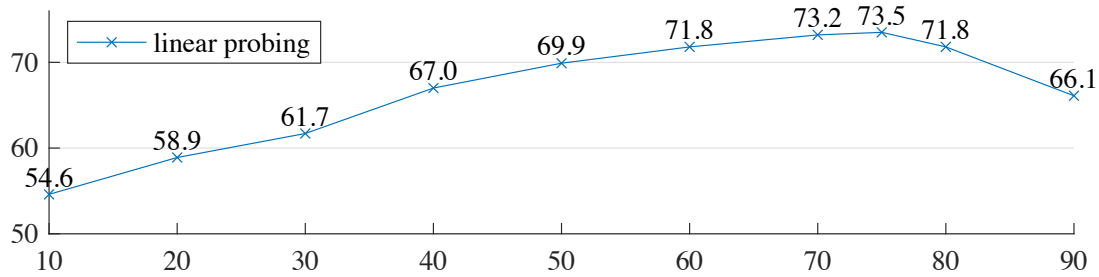


He et al., "Masked Autoencoders Are Scalable Vision Learners", CVPR, 2022

Core idea: Mask out 75% of image patches, train to reconstruct them

Key insight: High masking ratio creates a challenging task that requires understanding image content

MAE ImageNet Results



He et al., "Masked Autoencoders Are Scalable Vision Learners", CVPR, 2022

Surprising findings:

- Use an internal layer of the network as a representation
- Linear probing achieves 73.5%—comparable to SimCLR!
- Minimal data augmentation needed

MAE Reconstruction Examples

- 85% masking: model still produces plausible reconstructions
- The model must understand: object shapes, textures, semantic content
- This understanding becomes the learned representation

He et al., CVPR, 2022

Self-Supervised Methods Summary

Method	Supervision Signal	Pretext Task
SimCLR	Augmentation	Match similar images
CLIP	Text descriptions	Match text-image pairs
MAE	Image itself	Reconstruct masked patches

Common theme: No manual labels needed

All achieve strong ImageNet performance

Key question: How do we apply this to cell painting images?

The Phenomics Challenge

Two fundamental challenges:

1. Quantifying cellular responses to genetic and chemical perturbations by predicting the outcomes of various assays
2. Featurizing cell painting microscopy images to support this prediction challenge

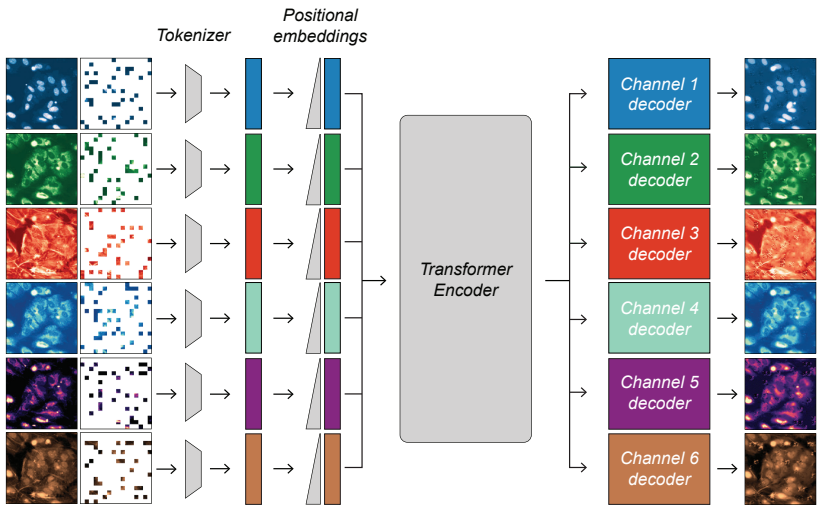
Kraus et al., "Masked Autoencoders for Microscopy", 2024

Question: Can self-supervised methods help with cell painting featurization?

MicroMAE—Masked Autoencoders for Microscopy

Training data:

- **MAE:** 75–85% masking ratio
- **RPI-93M (private):** 92.8M images, 3.9M perturbations
- **Image input:** 2,048 × 2,048 × 6 pixels (6-channel Cell Painting) cropped to 256 × 256

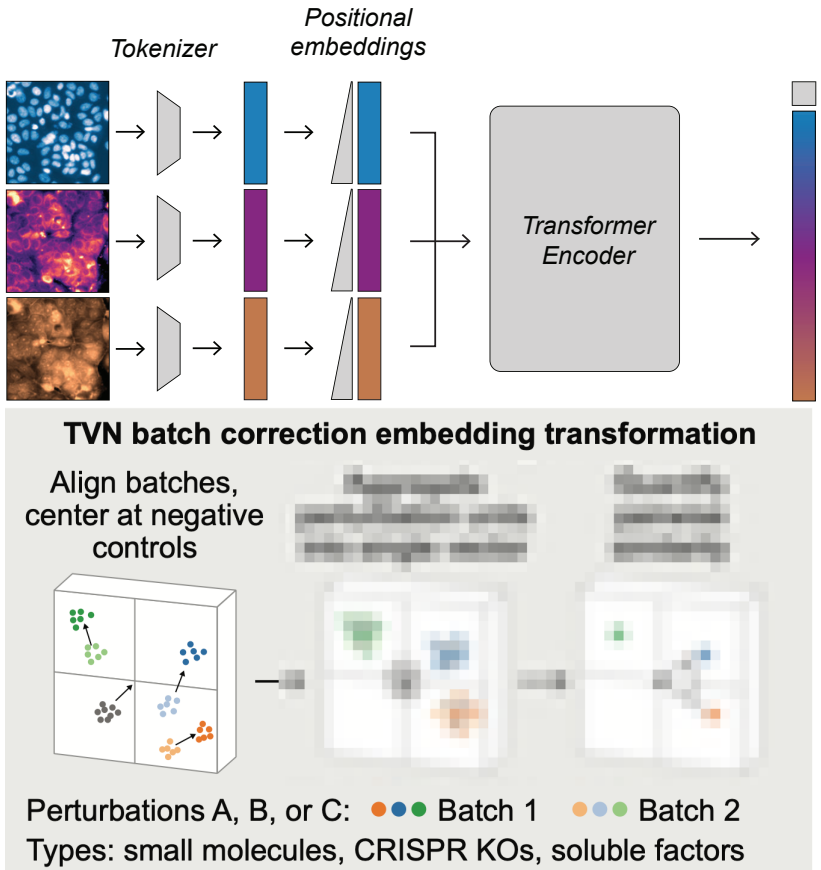


Kraus et al., "Masked Autoencoders for Microscopy are Scalable Learners of Cellular Biology", 2024

MicroMAE Architecture

Typical Variation Normalization (TVN) to remove batch effects

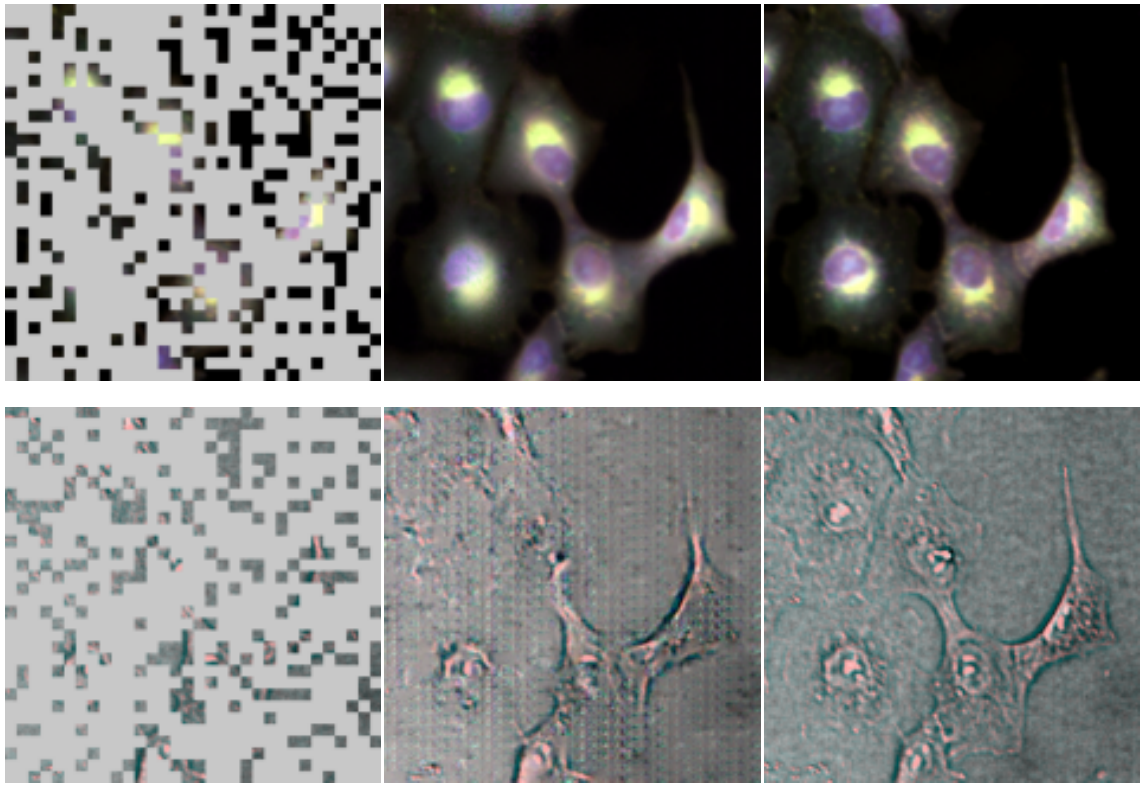
- Use negative controls across the experiment to remove natural correlation between embedding dimensions
- For each batch, remove further correlations found in negative controls for that batch



Kraus et al., 2024

Goal: Remove batch-wise variation

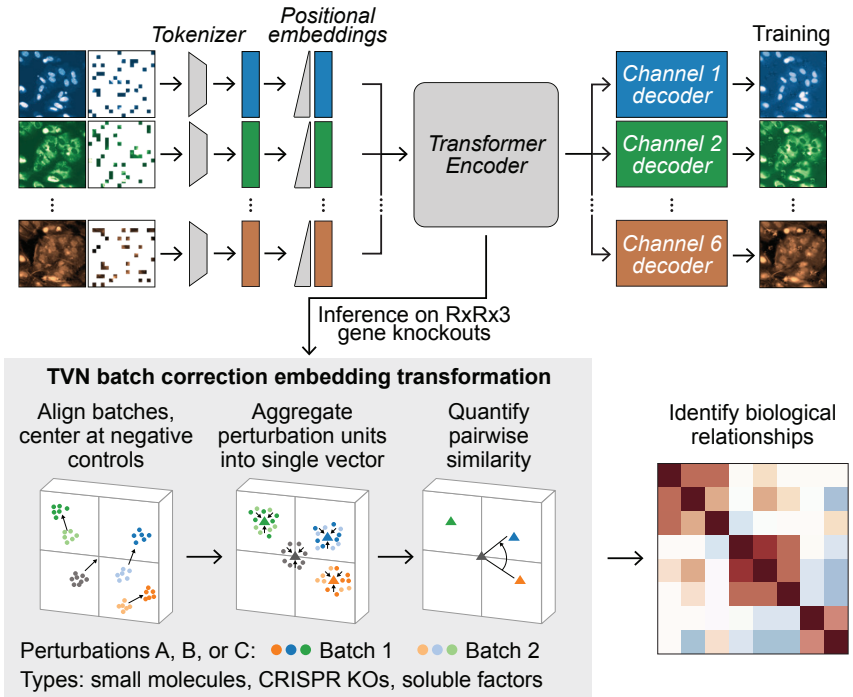
MicroMAE Reconstructions



Kraus et al., 2024

- The model learns meaningful cellular structure:
  - Nucleus shapes and boundaries
  - Organelle distributions
  - Cytoskeletal patterns

Evaluating Biological Relationship Recall



Kraus et al., 2024

Method:

1. Compute a single, aggregated embedding for each perturbation
2. Compute cosine similarity between aggregate embeddings
3. For each gene, check whether the top 5% most aligned gene embeddings and bottom 5% most anti-aligned are related genes

MicroMAE Outperforms Alternatives

Model	Dataset	CORUM	hu.MAP	Reactome	StringDB
Random	N/A	.100	.100	.100	.100
Pixel stats	N/A	.280	.260	.160	.270
ImageNet ViT-L/16	ImageNet-21k	.531	.360	.228	.409
WSL ViT-L/16	RxRx3	.532	.353	.196	.402
MAE ViT-L/16	**RxRx3**	**.560**	**.374**	**.231**	**.427**

Kraus et al., 2024, Table 3

Surprising Finding—ImageNet Models Work!

Counterintuitive result: ImageNet-pretrained models outperform many cell-specific models

Models trained on cats, dogs, and cars transfer surprisingly well to microscopy!

Why? Visual primitives (edges, textures, shapes) are universal?

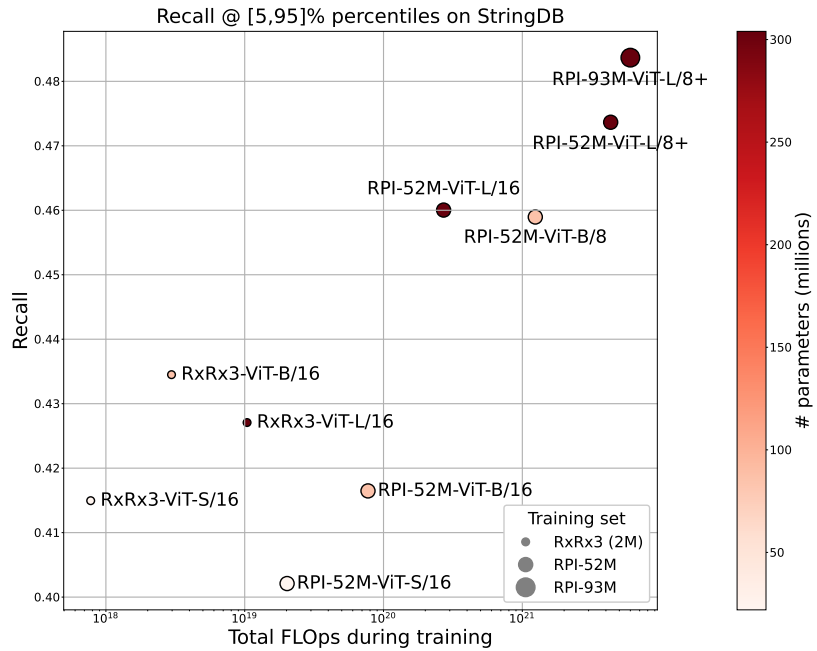
Implication: General-purpose representations have broad applicability

BUT: Cell-specific MAE training still does better when scaled appropriately

Kraus et al., 2024

Scaling Laws for Biology

Key finding: MAE performance scales as we train bigger models on more data



Kraus et al., 2024

Transfer to External Data (JUMP-CP)

Test: Does MicroMAE transfer to data from different labs/protocols?

JUMP-CP: External dataset from 12 different centers

Results (ability to retrieve replicate perturbations against the background of negative control samples):

Model	**Average Precision**
CellProfiler	.53 ± .30
ImageNet ViT	.88 ± .09
WSL ViT	.84 ± .08
MAE ViT-L/8+	.78 ± .13
CA-MAE ViT-L/16+	**.95 ± .05**

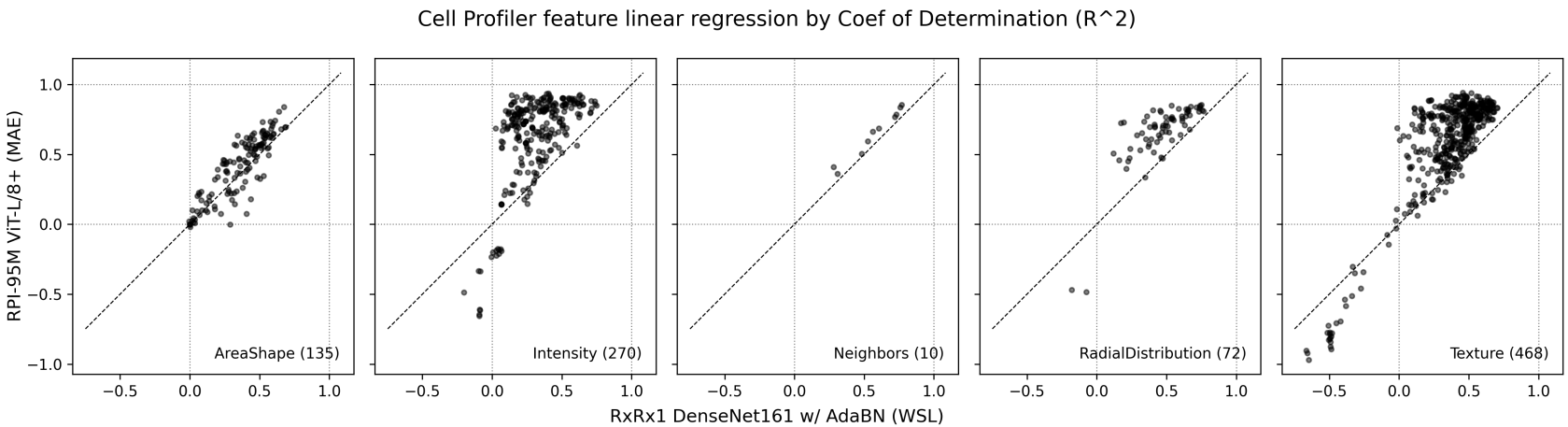
CA-MAE (Channel-Agnostic) achieves best transfer—representations generalize beyond training data

Kraus et al., 2024

What Information Do MAE Embeddings Capture?

Test: Can MAE embeddings predict CellProfiler features?

Method: Linear regression from embeddings to CP features and compare coefficient of determination



Kraus et al., 2024

Application—Predicting Intervention Effects

Practical application for drug discovery:

1. Screen compounds with Cell Painting
2. Extract embeddings with MicroMAE
3. Compare compound embeddings to genetic perturbation embeddings
4. If drug X embedding similar to gene Y knockout embedding:
   - Drug X may target gene Y's pathway
   - Enables target identification without biochemical assays

Key insight: If gene A's knockout produces a similar embedding to drug B's effect, then drug B might target gene A's pathway.

This connects back to our opening question about causal effects of interventions.

Bringing It All Together

The journey of this lecture:

1. **Problem:** Causality in cell biology requires interventional experiments, but we can't measure every outcome for every intervention
2. **Cell Painting:** Universal outcome measure that scales to millions of experiments, but images difficult to interpret and difficult to use with ML
3. **Self-supervised learning:** Learn meaningful features without labels (SimCLR, CLIP, MAE)
4. **MicroMAE:** Apply MAE to cell painting, achieves biological relationship recall

Question

Have we solved our original problem?

The Interpretability Challenge

Problem with MAE embeddings:

MAE helps us predict whether two perturbations are similar via cosine similarity, but we cannot tell *why* or *in what ways* they are different. We collapse multidimensional representations down to a single similarity score.

MAE is an effective "black box" but:

- What biological concepts does it encode?
- How does it distinguish perturbations?
- Can we extract interpretable features?

There's still work to do! See Donhauser et al., "Interpretable Concept Bottlenecks for Phenomics", 2025

Key Takeaways

1. **Causality requires intervention:** Simpson's Paradox shows why observational data is insufficient; perturbation experiments are essential for drug discovery
2. **Cell Painting is a universal outcome measure:** 6 dyes, 8 organelles, $\sim$1,500 features per cell, enabling "run once, query many" experiments
3. **Self-supervised learning extracts universal features features:** MAE learns to reconstruct images without labels, learning meaningful biology in the process
4. **But interpretability is still an open question.**