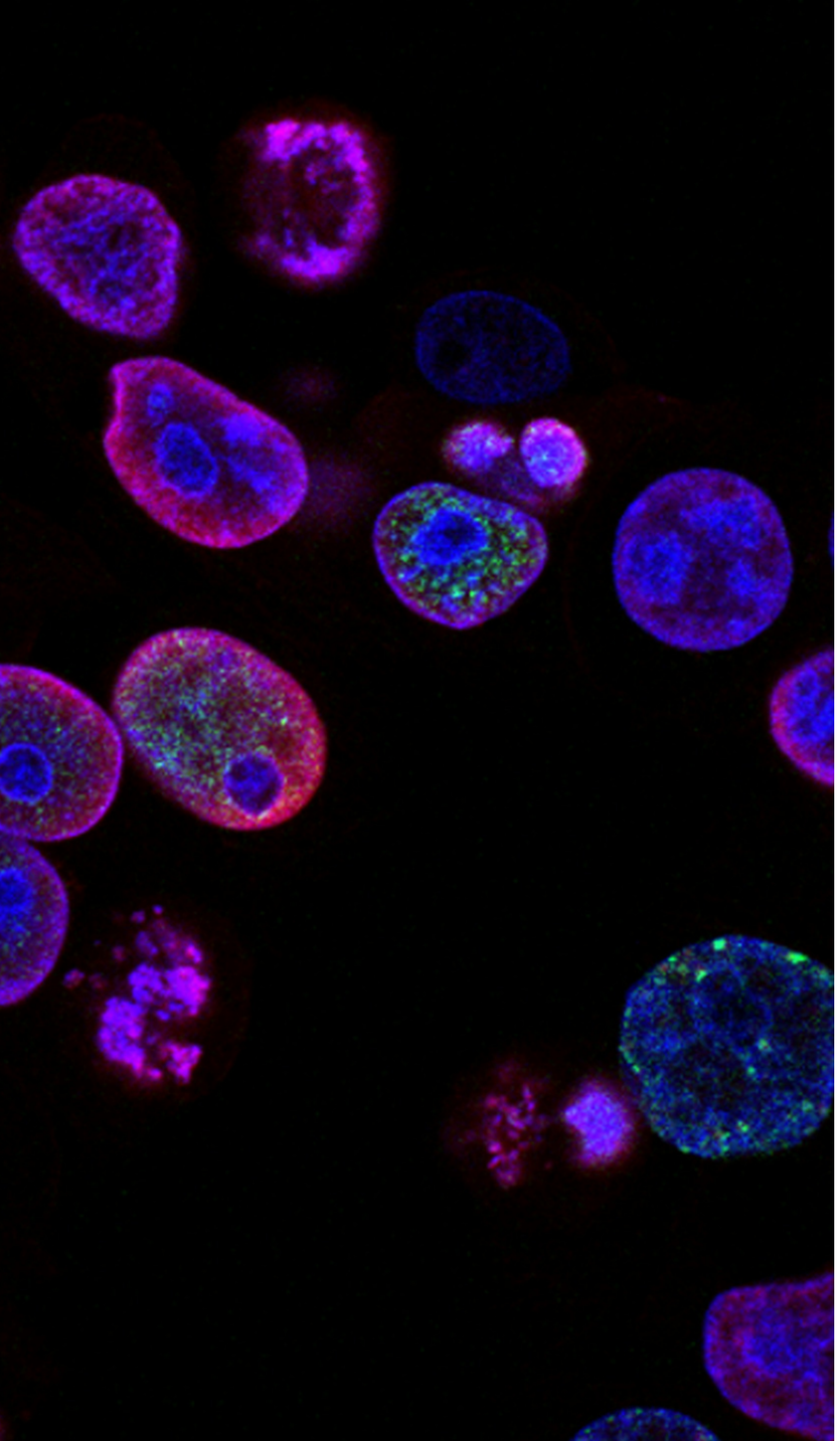
# Machine Learning for Cancer Genomics

Cait Harrigan
March 23, 2024

University of Toronto, Vector Institute
*cait.harrigan@mail.utoronto.ca*

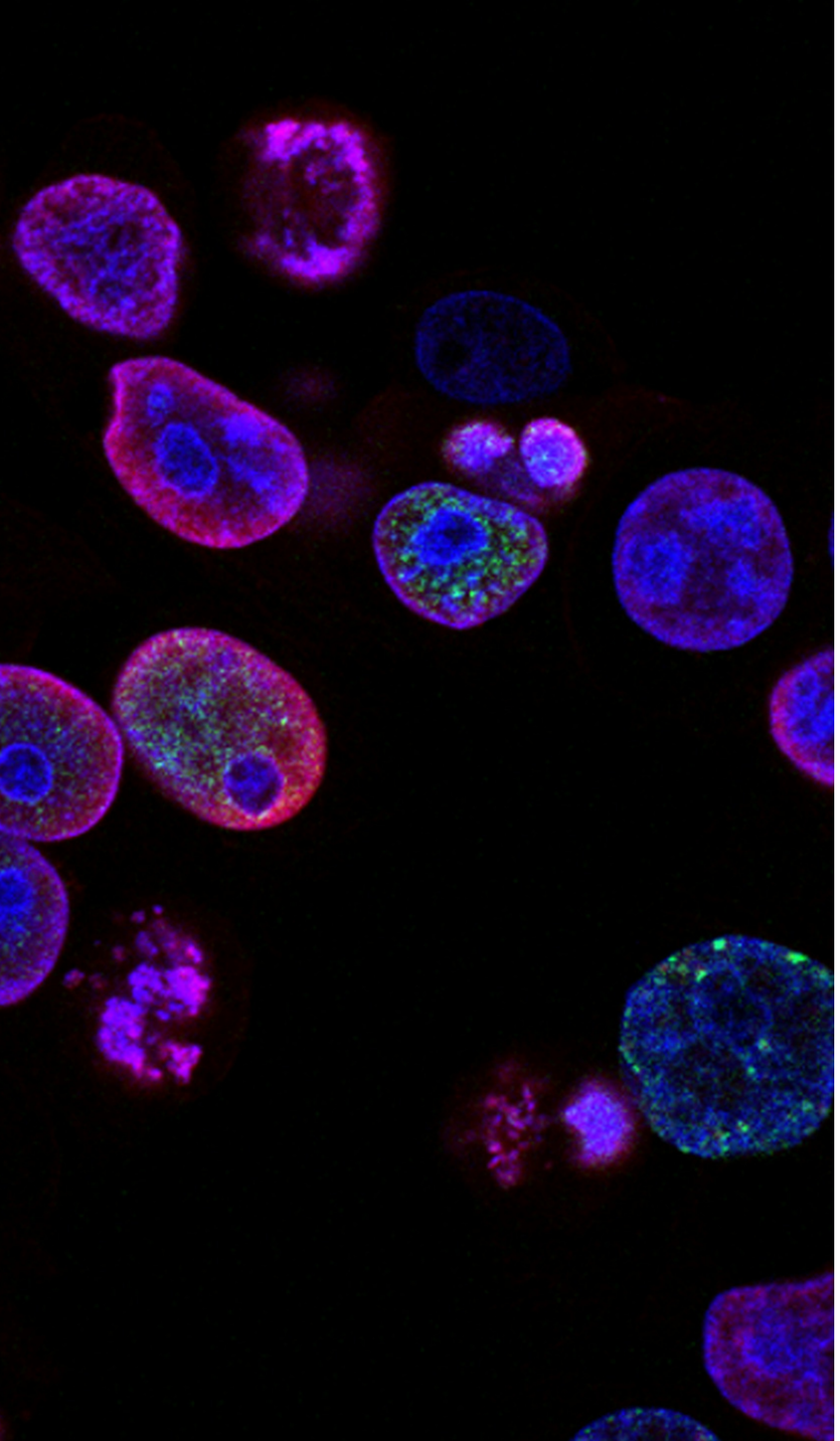


# Agenda

Part 1: Brief overview of the field

Part 2: My research - cancer evolution

Part 3: A day in the life of a PhD student



# Agenda

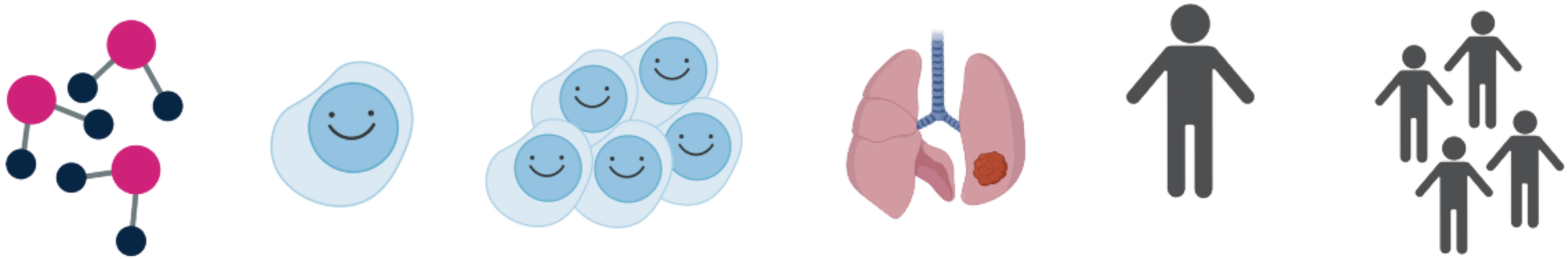
Part 1: Brief overview of the field

Part 2: My research - cancer evolution

Part 3: A day in the life of a PhD student

## Computational biology

## ML for health

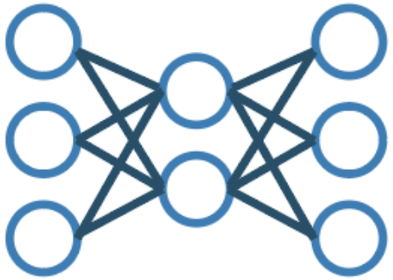
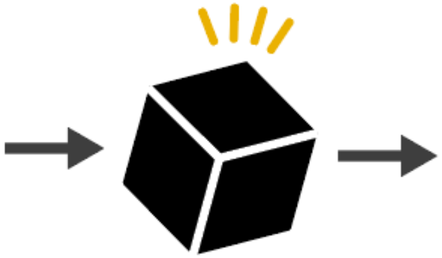


# Challenges in biomedical ML applications

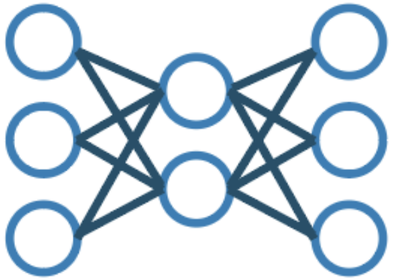
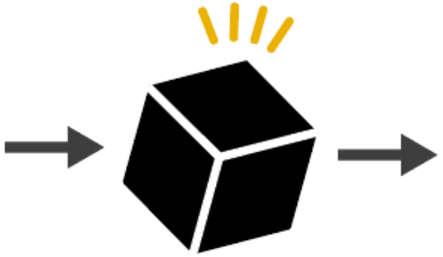
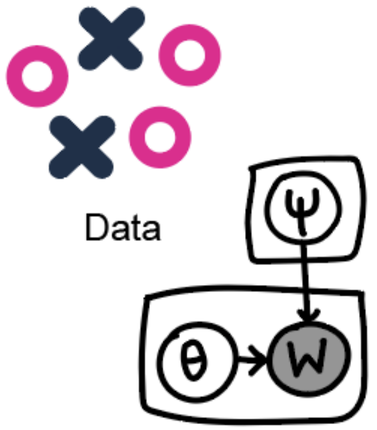
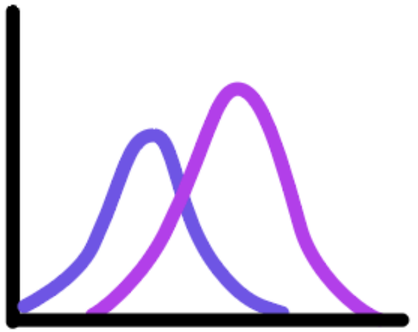
- Data size
- Labels are expensive
- Missing “ground truth”
- Generalization
- Interpretability and causal inference

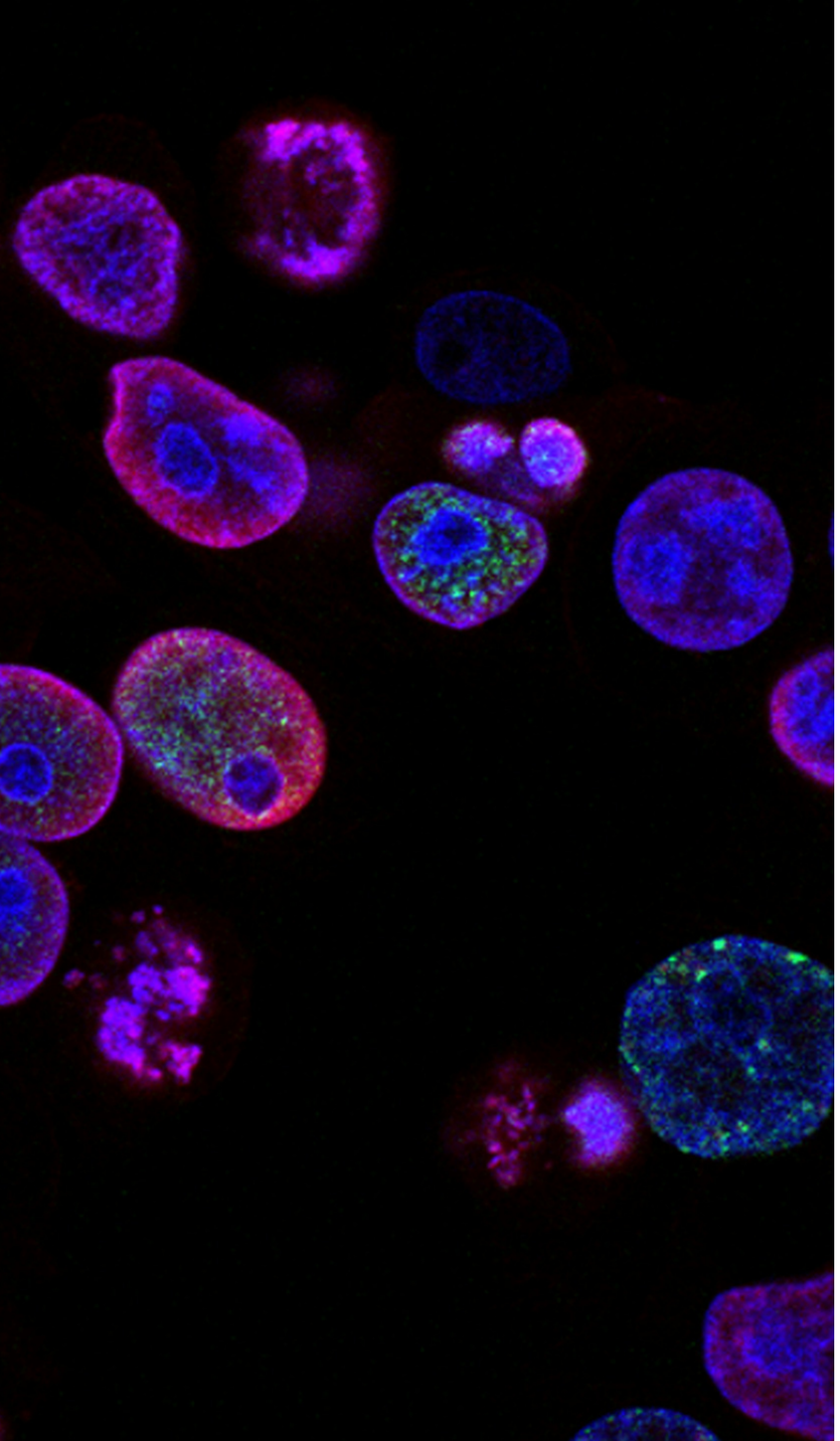
# Two common modelling paradigms

# Two common modelling paradigms

| | Input | Training loop | Final product |
|---|---|---|---|
| **Deep learning for prediction** |  Data |  Update weights |  Black box predictor |

# Two common modelling paradigms

| | Input | Training loop | Final product |
|---|---|---|---|
| **Deep learning for prediction** |  Data |  Update weights |  Black box predictor |
| **Bayeisan inference** |   Probabilistic model | $\hat{\theta}_i = U(\hat{\theta}_{i-1})$ Update parameter estimates |  Posterior distributions |



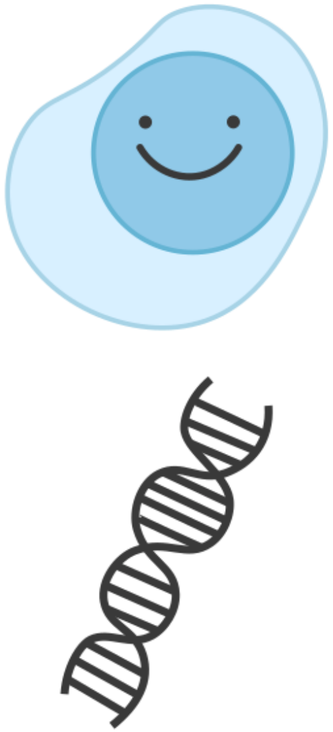
# Agenda

Part 1: Brief overview of the field

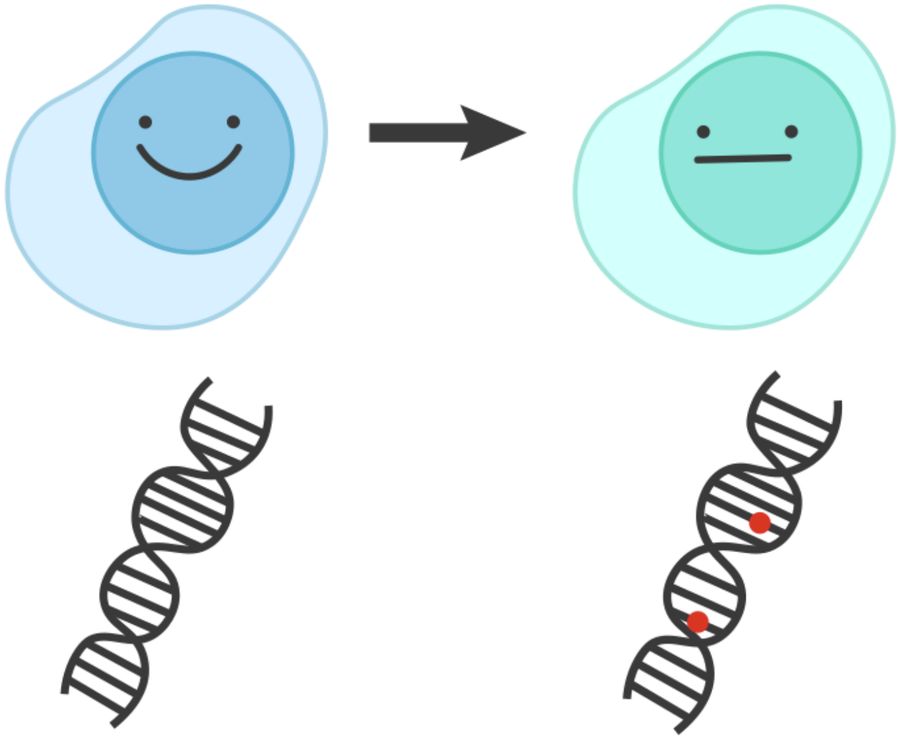
Part 2: My research - cancer evolution

Part 3: A day in the life of a PhD student

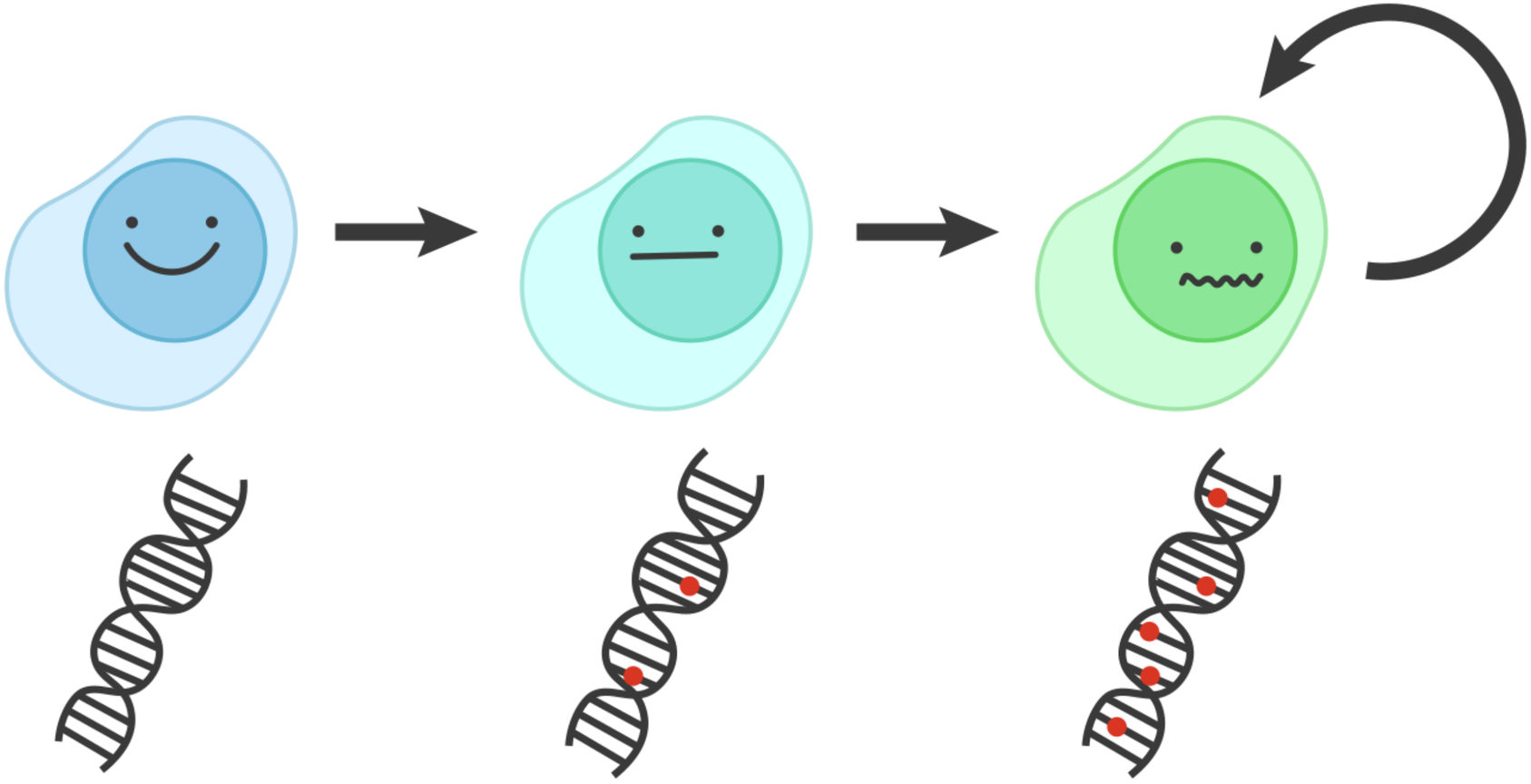
# DNA = set of instructions which tell a cell what to do



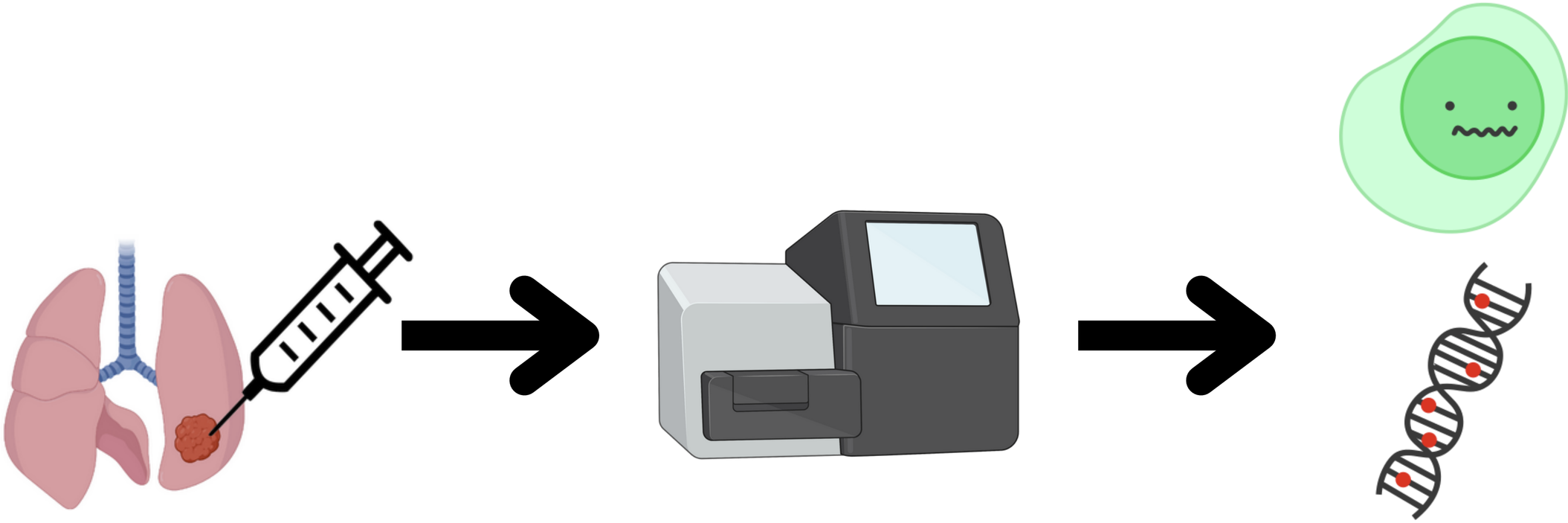
# DNA = set of instructions which tell a cell what to do



# DNA = set of instructions which tell a cell what to do



# With whole genome sequencing, get a list of mutations in a sample



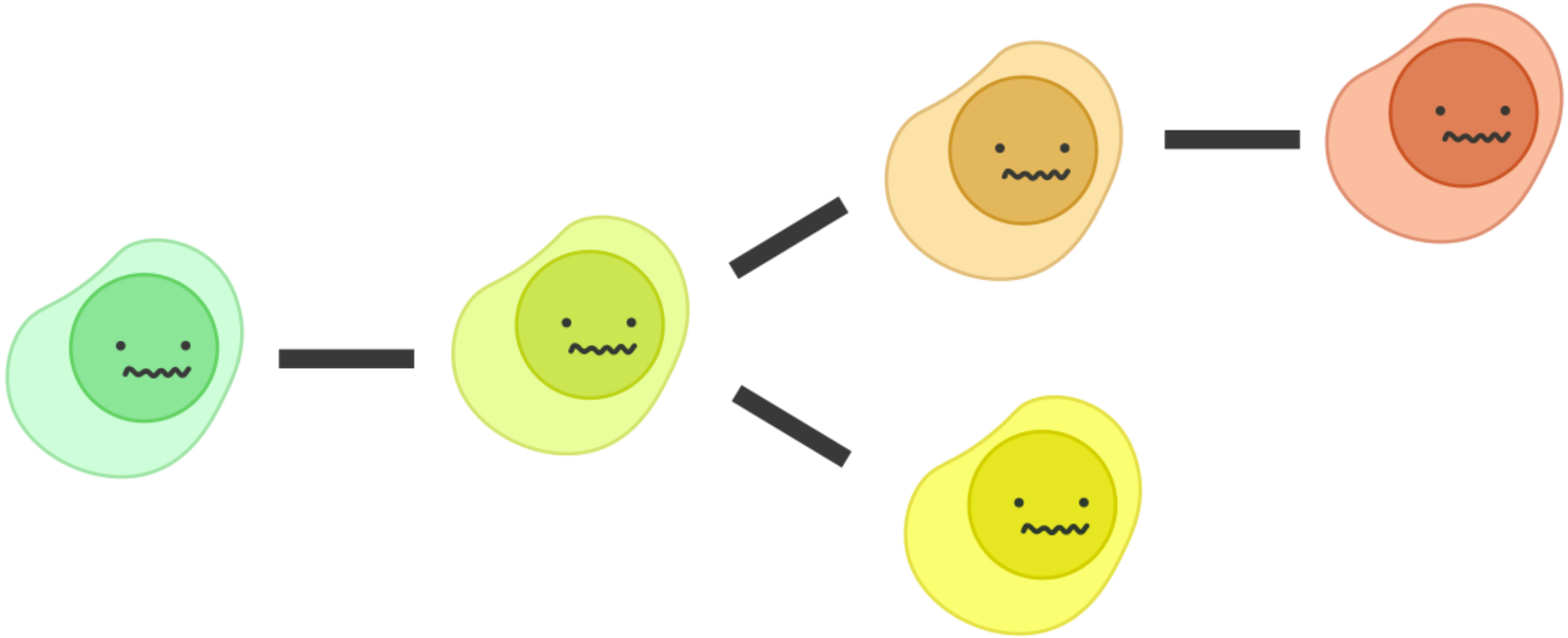
Sample

Sequence DNA

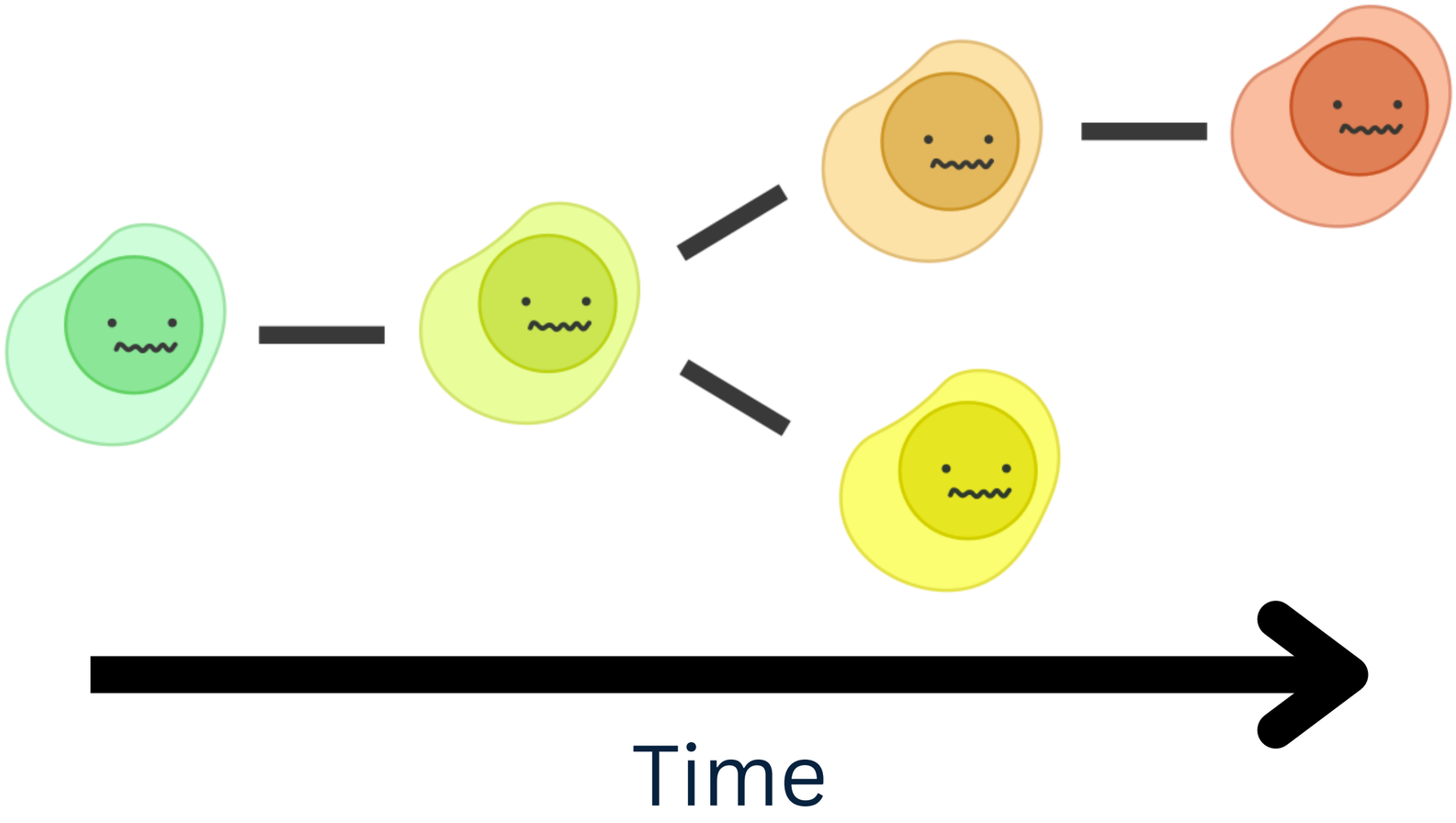
Identify mutations

# A tumour sample contains a mixture of cells!

# Cells in a tumour are all related

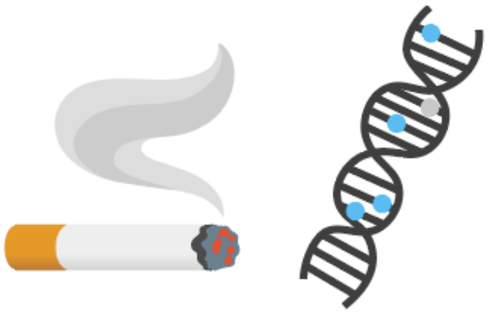


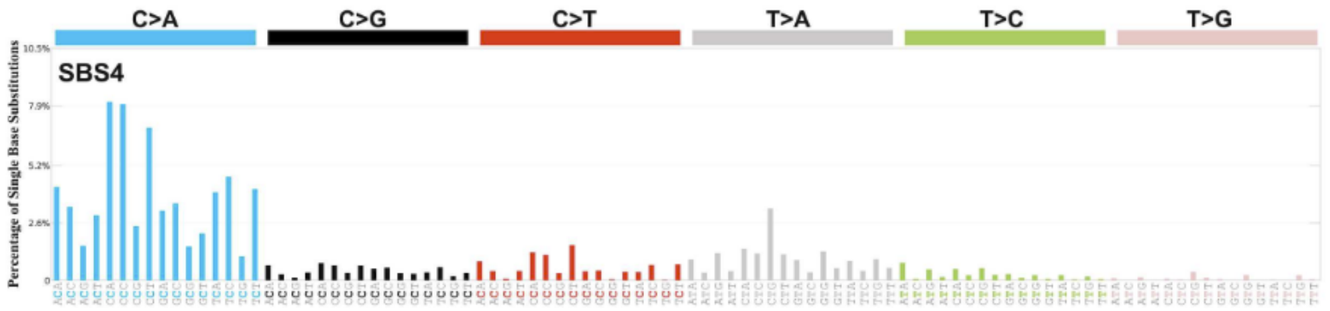
# Cells in a tumour are all related

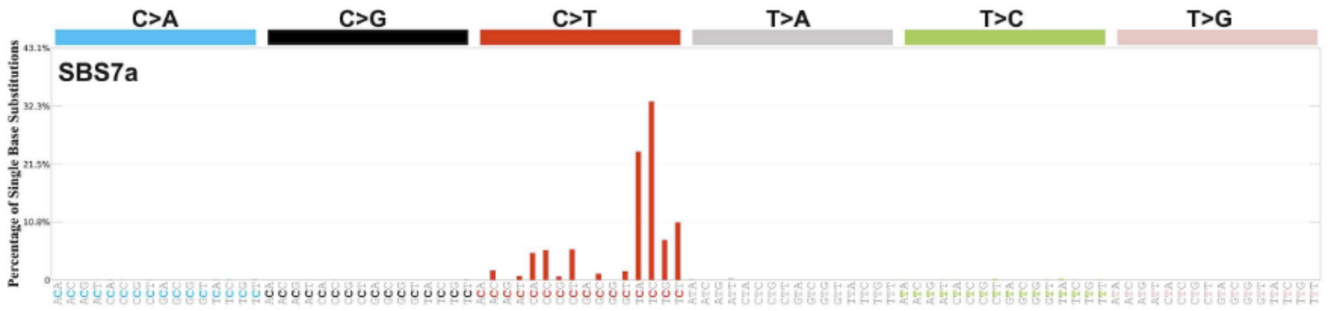


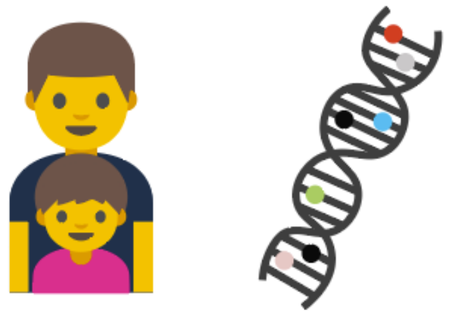

Mutations act as a “fossil record” of everything that happened to the tumour over its development.

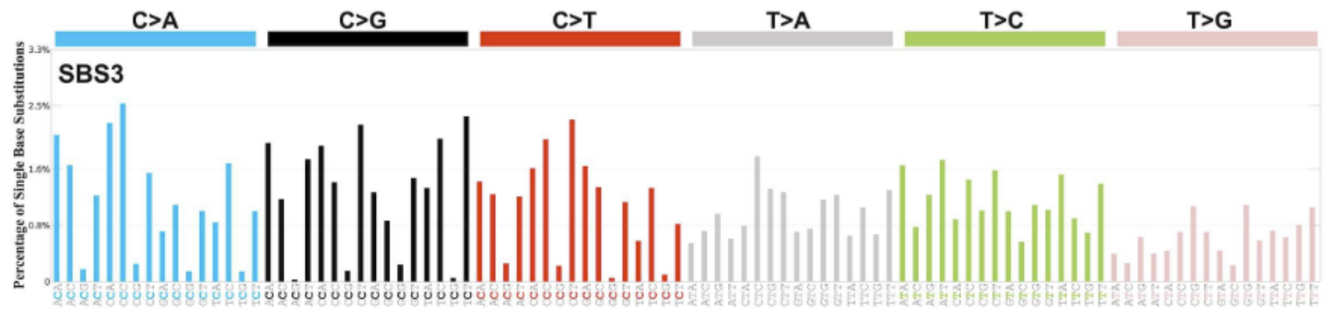
# Mutagens and DNA repair defects can cause distinct patterns of mutations



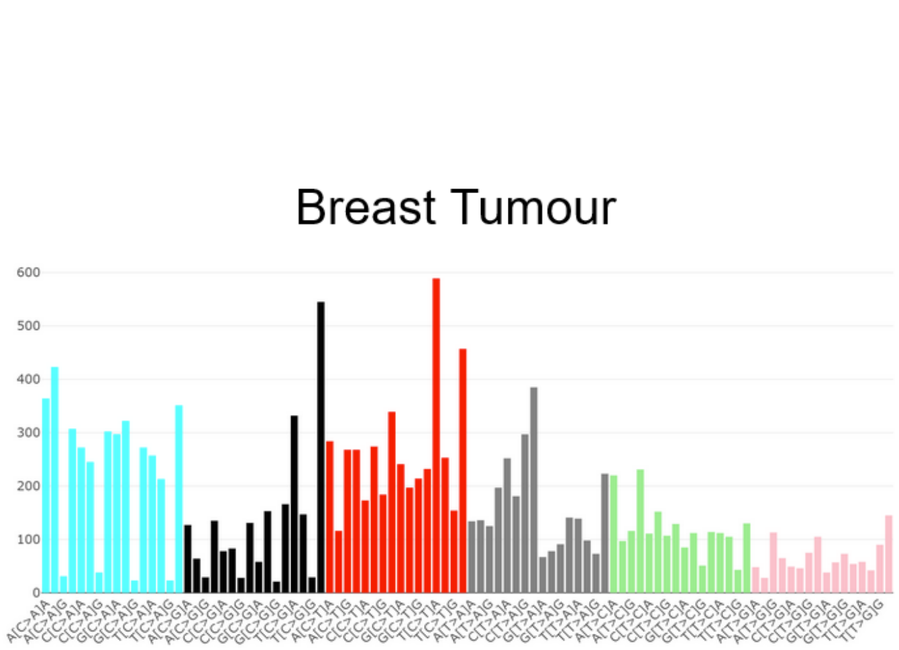




# Mutational signature analysis = identify these patterns



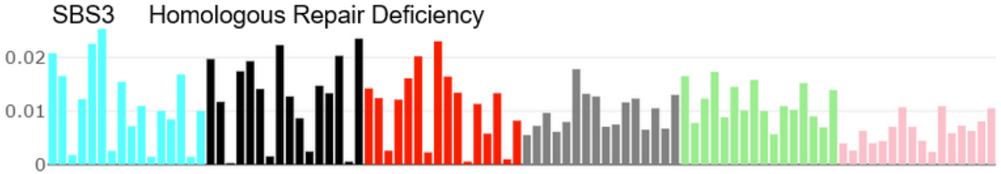

25%

30%

15%

23%

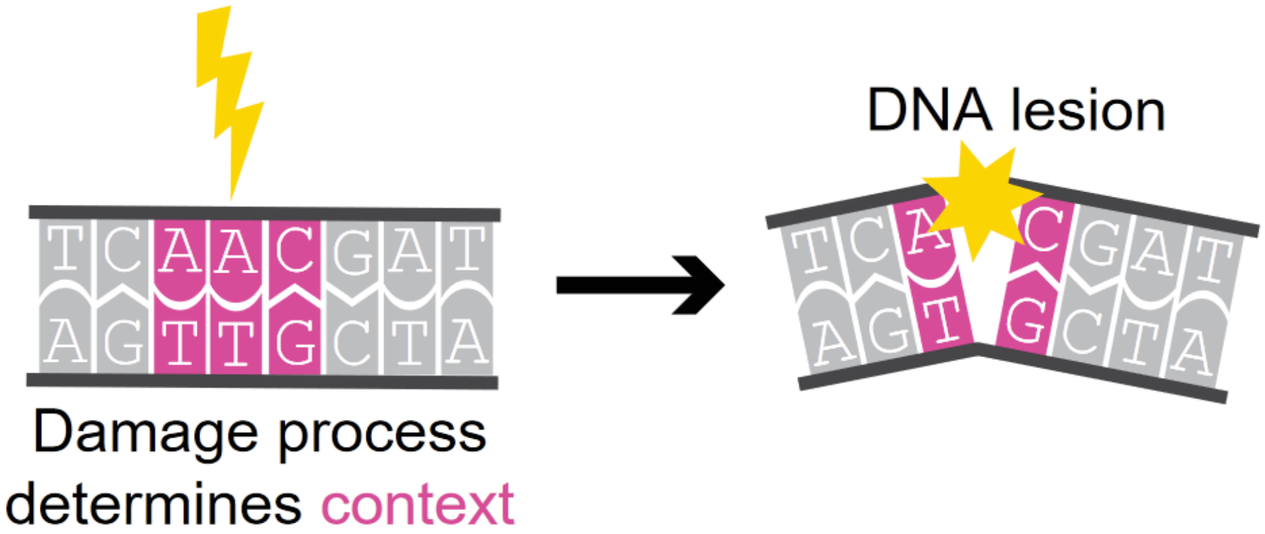
7%

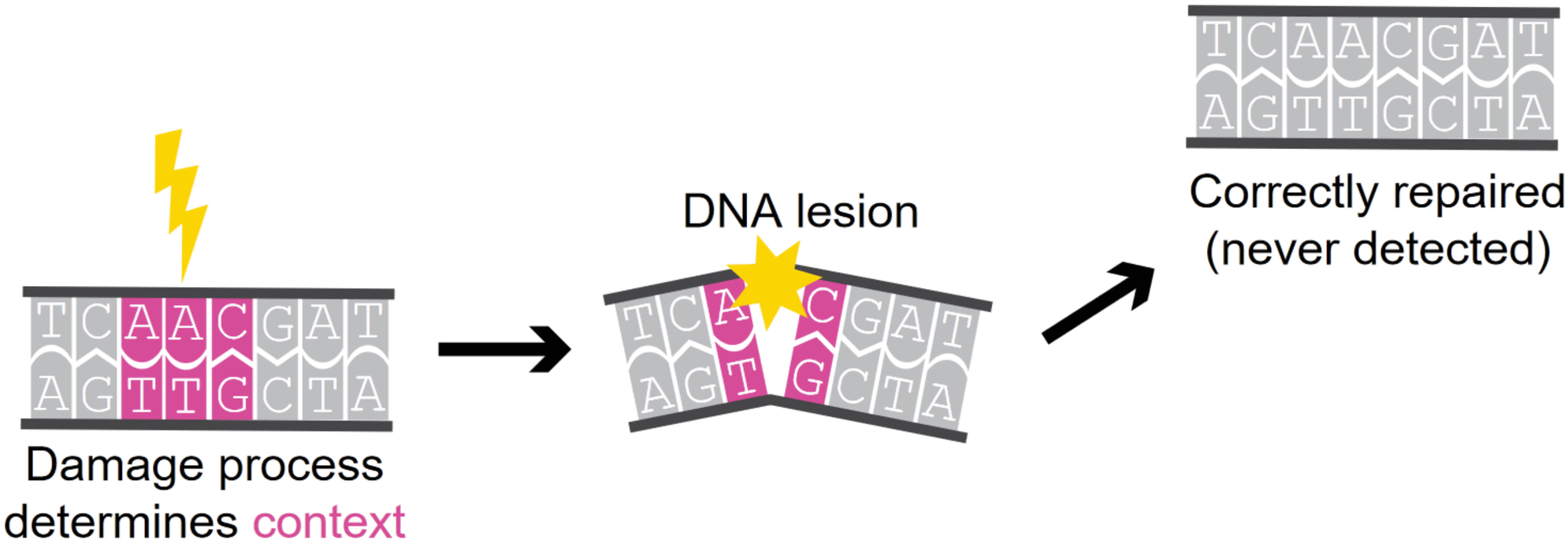



Other

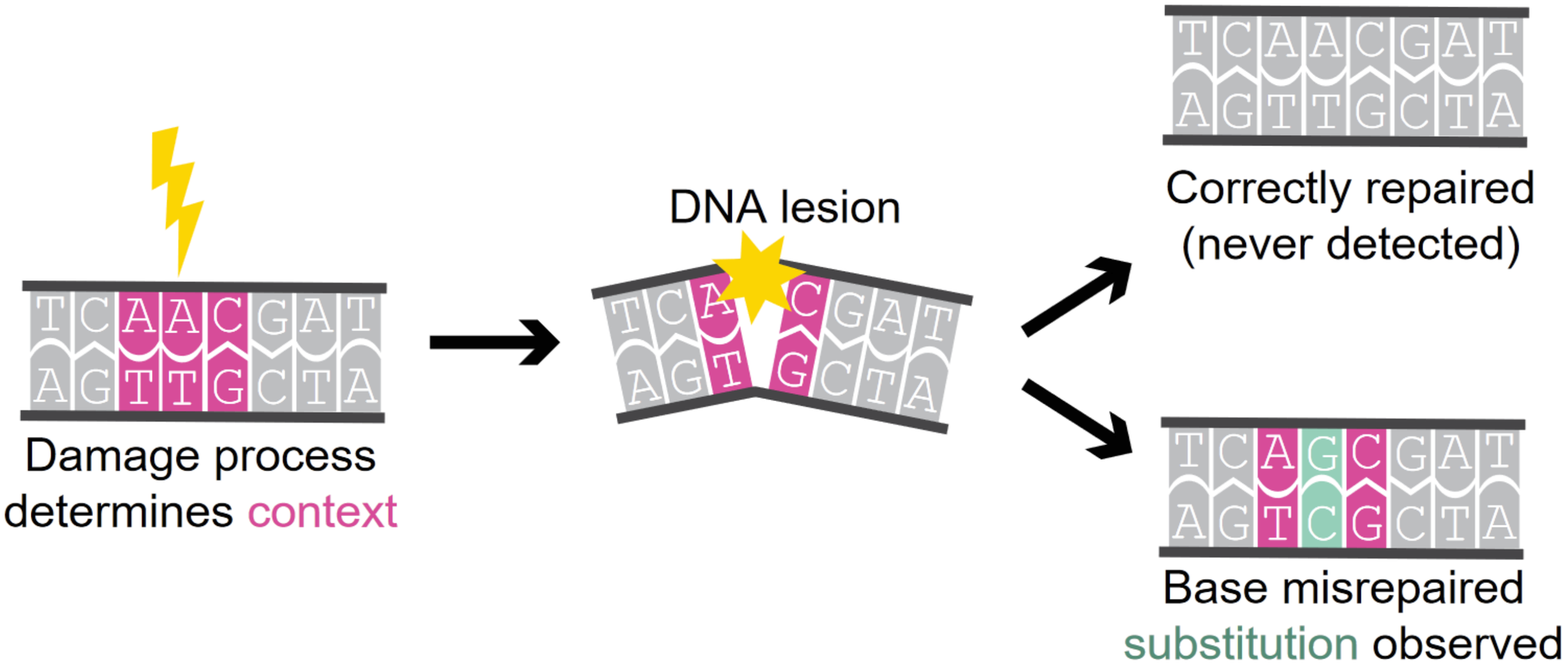
# Problem: current standard model mixes the effects of damage and misrepair processes



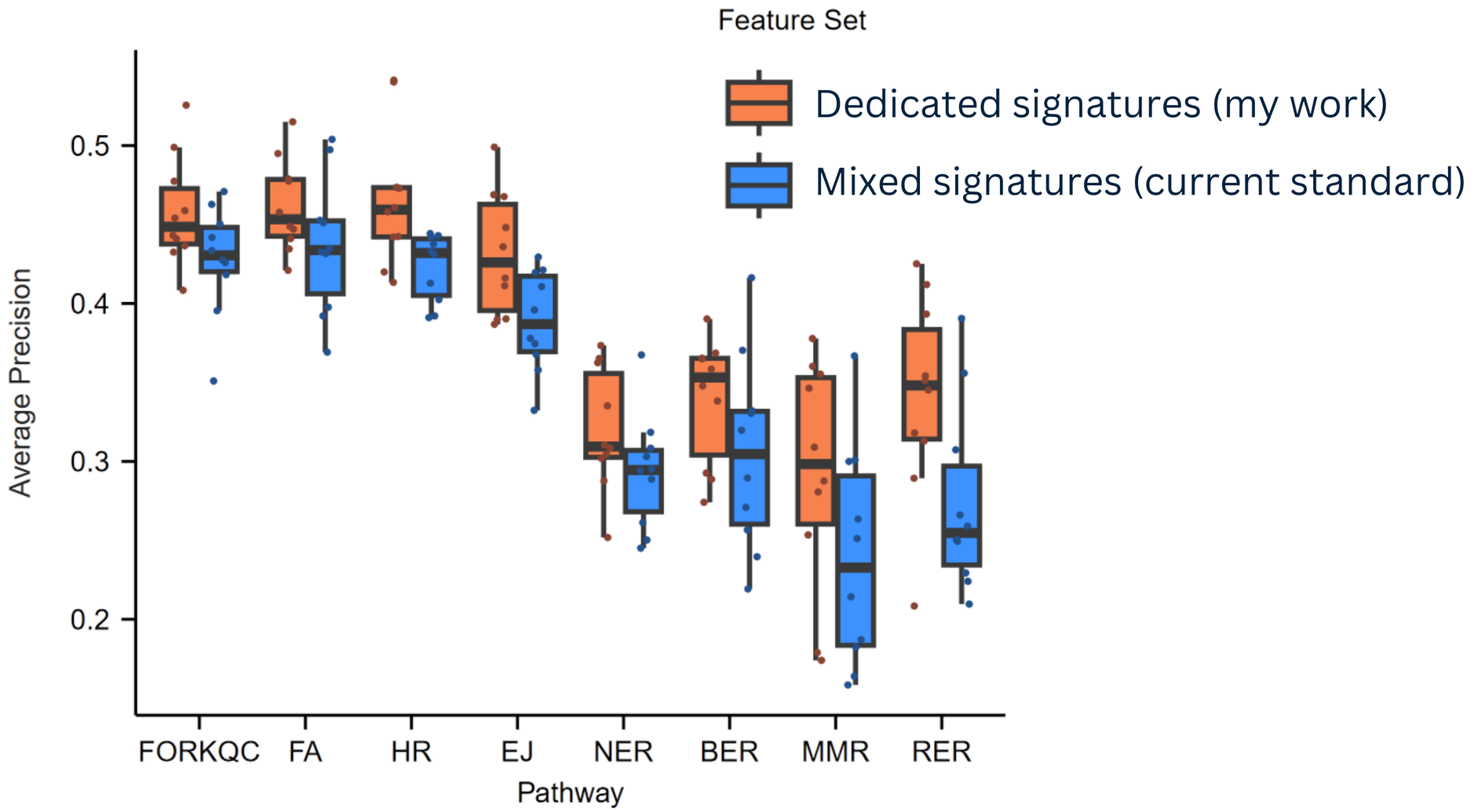
# Problem: current standard model mixes the effects of damage and misrepair processes

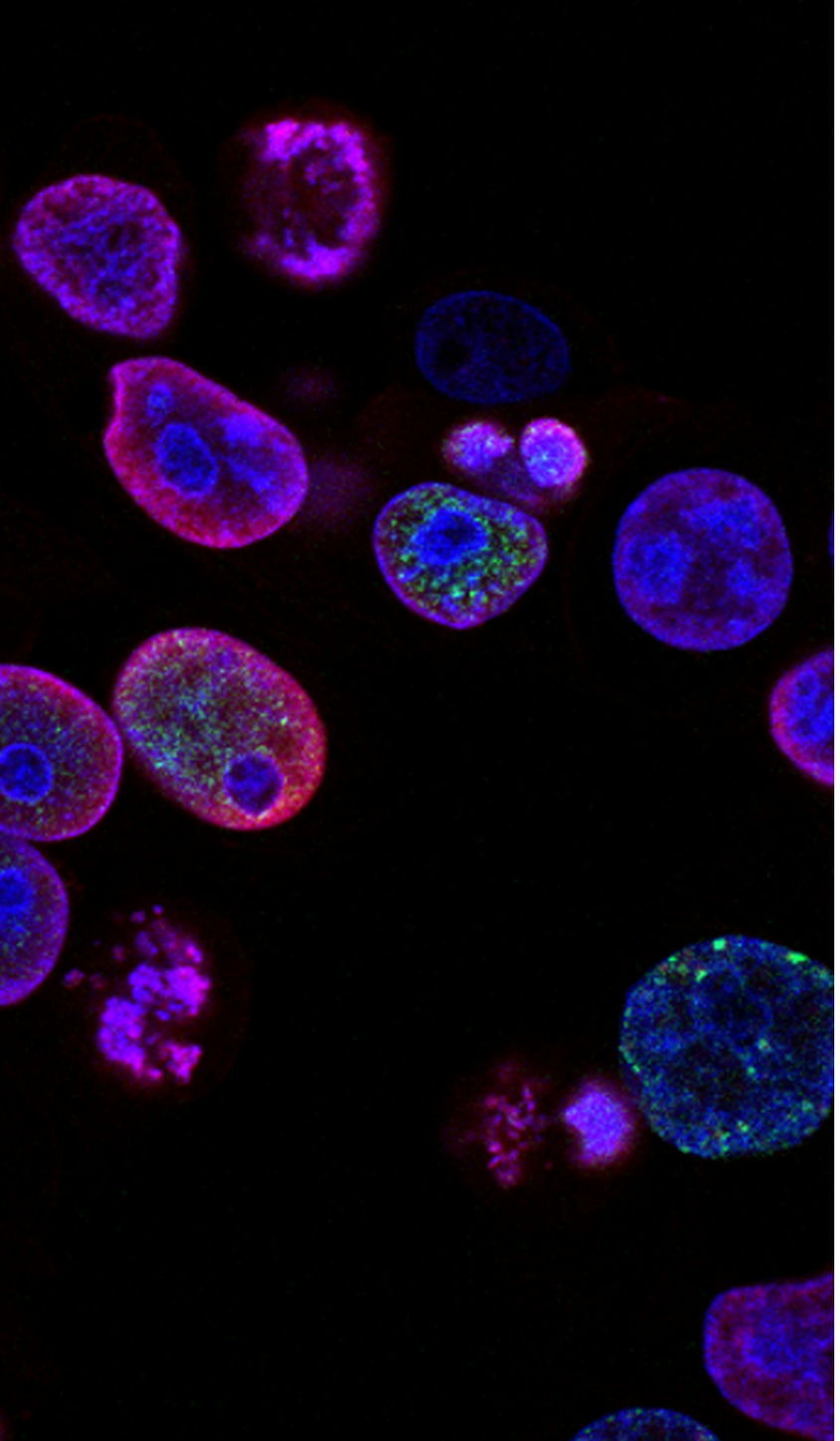


# Problem: current standard model mixes the effects of damage and misrepair processes



# Having dedicated damage and repair signatures helps predict repair deficiencies in cancer





# Agenda

Part 1: Brief overview of the field

Part 2: My research - cancer evolution

Part 3: A day in the life of a PhD student

## PhD life

The good: you get to work with amazing people, on problems you think are cool.

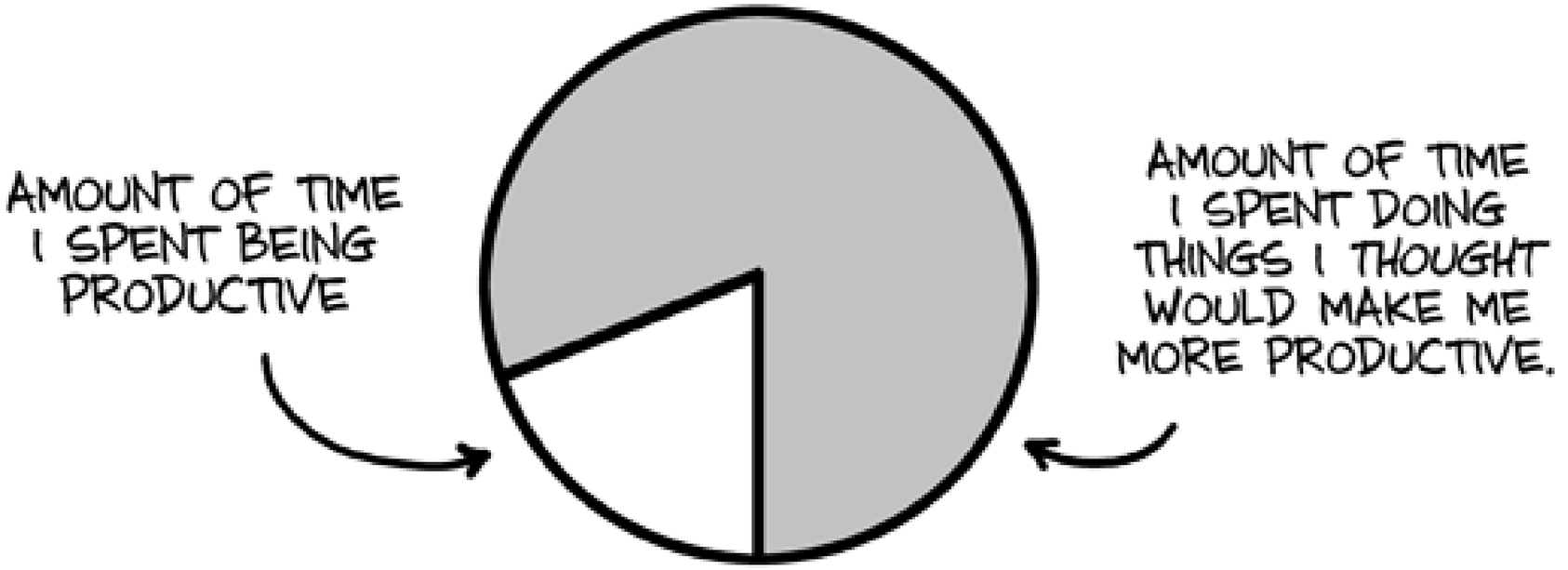
The bad: It can feel like a group project where you're the only one who cares if you get a good grade

The ugly: it takes 4+ years

# How I spend my time

- A **lot** of reading papers
- Writing code, doing data analysis
- Making plots & figures
- Writing results and scholarship applications
- Talking to other scientists

# HOW MY WEEK WENT:




WWW.PHDCOMICS.COM

JORGE CHAM © 2014

## Things you can do now to prepare for grad school

- Take a project-based class to get good at working independently
- Read papers
- Attend talks (ex. monthly meeting of torbug.org)
- Join a research lab

## Things you can do now to prepare for grad school

- Take a project-based class to get good at working independently
- Read papers
- Attend talks (ex. monthly meeting of torbug.org)
- Join a research lab

**Think about reference letters!**

**Email**

cait.harrigan@mail.utoronto.ca

**Web**

caitharrigan.ca