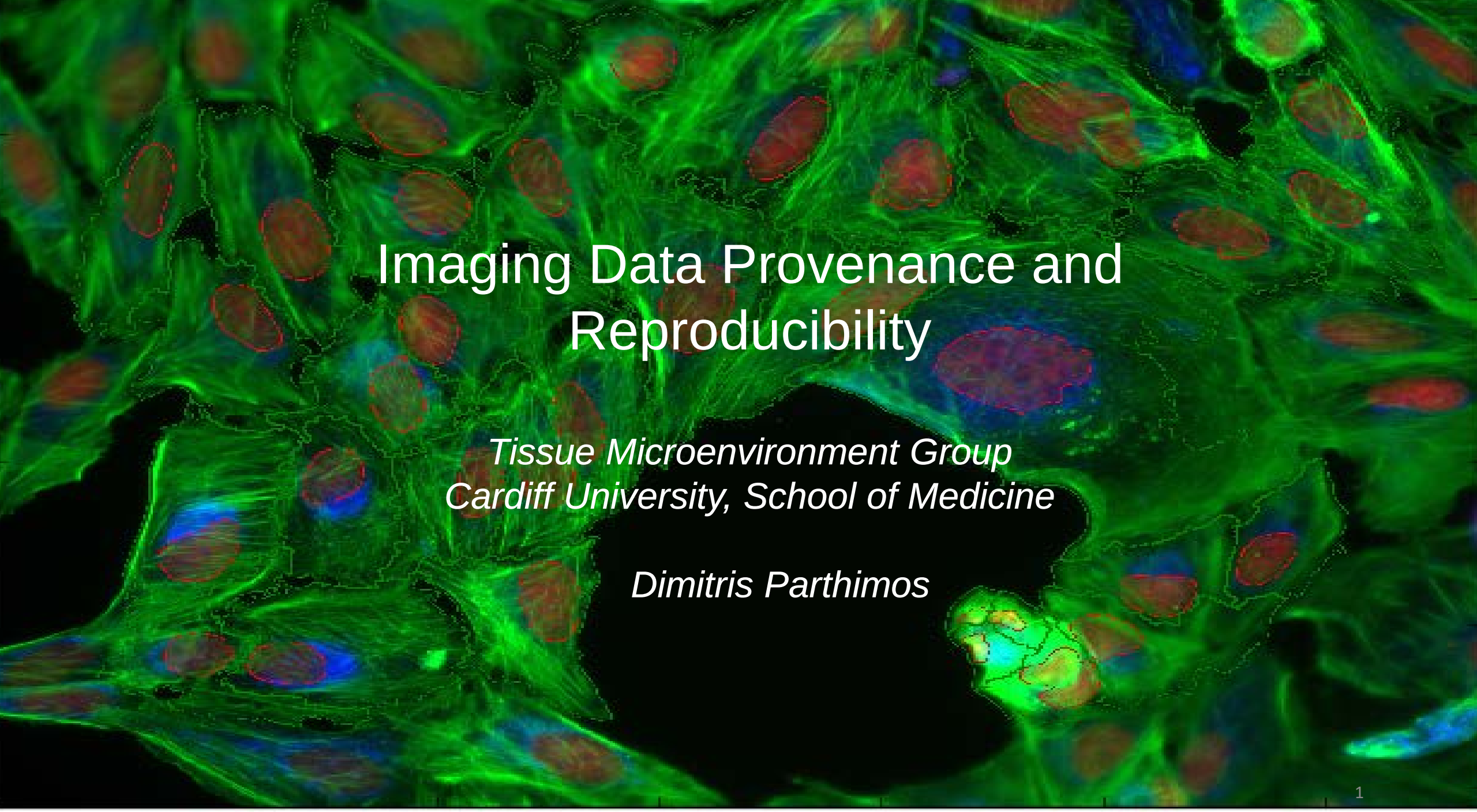
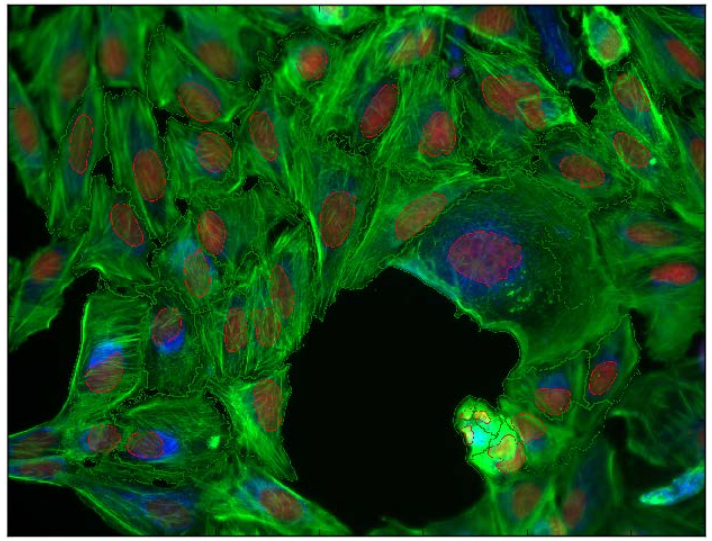


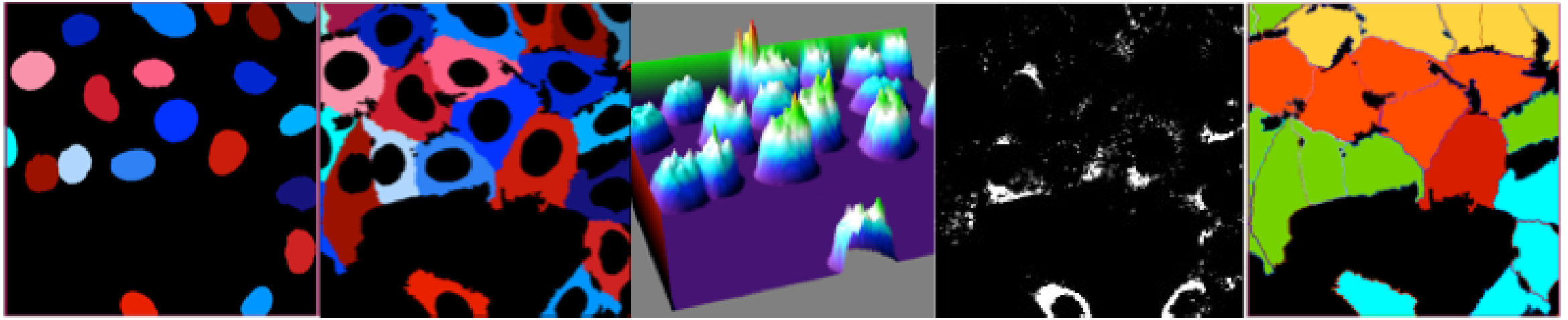
Imaging Data Provenance and Reproducibility

Tissue Microenvironment Group
Cardiff University, School of Medicine

Dimitris Parthimos

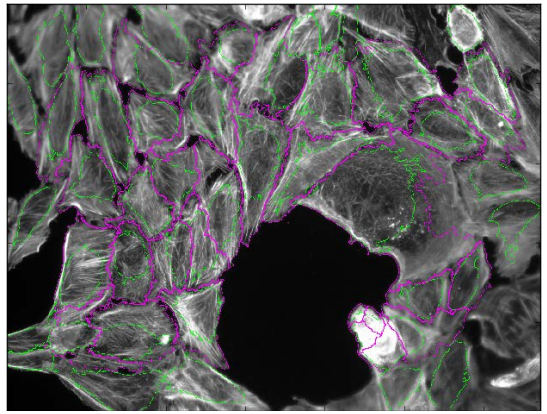
Typical analysis pipeline

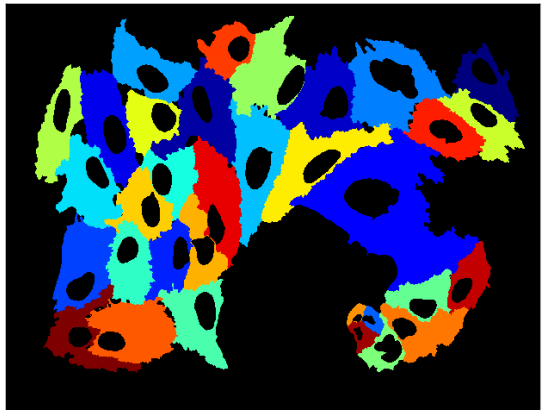


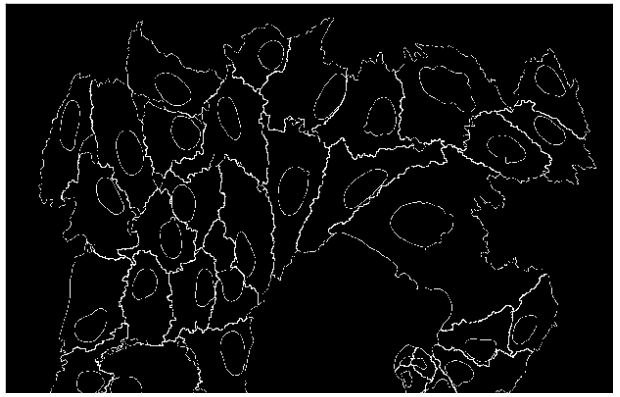


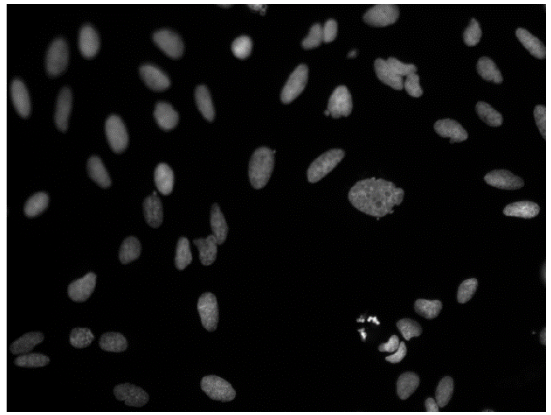
Counts, Sizes, Shapes, Intensities, Textures, Correlations, Neighborhoods

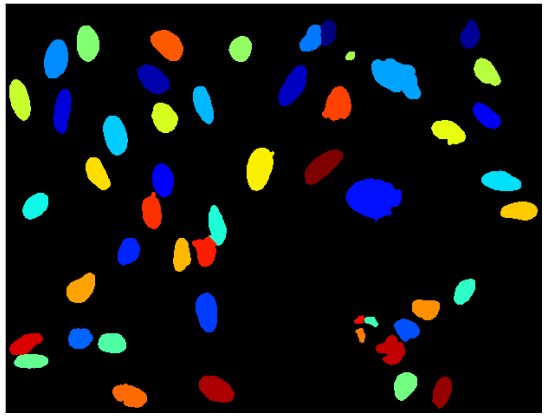
Image analysis - Pipeline in Action (CellProfiler)

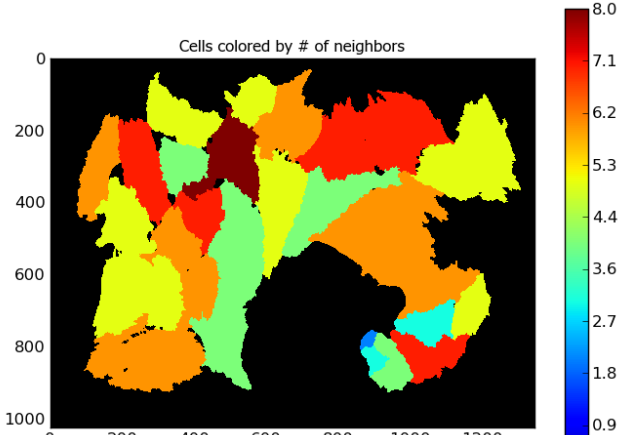


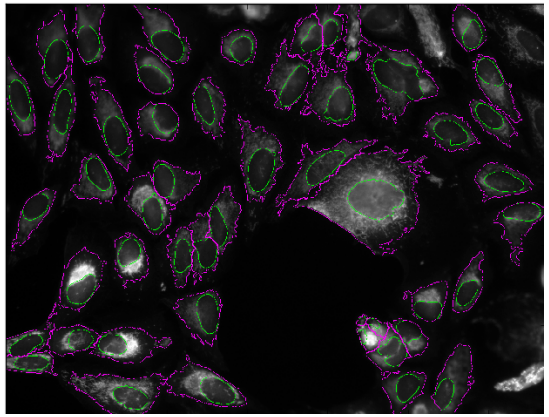


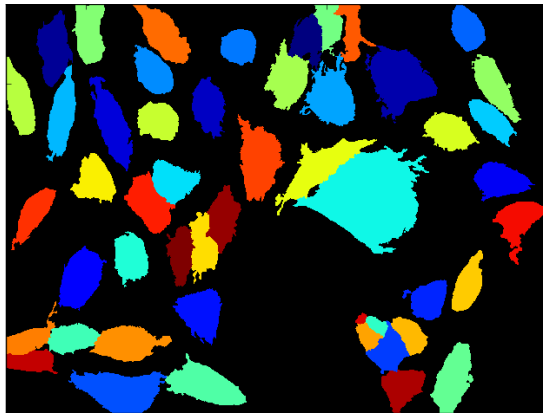


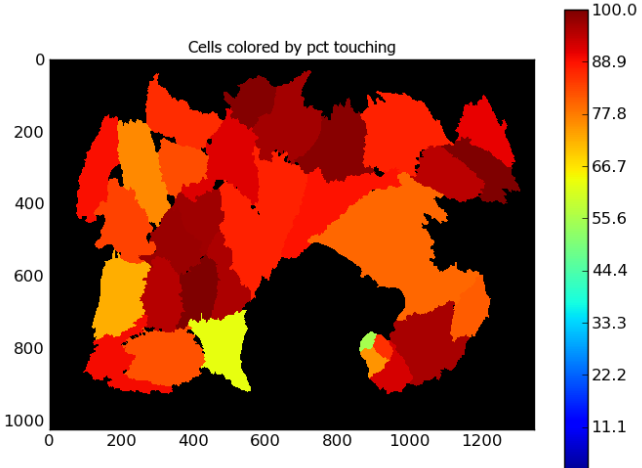




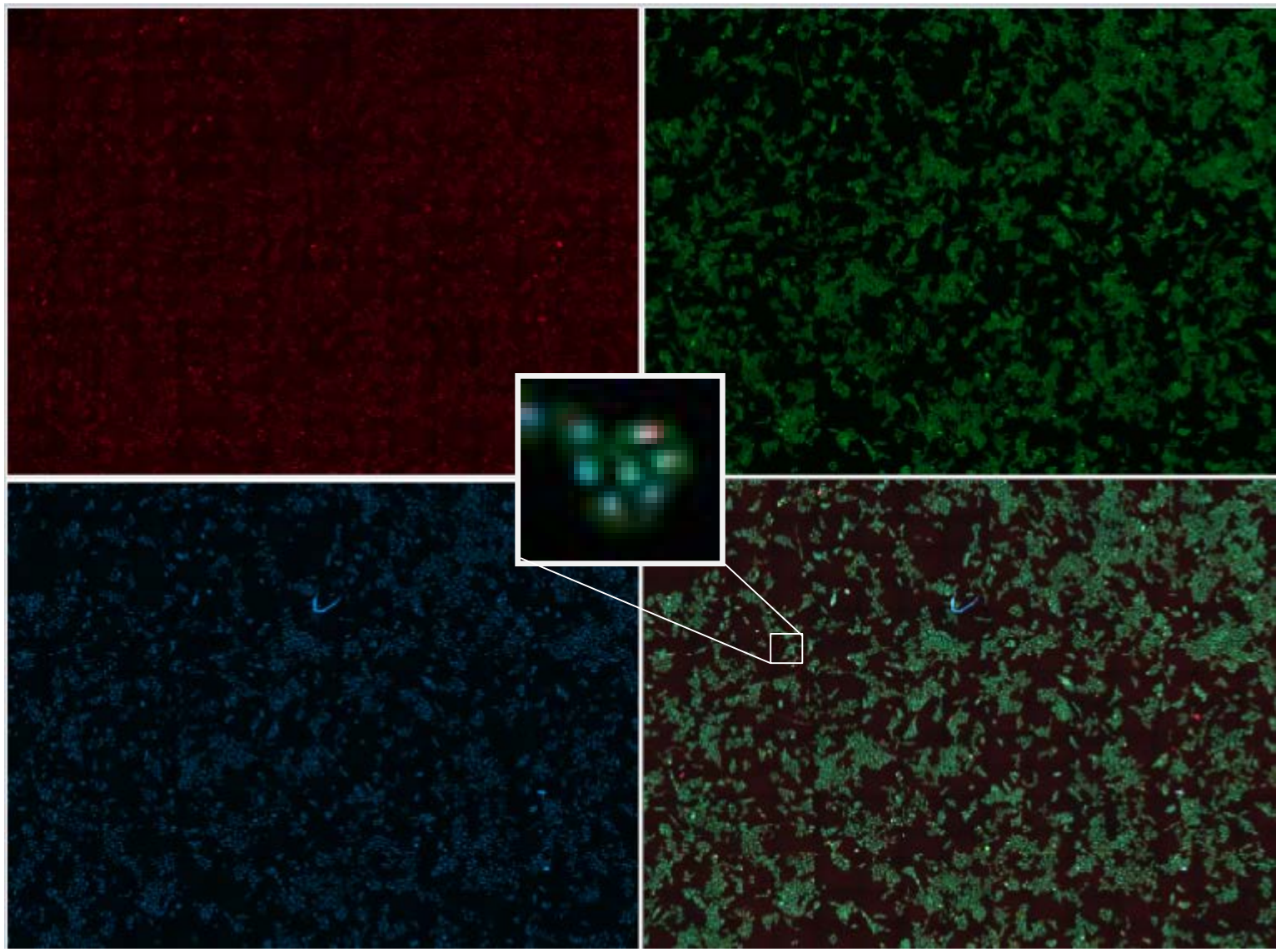








Data collected at NIST: Static data



design/meta data

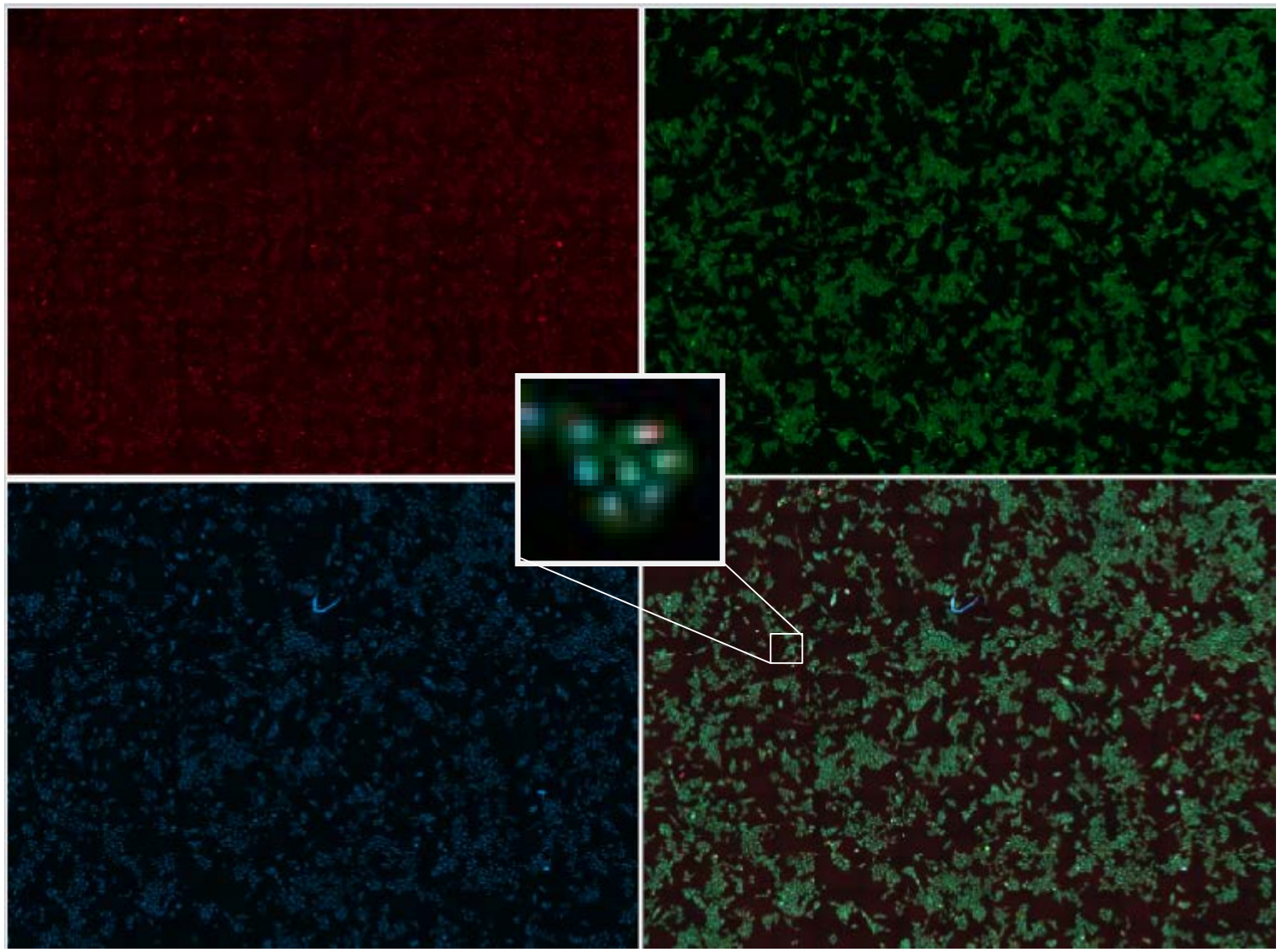
reference material/data

image analysis algorithms

new cell paints

patient cells

Data collected at NIST: Testing stability of data



Original data
190 GB

Reference data
3.5 GB

In silico
generated data
569 GB

Other channels
x3

Multiple
experimental
data

Case Study 2

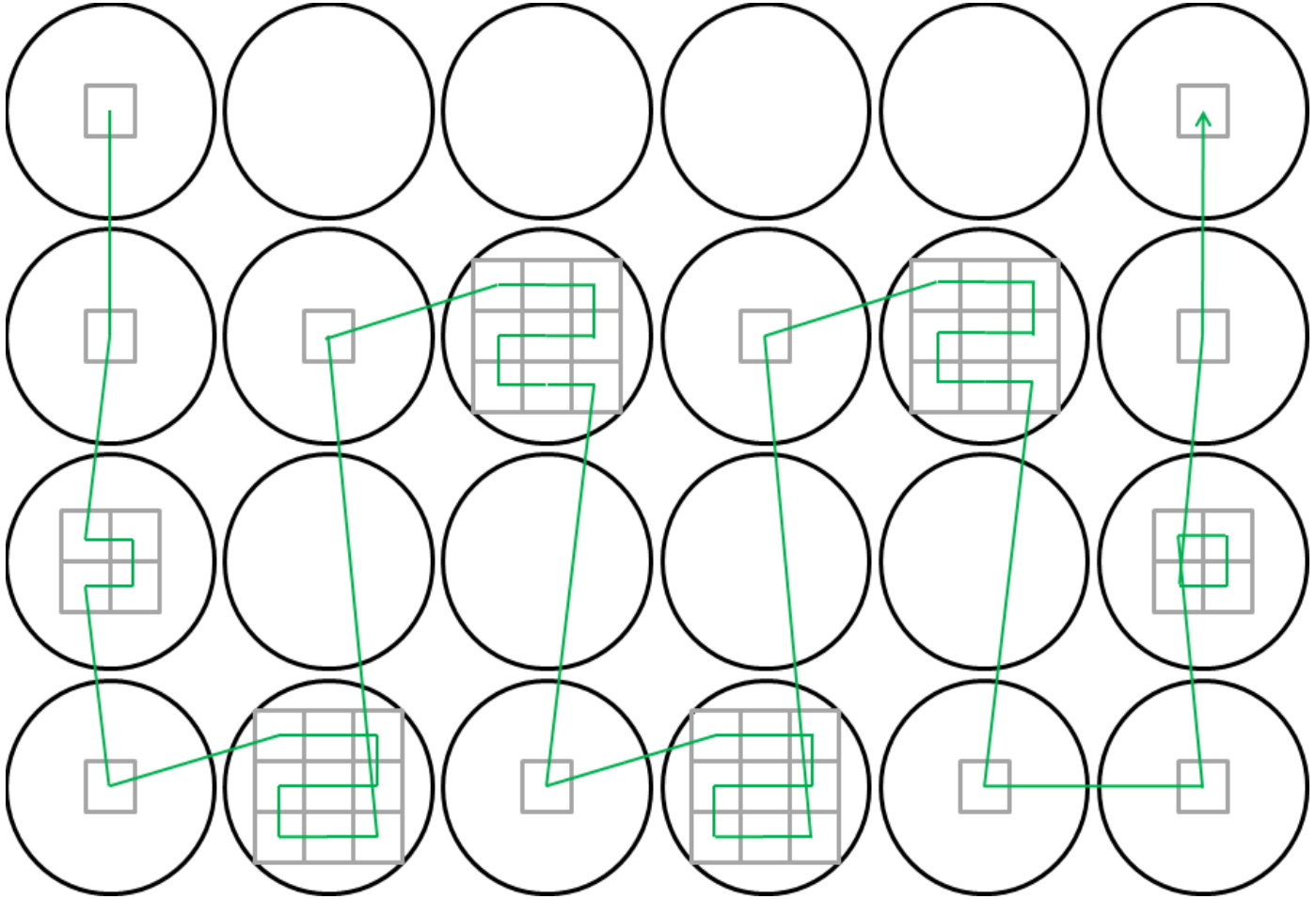
Timelapse microscopy of live cells responding to drug

R.A. Howard-Jones, Marie Wiltshire, A.J. Sloan and R.J. Errington
Cardiff University, School of Medicine,
M.R. Brown and Matthieu Duteil
Swansea University, School of Engineering

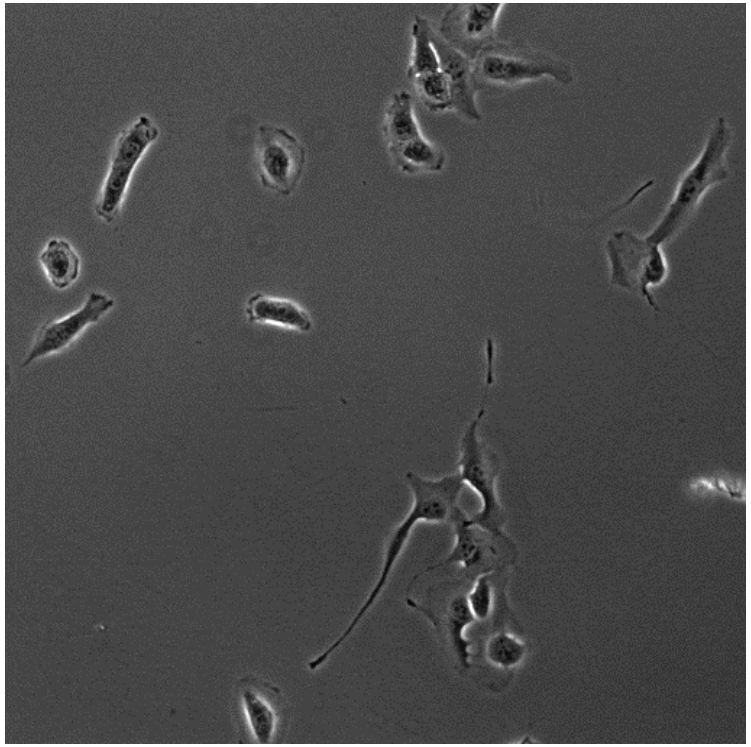
J Elliott and Michael Halter
Cellular Biosystems, NIST

Peter Bajcsy and team
Information technology, NIST

Collected on a high-throughput screening instrument designed to automatically collect an image sequence as a movie



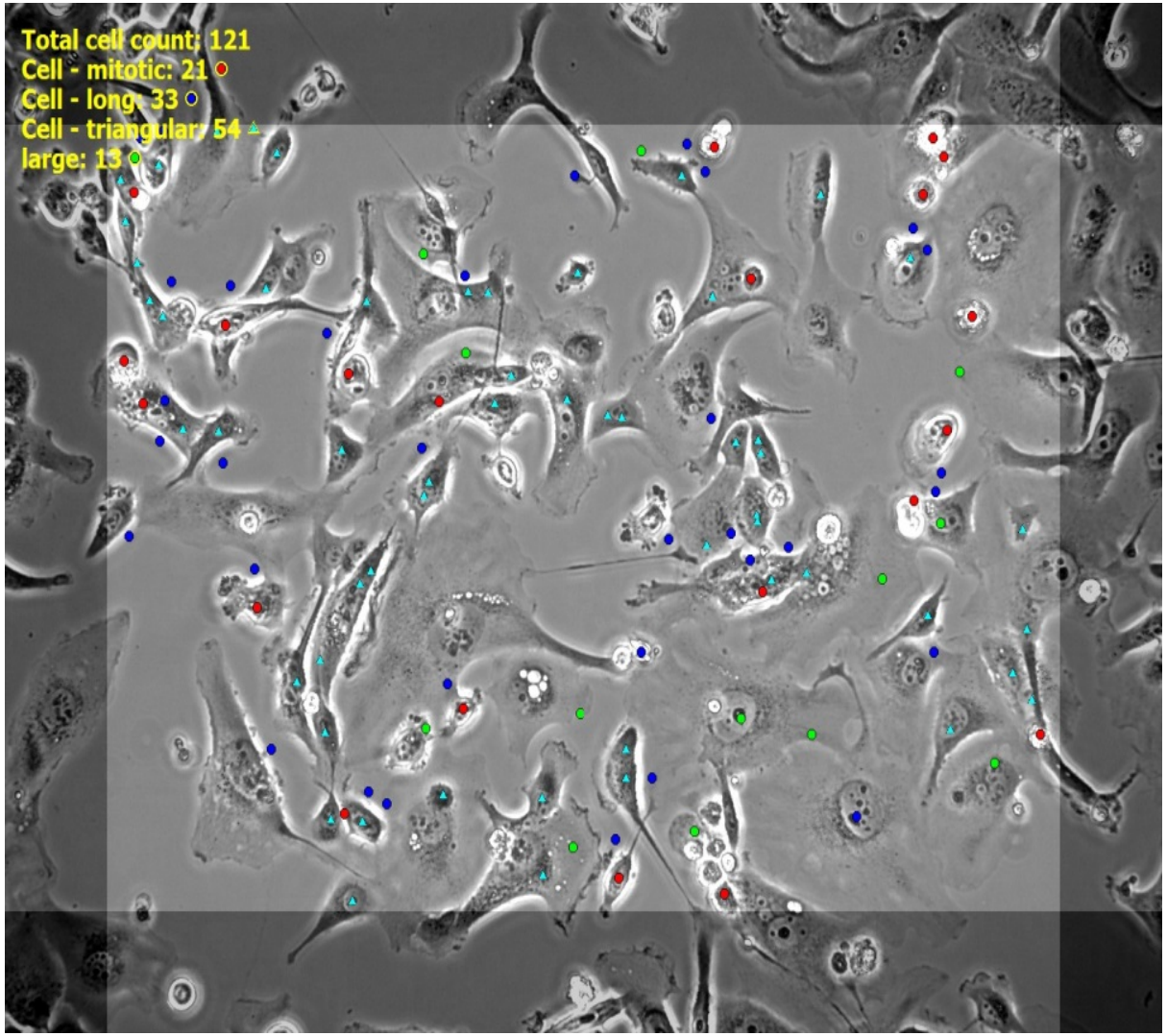
Typical video time lapse of cancer cell line



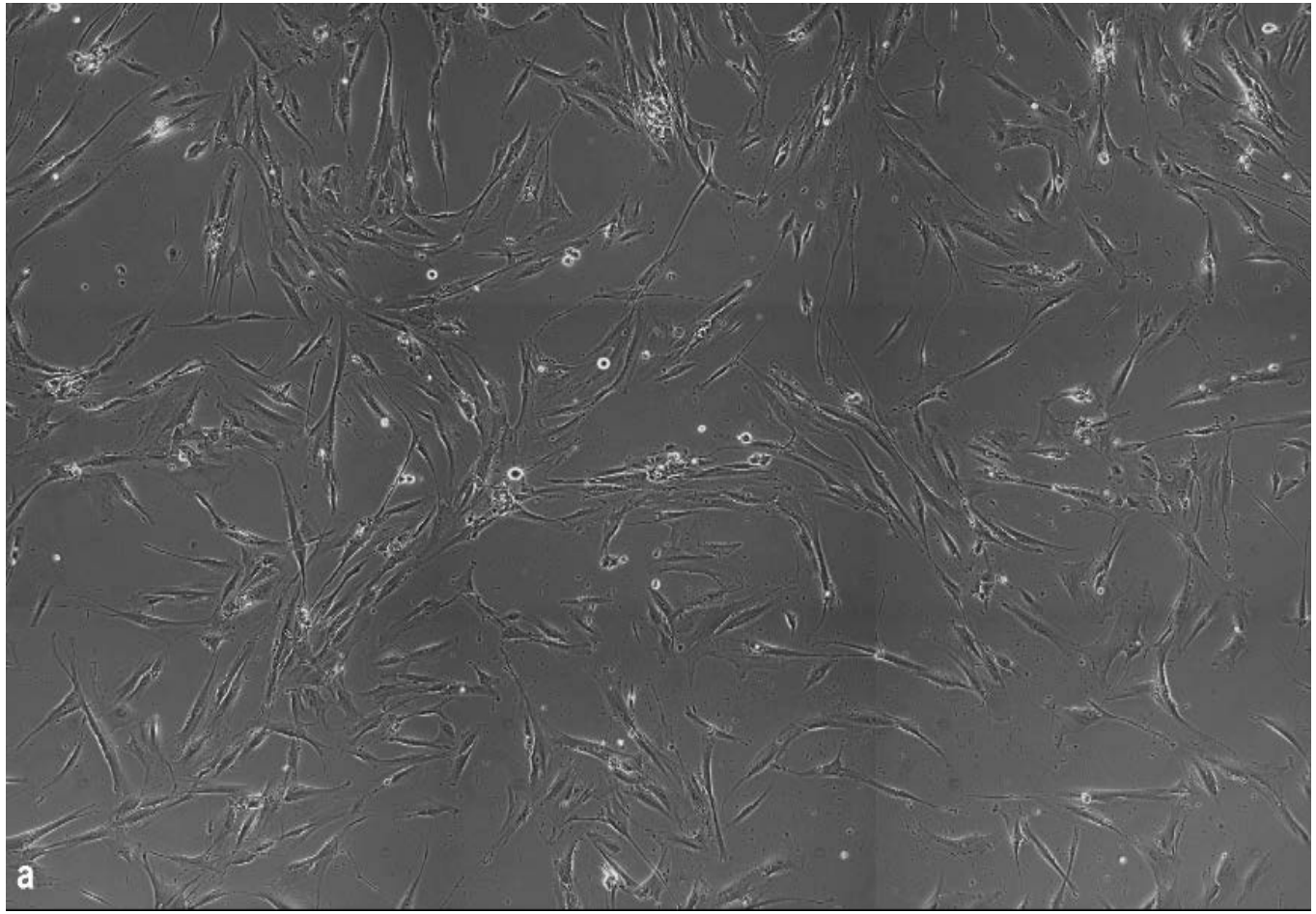
What do we want to extract from such timelapse image data

- Derive behavioural features of cells in control and perturbed conditions
- Obtain features of clonal expansion and understand this expansion at different granularities, including lineages
- In cancer cells: understand the derivation (dynamics) of the polyploidy phenotype, and measure the neighbourhood in which these cells survive.
- Patterns of drug resistance, temporal and spatial characteristics of which are poorly understood

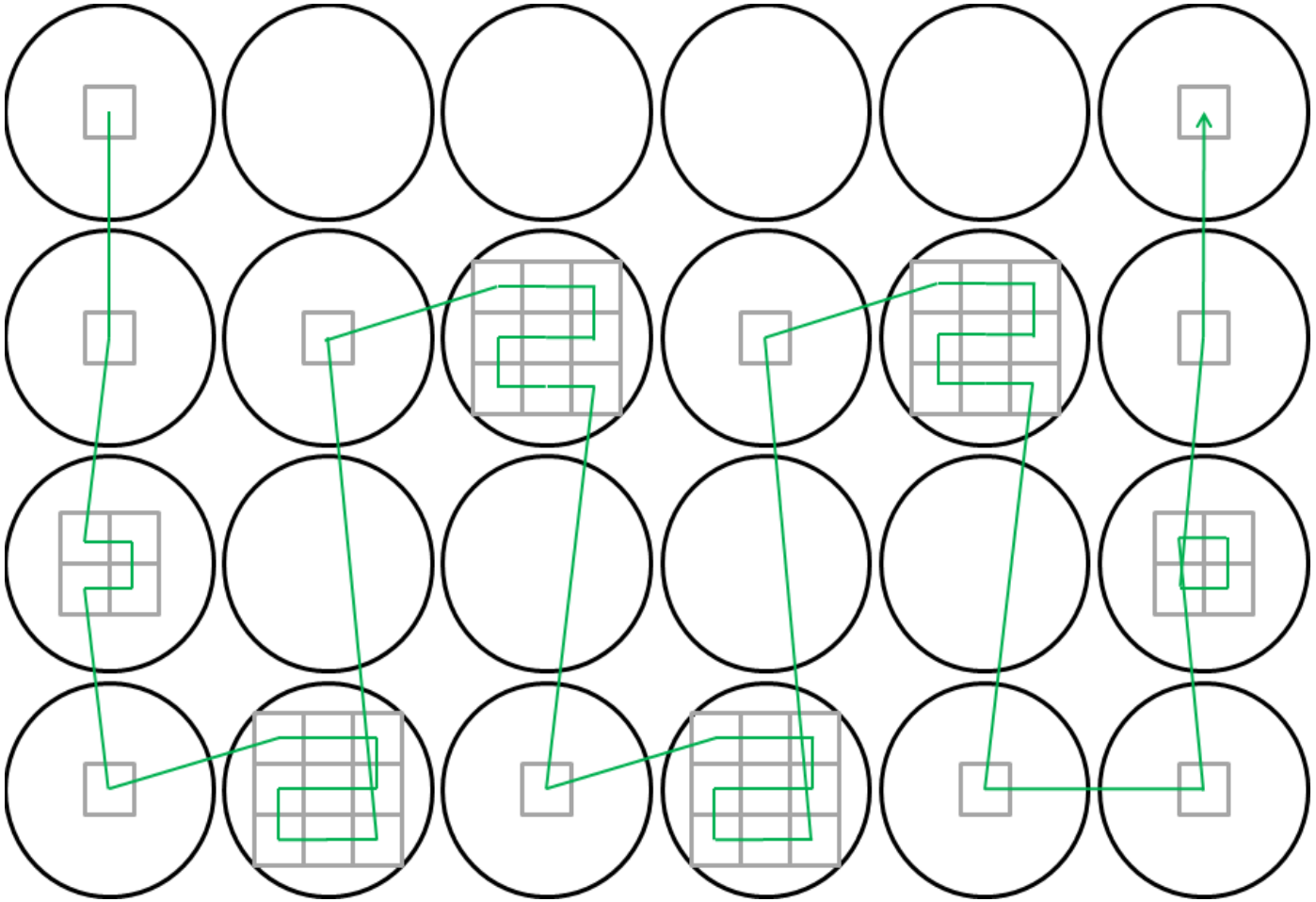
Automated cell detection and tracking



Tiled 3x3 FOV



Collected on a high-throughput screening instrument designed to automatically collect an image sequence a movie



For this one experiment: 108 FOV collected every 30 minutes for 120.5 hours: 2 channels
Each FOV = 1392*1040 therefore entire dataset is 72 Gigabytes

2 datasets one for cancer cell line and one for primary mesenchymal stem cells – 144GB

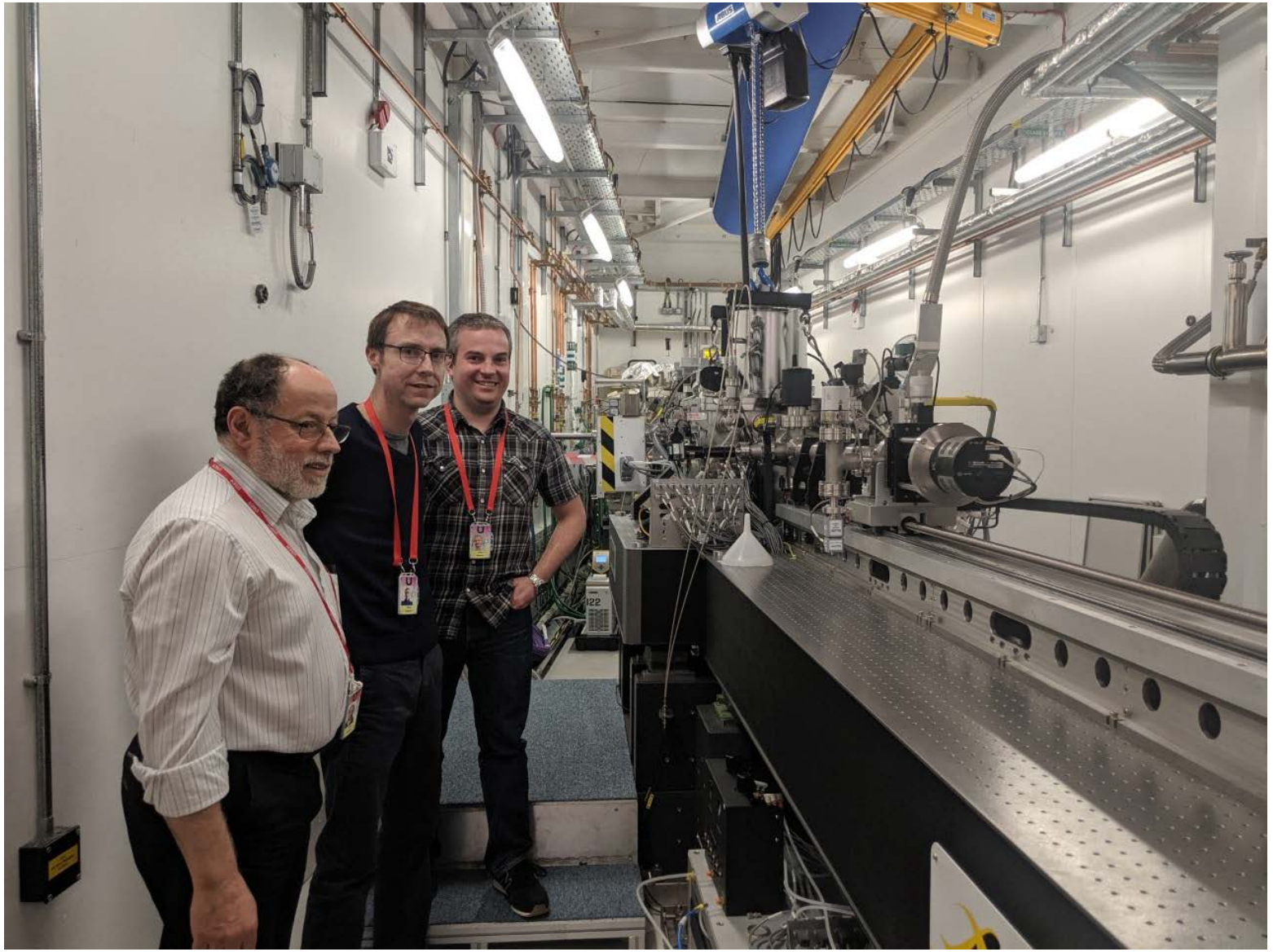
Case Study 3

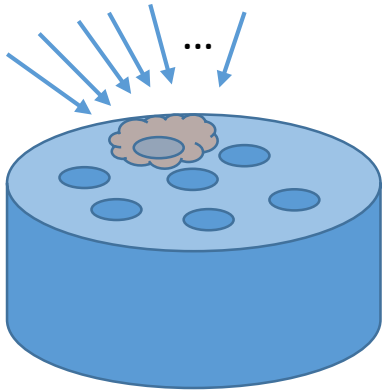
Extracellular vesicle activity associated with cancer resistance

Aled Clayton, Jason Webber
Cardiff University, School of Medicine,

Diamond

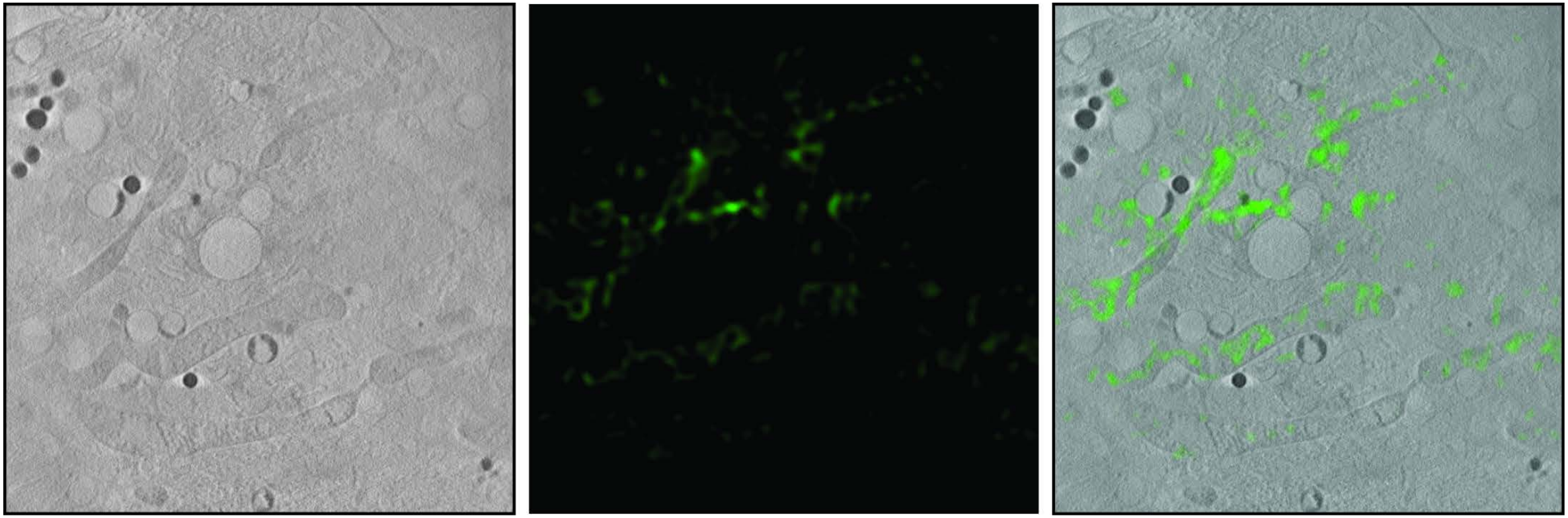
Fluorescent microscope overlaid with x-ray beam





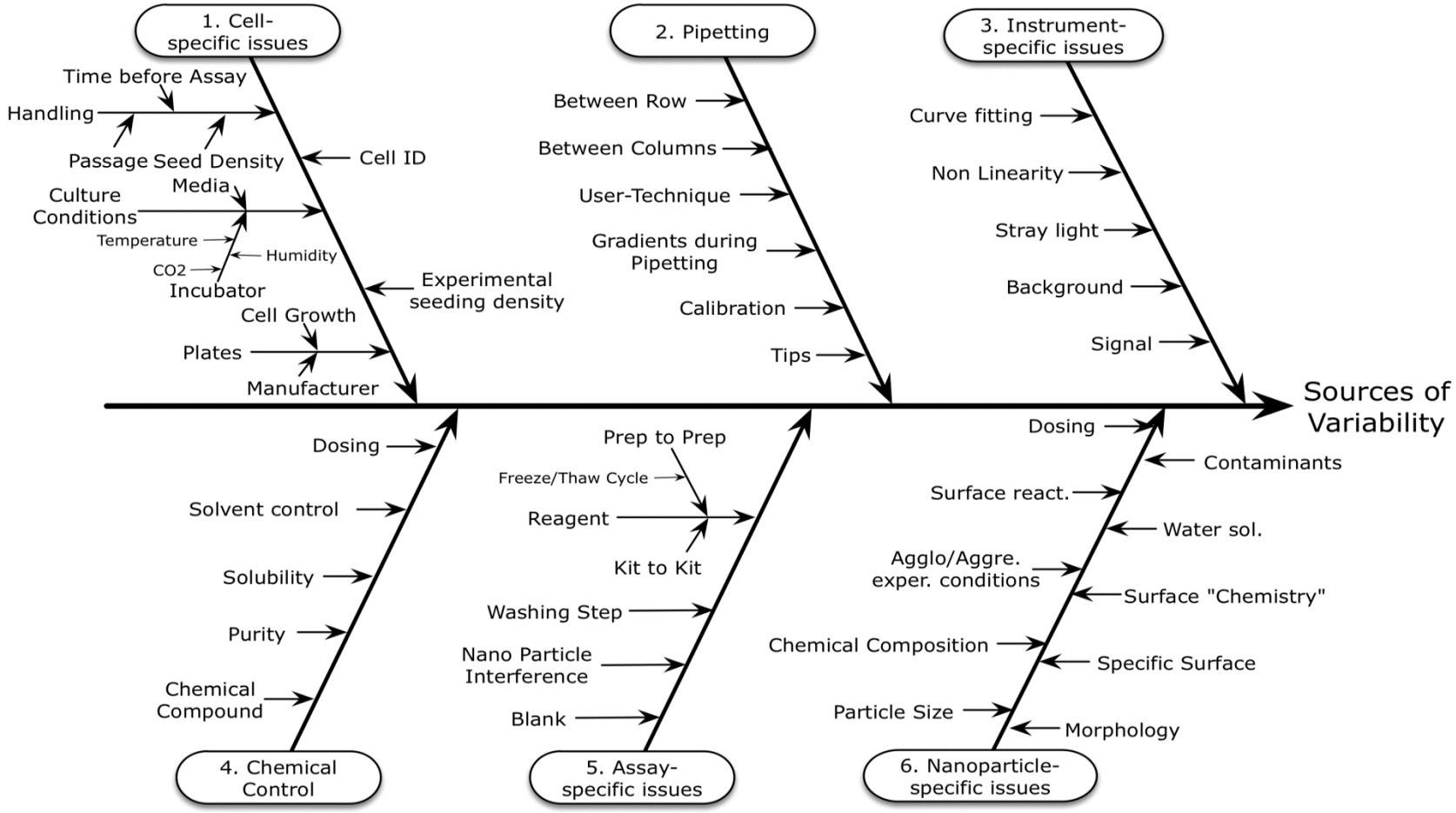
- *Cryo Soft X-ray Tomograph*
- *Cryo Structured Illumination Microscope*

Fluorescent microscopy overlaid with x-ray beam image



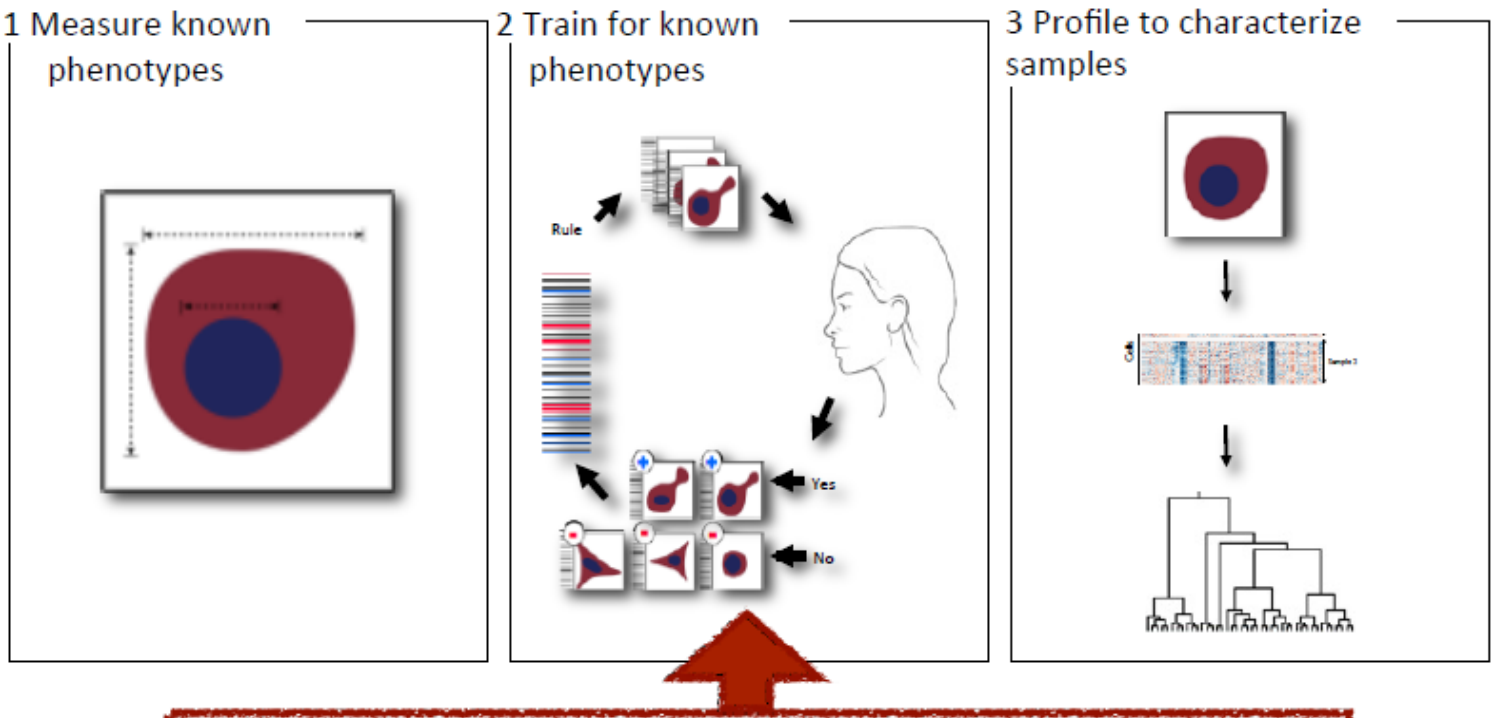
±60º at 0.2º equivalent to 1.2 GB data size (x 100 cells x treatments)

Identify sources of variability in assay



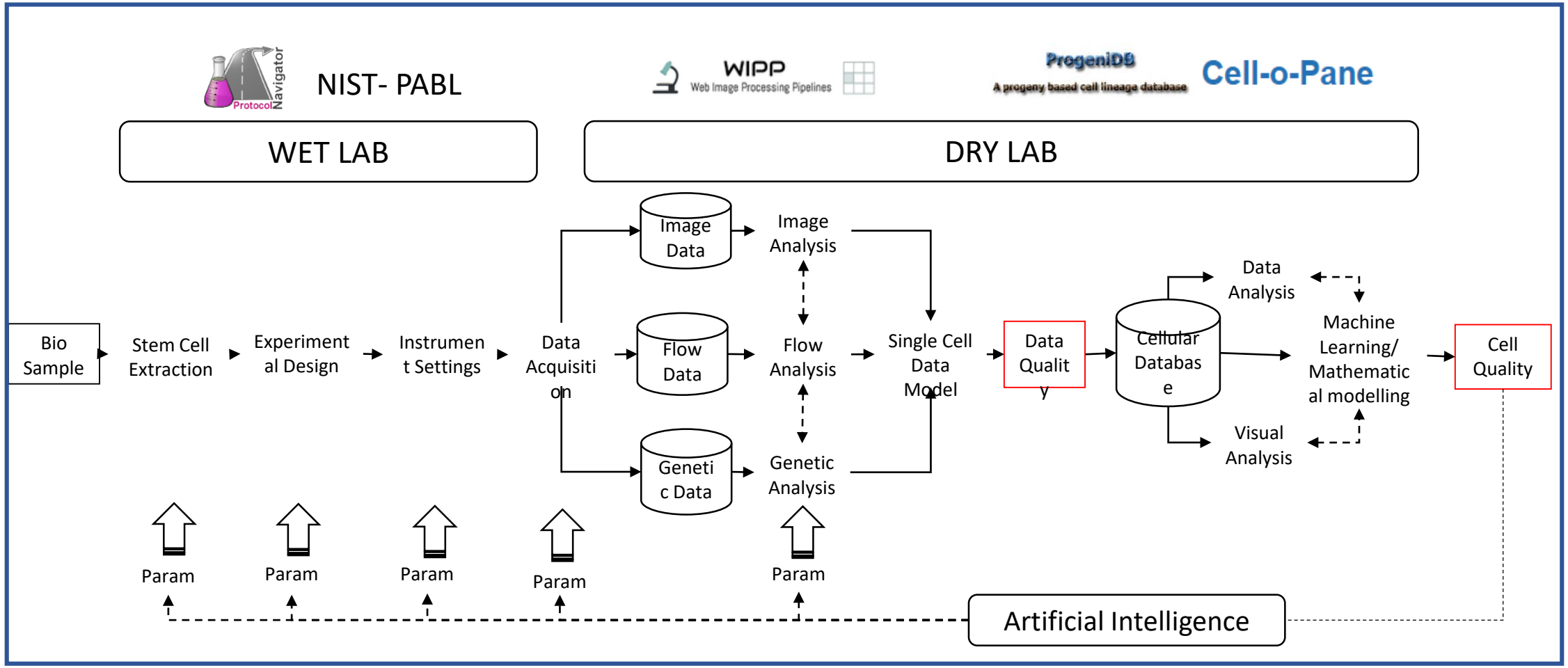
Cause and effect diagram

Three waves of quantitative image analysis



Measure everything, then use machine learning to distinguish a phenotype of interest

BROAD INSTITUTE


NIST- PABL
WIPP
Web Image Processing Pipelines
ProgeniDB
A progeny based cell lineage database
Cell-o-Pane
WET LAB
DRY LAB
Bio Sample
Stem Cell Extraction
Experimental Design
Instrument Settings
Data Acquisition
Image Data
Flow Data
Genetic Data
Image Analysis
Flow Analysis
Genetic Analysis
Single Cell Data Model
Data Quality
Cellular Database
Data Analysis
Visual Analysis
Machine Learning/ Mathematical modelling
Cell Quality
Param
Param
Param
Param
Param
Artificial Intelligence

Summary

- Bioimaging community faces an explosive growth in data size and complexity
- Reproducibility is a key concern; traceability of algorithm pipelines
- Fundamental need for tools such as Web Image Processing Pipeline (WIPP) and Automated Bio-Imaging Laboratories (NIST)
- Interactive measurements and AI-assisted discoveries over large image banks