# Machine learning in translational cancer research

**Krzysztof Pastuszak**[1,2,3] | **krzpastu@pg.edu.pl**

[1] **Department of Algorithms and System Modeling**

Gdansk University of Technology

[2] **Laboratory of Translational Oncology**

Institute of Medical Biotechnology and Experimental Oncology

University of Gdańsk and Medical University of Gdańsk

[3] **Center of Biostatistics and Bioinformatics**

Medical University of Gdańsk

RESEARCH UNIVERSITY
EXCELLENCE INITIATIVE
Ministry of Science and Higher Education

GDAŃSK UNIVERSITY OF TECHNOLOGY

www.gumed.edu.pl

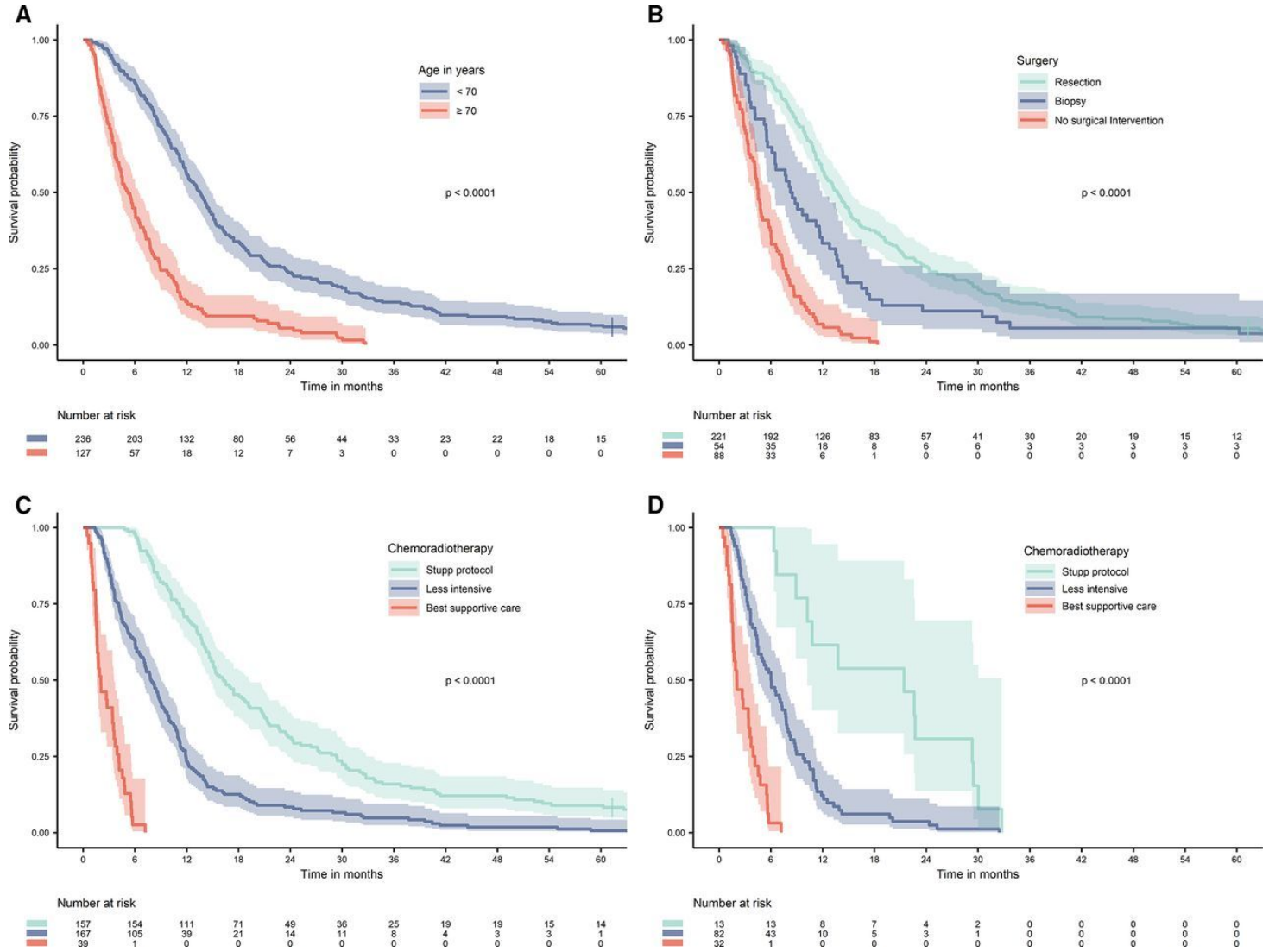
# Motivation

## Glioblastoma

- Grade 4 central nervous tumor
- Responsible for 14.5% of CNS tumors
  - 48.6% of malignant ones
- Prognosis remains very poor even with early diagnosis
- Different molecular mechanisms involved warrant research into drug combinations

# Motivation

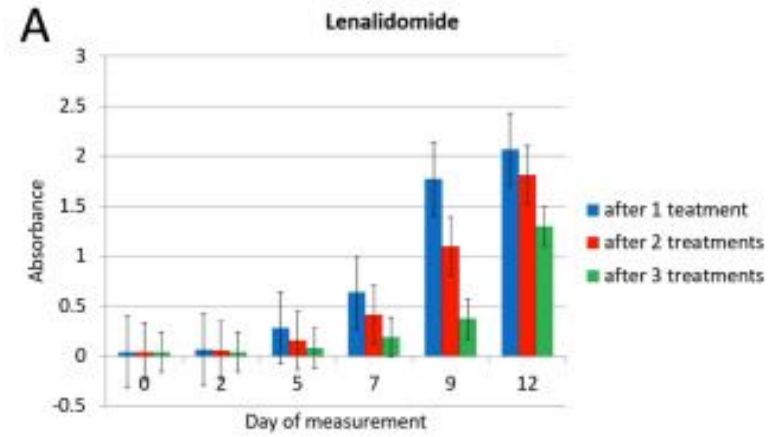
## Glioblastoma

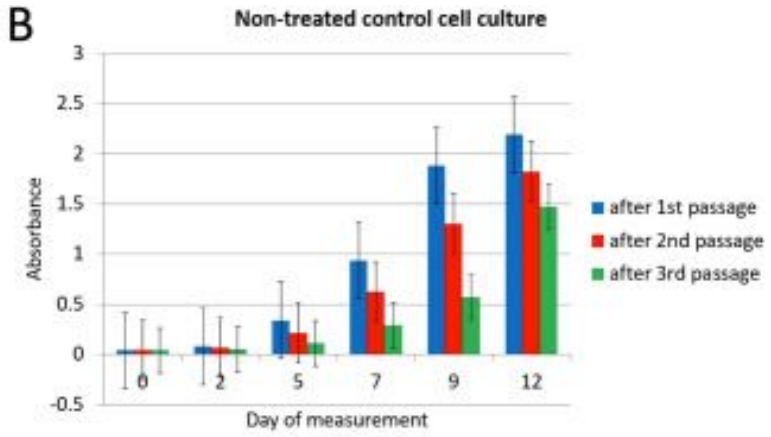


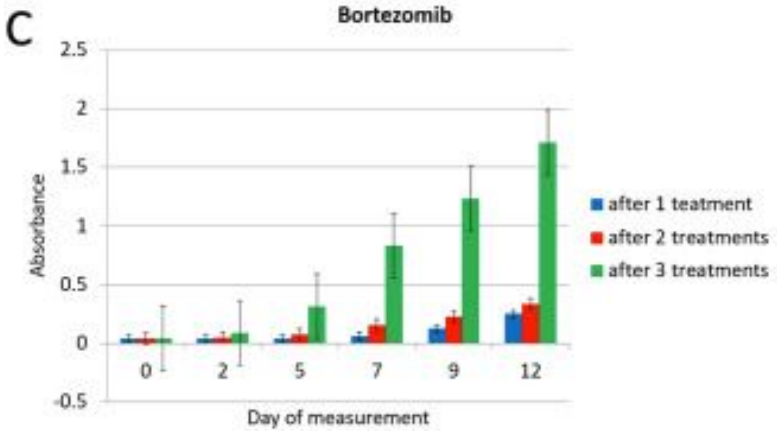
# Motivation

## Chemotherapy

- Sensitivity to chemotherapy may change over time



# Motivation

## Sensitivity to chemotherapy

- Typically, in vitro efficacy of drug combinations is measured by estimating cell viabilities of cell cultures undergoing treatment
  - they are compared to controls with no treatment
  - estimated viabilities are then used to calculate drug synergies
- Measurement is invasive
  - involves adding agent to the cultures
  - inclusion of agent may affect the behaviour of cells
  - repeated measurements are cost and work ineffective

# Objective

## Design a machine learning based method of estimating cell culture viabilities

- Estimation based on photos under the microscope
- Photos taken at macro (whole plate) scale
- Much faster and cheaper than traditional measurement
- Project in cooperation with VUMC Cancer Center Amsterdam

GDAŃSKI UNIWERSYTET MEDYCZNY

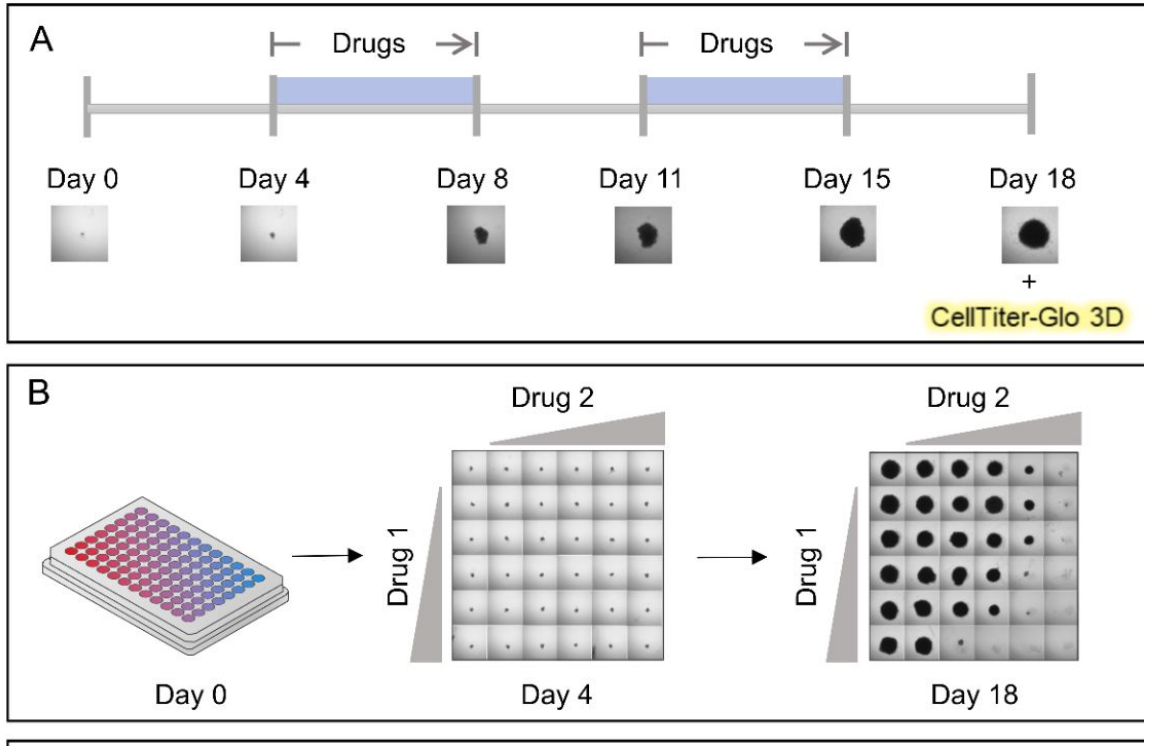
UCZELNIA BADAWCZA
INICJATYWA DOSKONAŁOŚCI

GDAŃSK UNIVERSITY OF TECHNOLOGY

# Data

## GBM in vitro drug synergy dataset

- Data obtained during study on drug synergies in GBM
- Different drug concentrations were used in each well



# Data

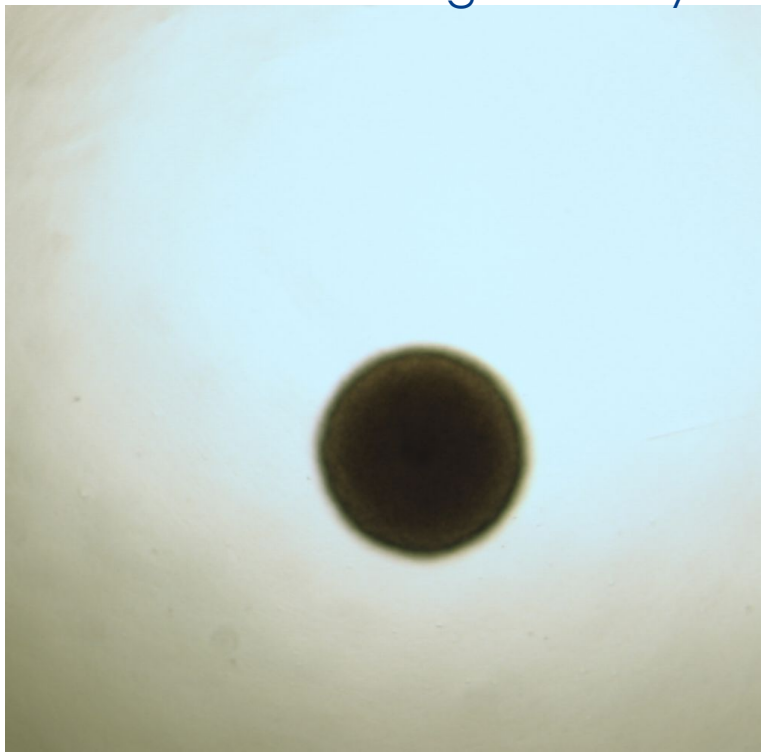
## GBM in vitro drug synergy dataset

- 3 cell lines were selected
- 16 drug combinations
- Two repetitions for each plate
- Photos taken at days:
  - 4 (start of 1st treatment)
  - 8 (end of 1st treatment)
  - 11 (start of 2nd treatment)
  - 14 (end of 2nd treatment)
  - 18 (reference cell viability measurement)

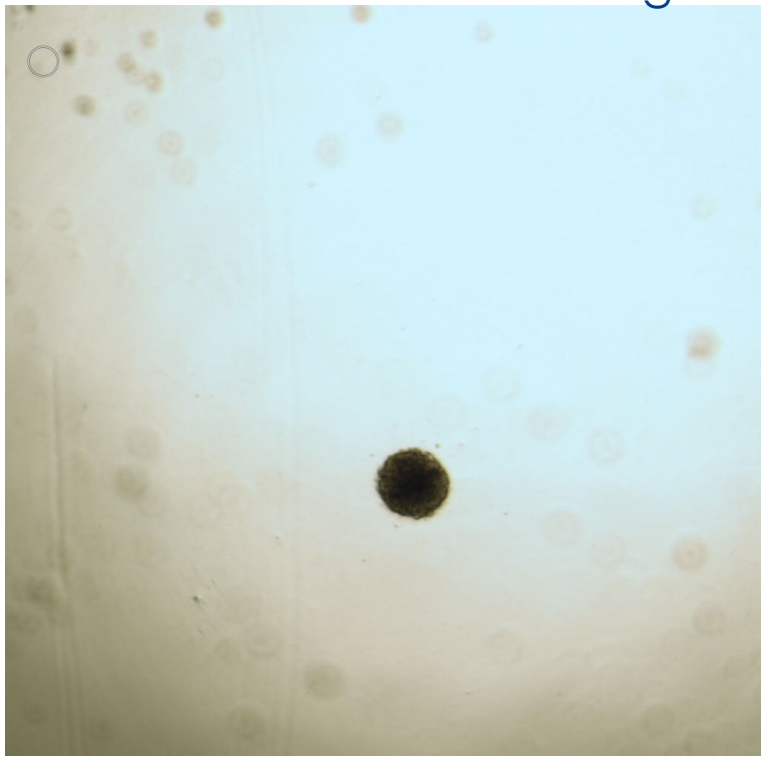
# Data

## GBM in vitro drug synergy dataset

Cell culture with high viability



Cell culture with inhibited growth



# Data

## GBM in vitro drug synergy dataset

- Real cell viabilities were available only for the day 18th
- Measurements or photographs for some plates were missing
- For cell viability estimation, missing data was excluded
- For drug synergy calculations, mean of two repetitions was used
  - ➢ if data for one repetition was missing, only the second one was used

# Data

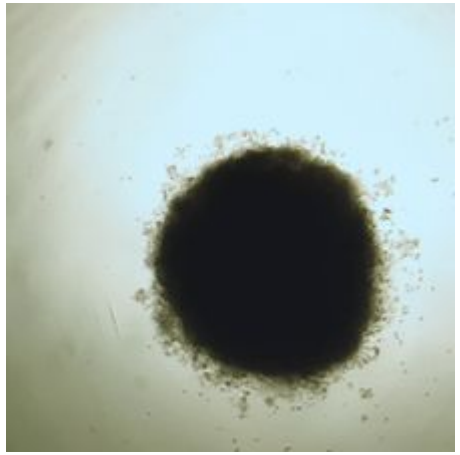
## GBM in vitro drug synergy dataset

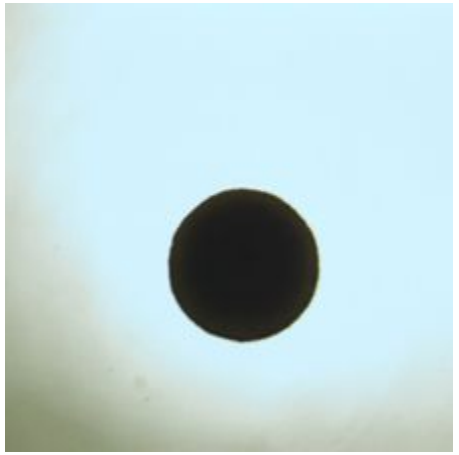
- Only day 18 data was used for training
- Each cell line was included in the training set
- 50% of drug combinations were selected for training purposes
  - ➢ remaining 50% of cells was used as the test dataset
- Either both repetitions were included in the training set or in the test set

# Methods and results

## Prediction of cell viabilities on day 18th

- Light conditions were different in some photos
- Artifacts (e.g. dust) were clearly visible in the background



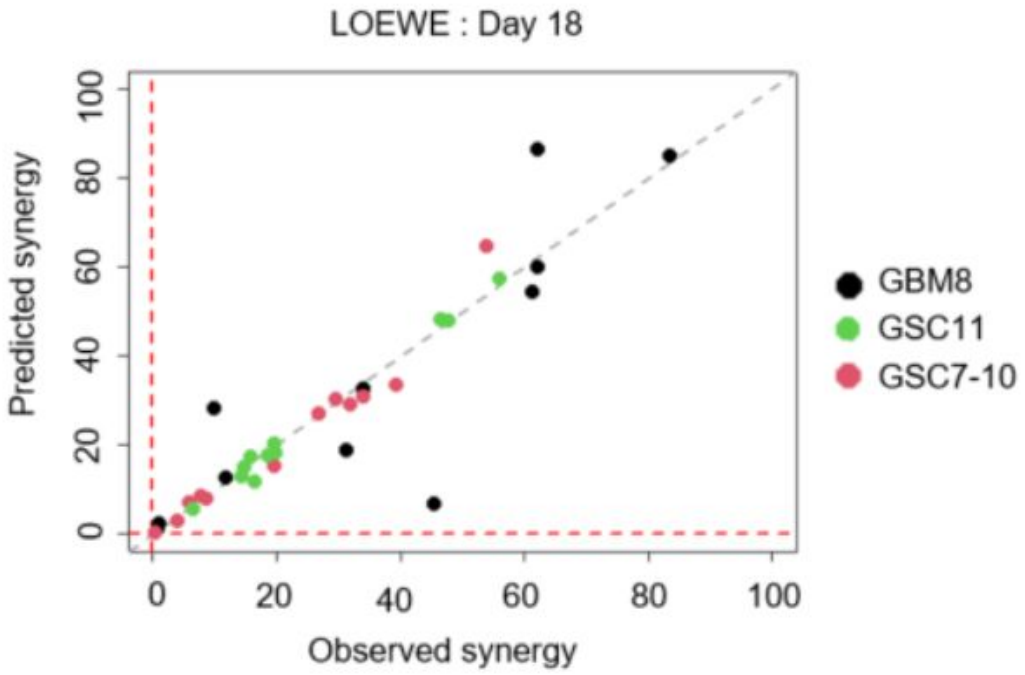


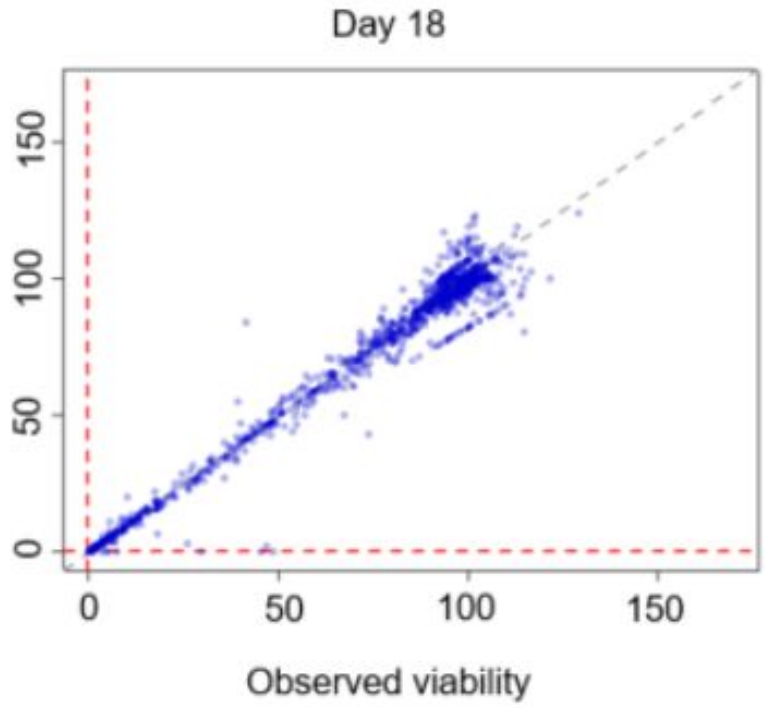
- After preprocessing, pictures were fed into CNN

# Methods and results

## Prediction of cell viabilities on day 18th

- Prediction on day 18 was relatively easy…



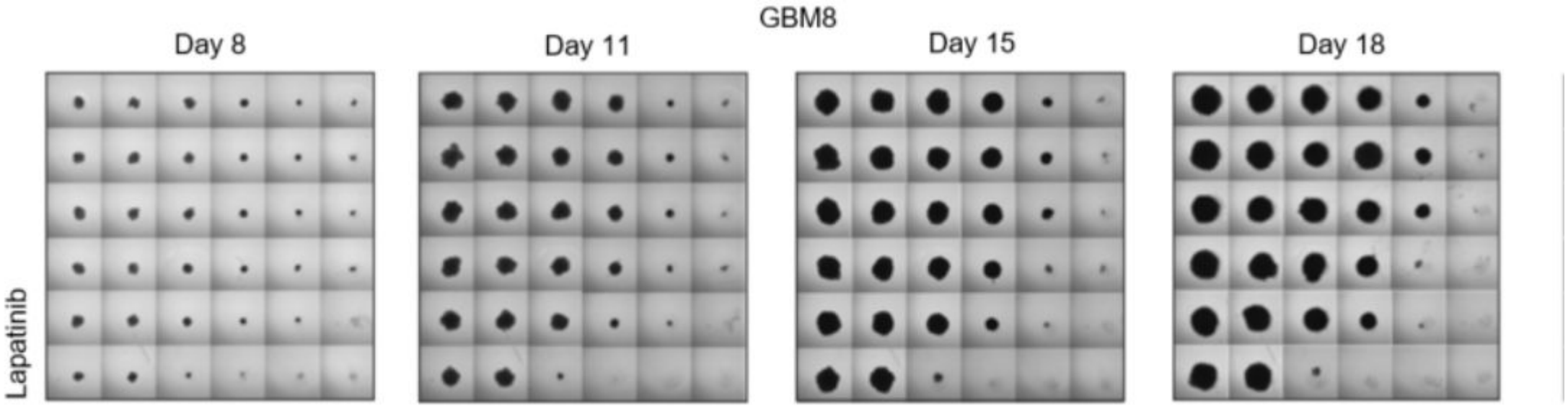




- … but previous days were more challenging

# Methods and results

## Prediction of cell viabilities on previous days

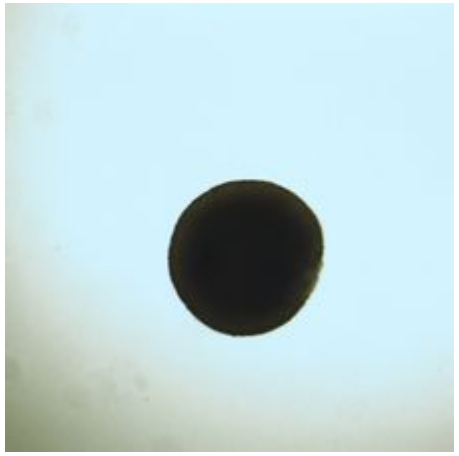
- Sizes of the cell cultures varied depending on day of measurement

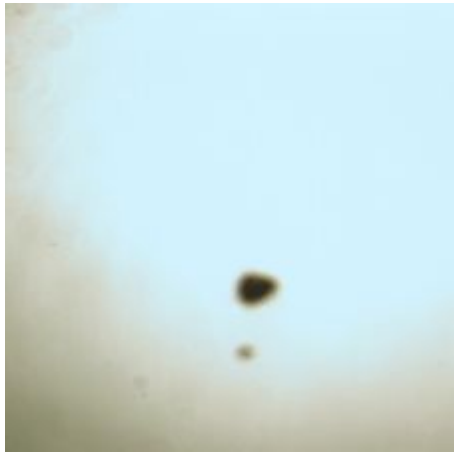


# Methods and results

## Prediction of cell viabilities on previous days

- For each plate, there was 1 reference well with no drugs added





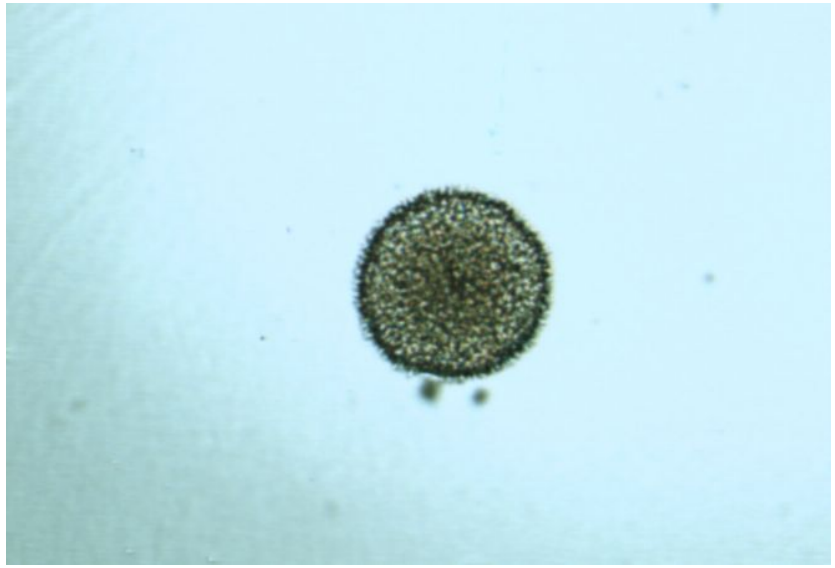
- We solved the problem by predicting from image pairs:
  - culture undergoing treatment + reference culture for each prediction
  - data augmentation was used to simulate different sizes of the cultures

# Methods and results

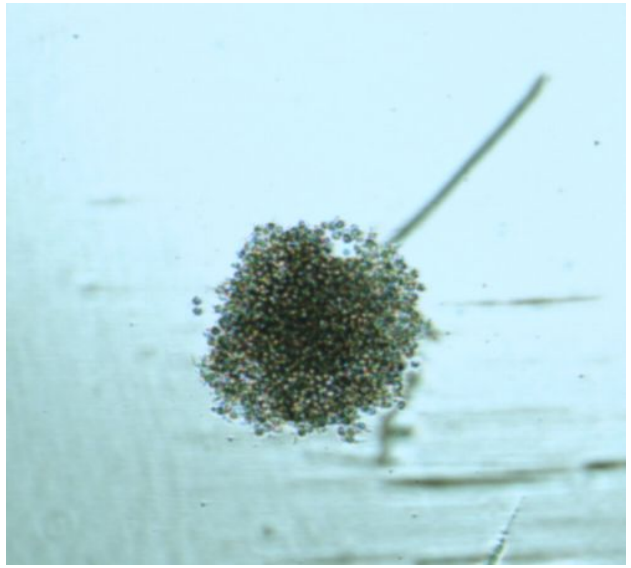
## Prediction of cell viabilities on previous days

- Sometimes, difference between high viability well and low viability well was quite subtle…

High viability well



Low viability well

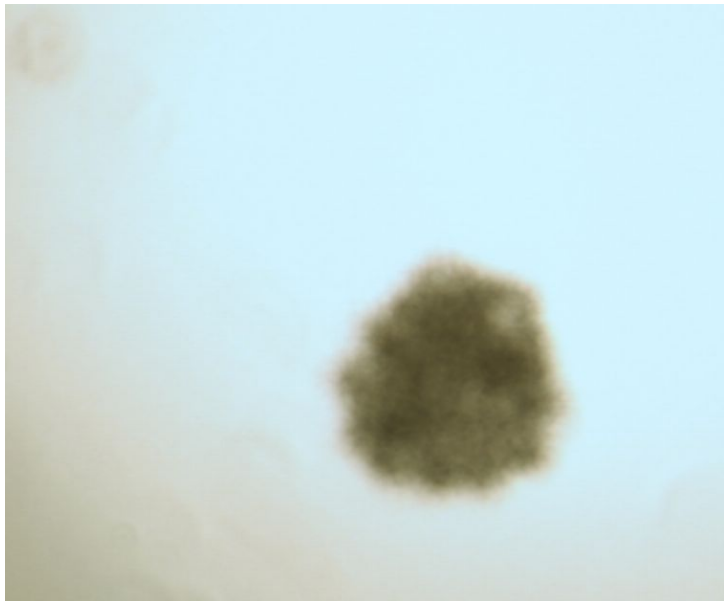


# Methods and results

## Prediction of cell viabilities on previous days

- Sometimes, cells that just died (viability close to 0) were still visible

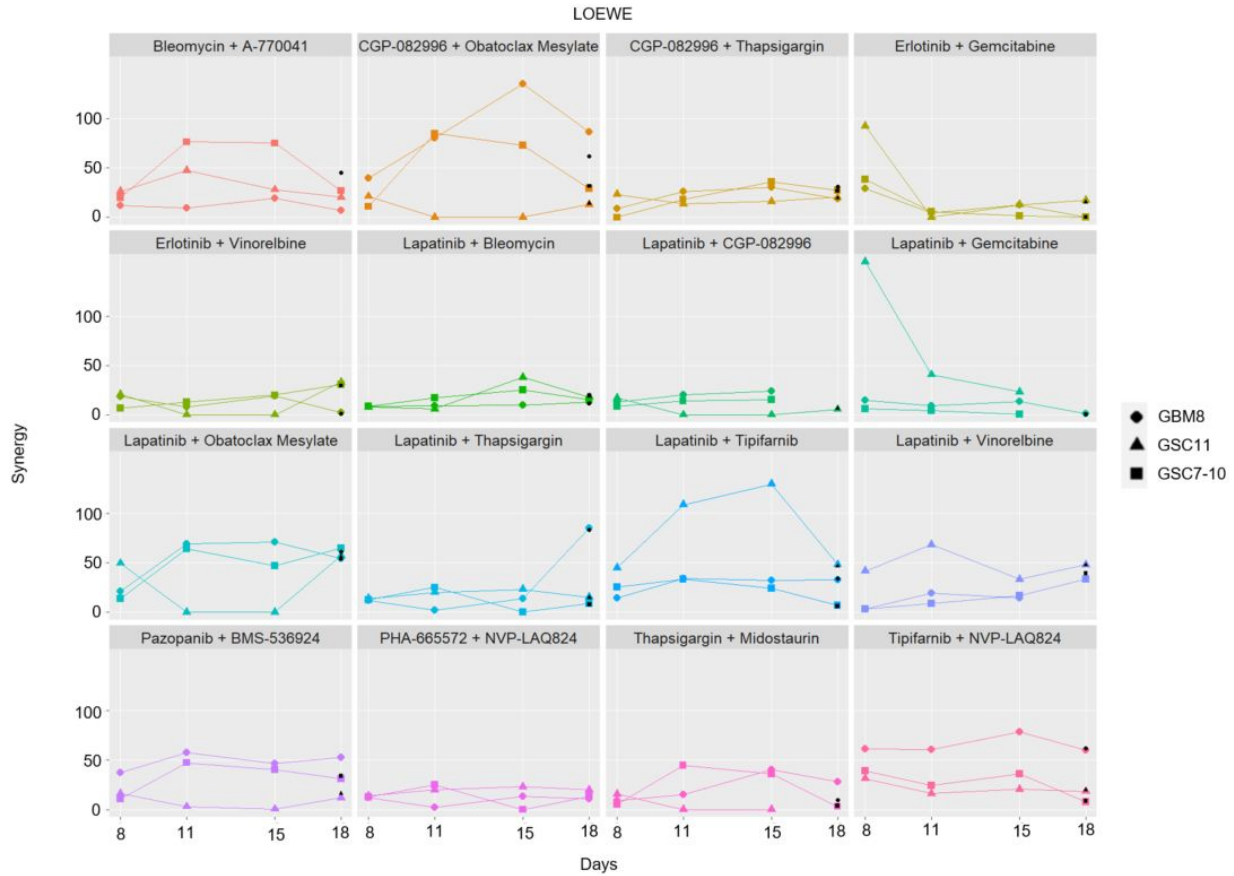
  Well with <1% viability



# Methods and results

## Prediction of cell viabilities on previous days

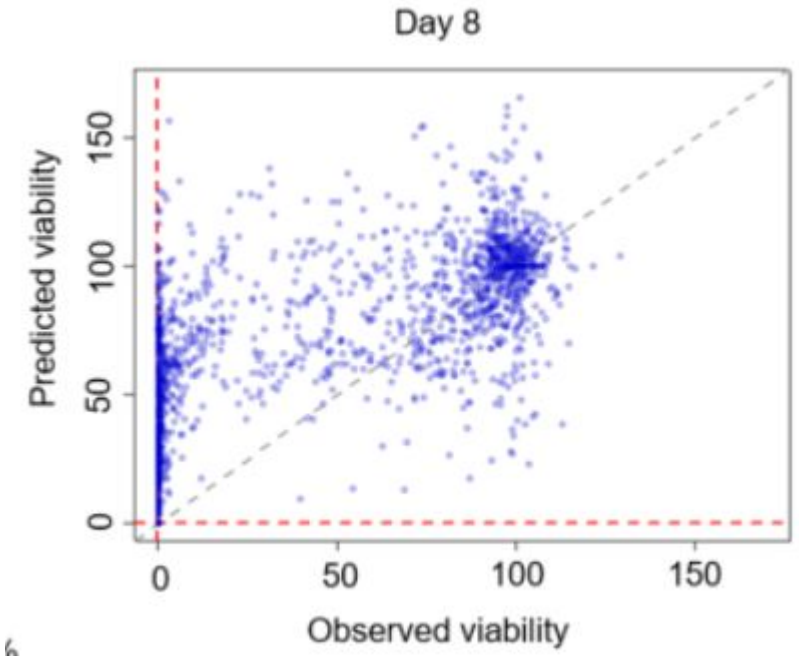
- Synergy estimates over time were quite consistent



# Methods and results

## Prediction of cell viabilities on previous days

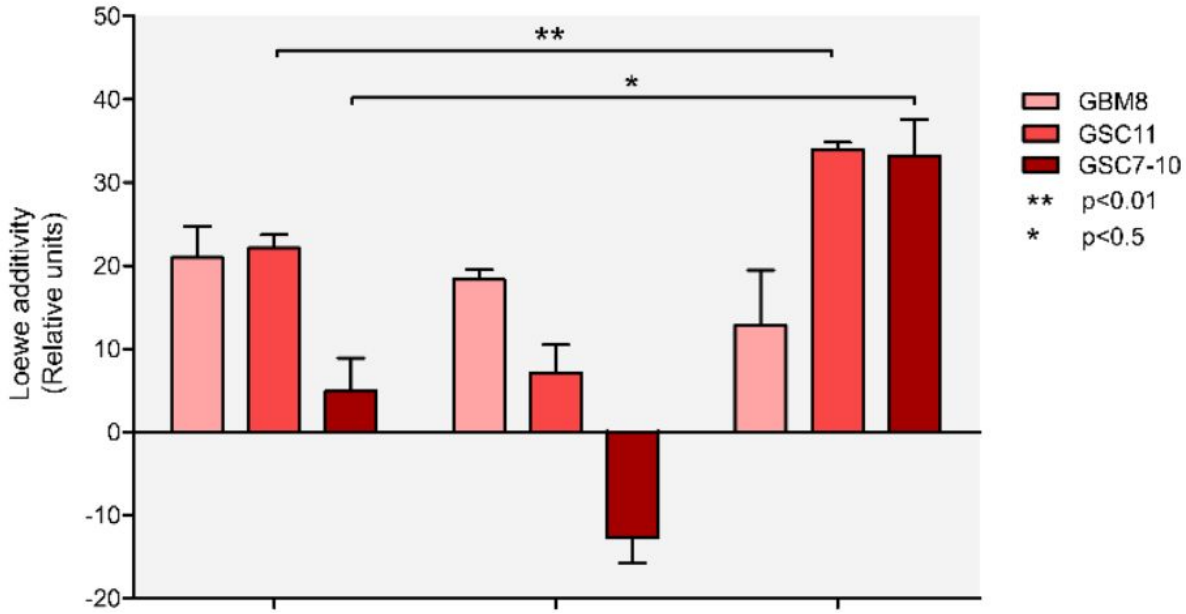
- Cell viability estimate were more error prone
- For comparison purposes, day 18 cell viabilities were used as the reference
- Viabilities were normalized against the viability of reference well (100%)



# Addendum

## Drug synergy study

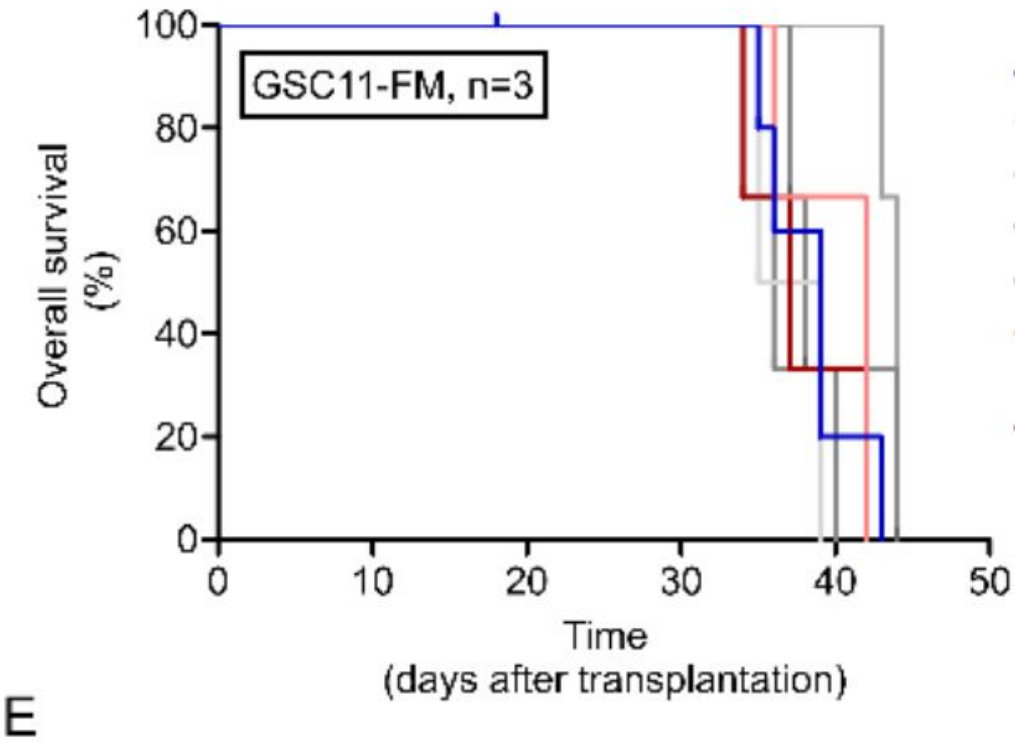
- 15 out of 43 combinations selected for further testing
- Promising 3-drug combination found



# Addendum

## Drug synergy study

- Improvements in mice survival were minor




E