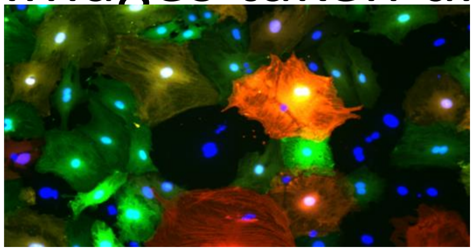
Machine Learning & High-Content Screening for Drug Development/ Repurposing

Sašo Džeroski
Jozef Stefan Institute, Ljubljana, Slovenia
Team: Dragi Kocev, Panče Panov, Tomaž Stepišnik, Martin Breskvar

Interreg
UNIONE EUROPEA
EVROPSKA UNIJA
ITALIA-SLOVENIJA
TRAIN
Progetto standard co-finanziato dal Fondo europeo di sviluppo regionale
Standardni projekt sofinancira Evropski sklad za regionalni razvoj

High-content screening

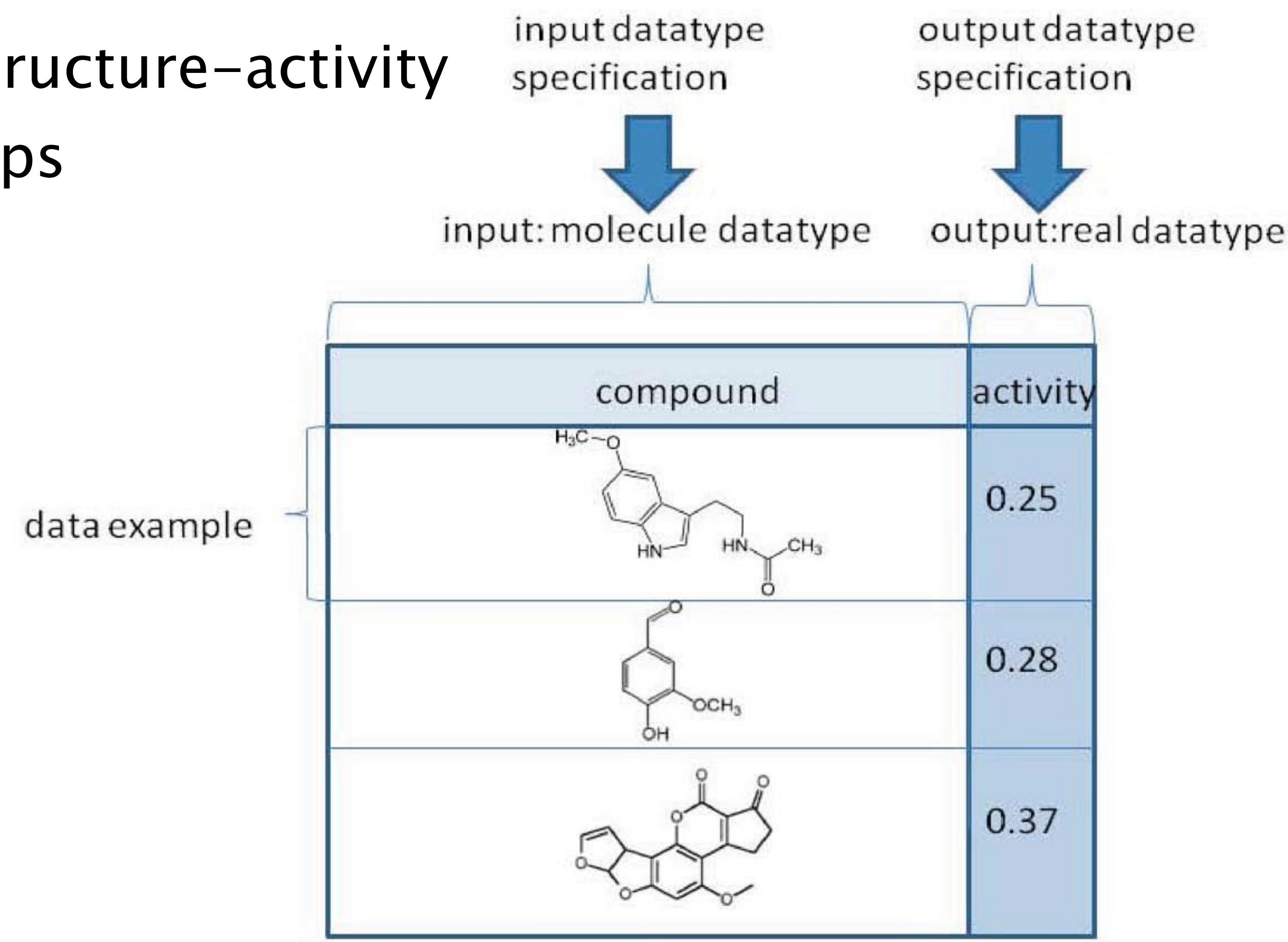
- Compound screens
- Genomic screens (miRNA screens)
- Images taken under the microscope



- Use machine learning to learn models for virtual compound screening, identify novel candidates

Learning from Compound Screens

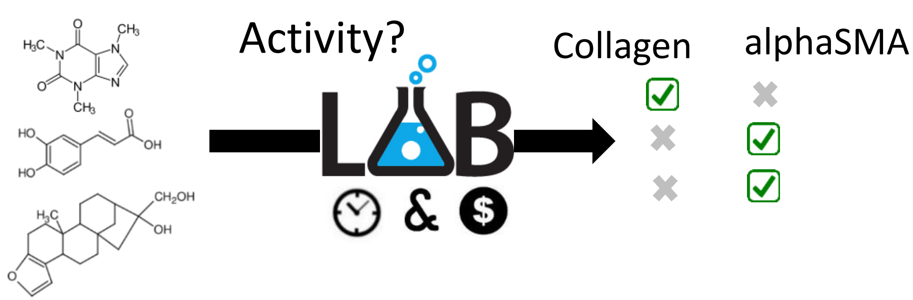
Learning structure–activity Relationships



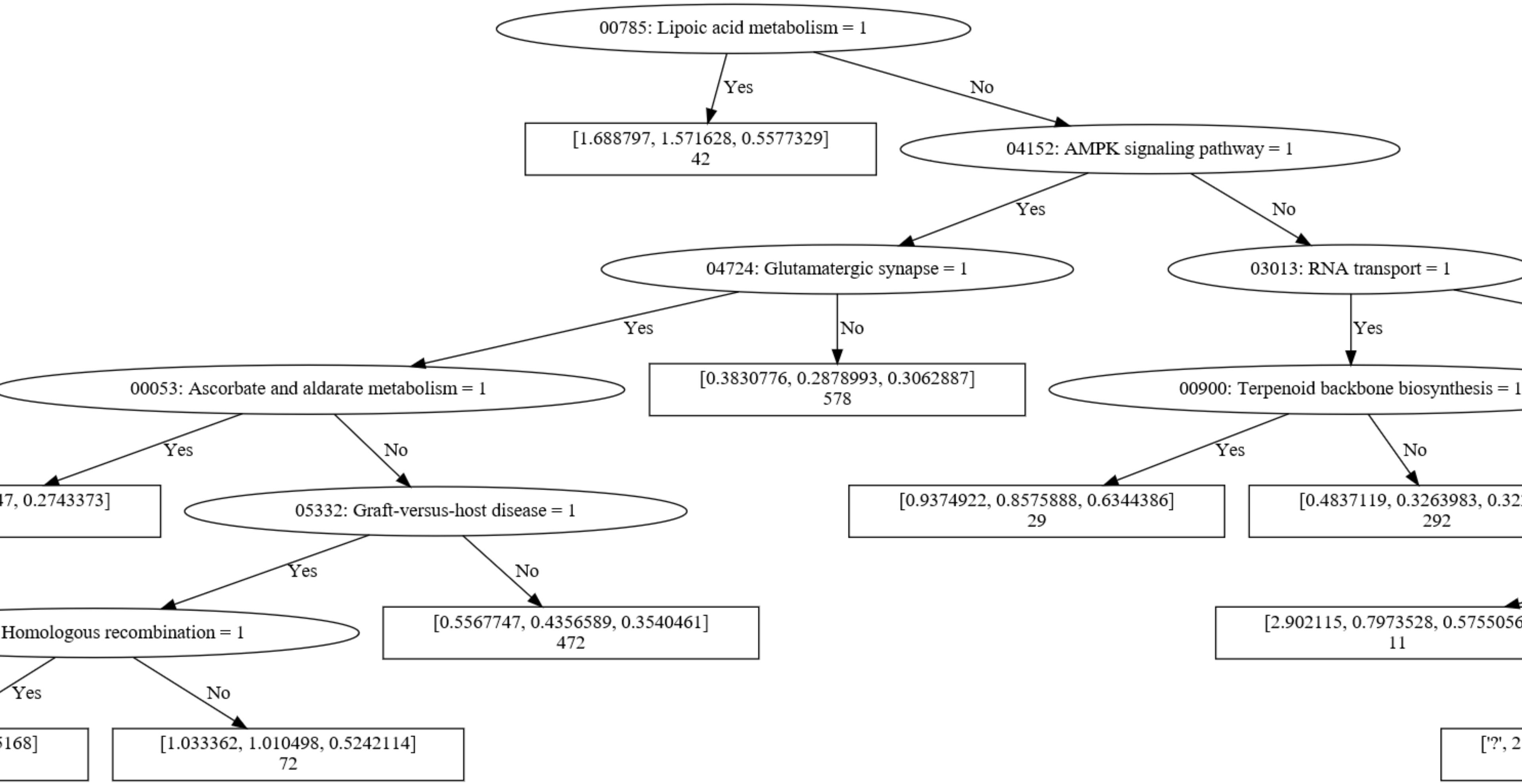
Learning from compound screens

Compound screen data

	Descriptive space				Target space	
Example 1	1	TRUE	0.49	0.69	0.68	3.91
Example 2	2	FALSE	0.08	0.07	0.56	7.59
Example 3	1	FALSE	0.08	0.07	0.10	7.57
Example 4	2	TRUE	0.49	0.69	0.08	8.86
...	...				...	...



The ML technology: Multi-target regression trees



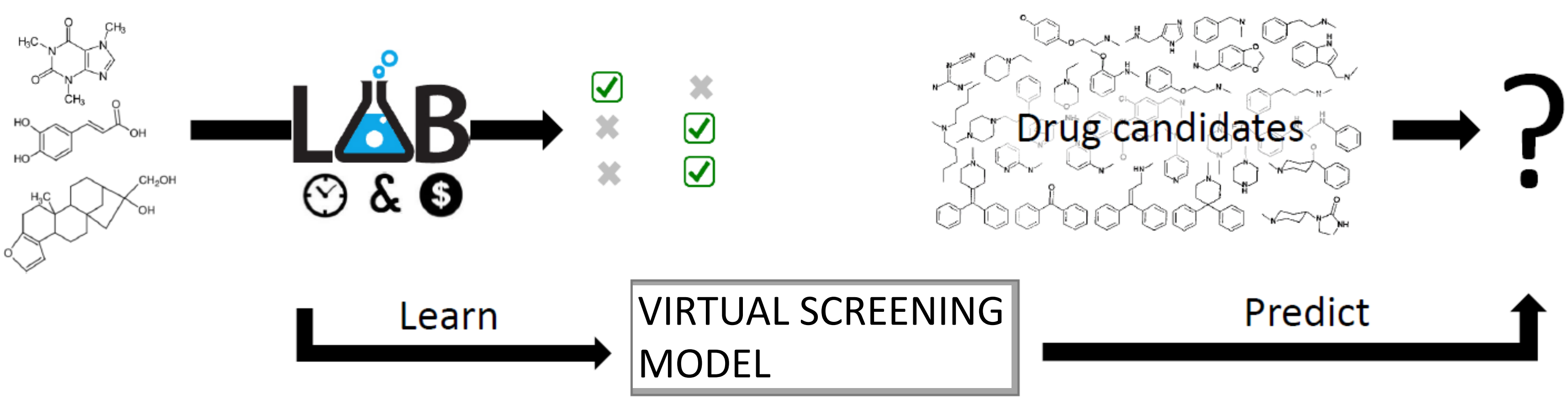
Virtual screening

Labeled data

	Descriptive space				Target space	
Example 1	1	TRUE	0.49	0.69	0.68	3.91
Example 2	2	FALSE	0.08	0.07	0.56	7.59
Example 3	1	FALSE	0.08	0.07	0.10	7.57
Example 4	2	TRUE	0.49	0.69	0.08	8.86
...	...				...	...

Unlabeled data

Example N+1	1	TRUE	0.86	0.35	?	?
Example N+2	2	FALSE	0.09	0.05	?	?
Example N+3	4	FALSE	0.07	0.01	?	?
Example N+4	2	TRUE	0.91	0.78	?	?
Example N+5	2	TRUE	0.42	0.69	?	?
...	...				...	...



Analyzing data from High-contents Screens

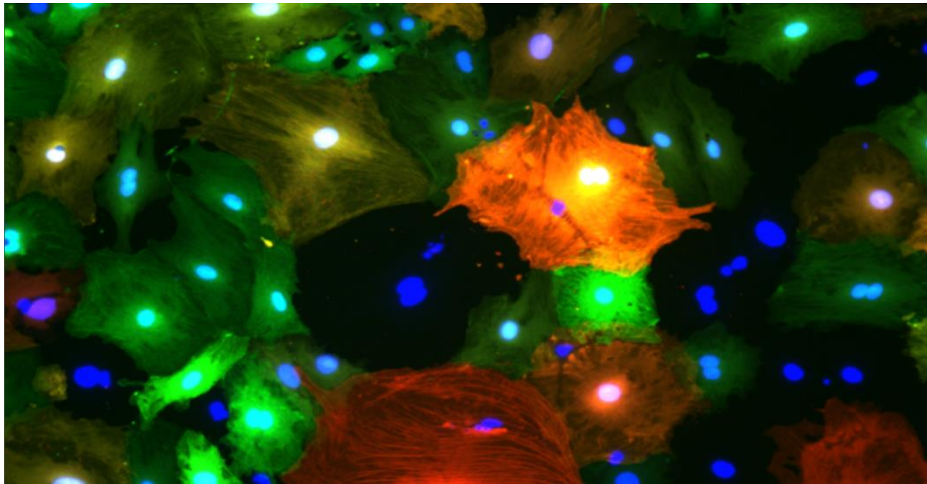
- Compounds described by fingerprints
- Generated by open-source chemoinformatics SW library RDkit
- The FCFP2 fingerprits were used (1024 features)
- Also considered profiles of targeted proteins
- These are the attributes

- Assays photographed under the microscope
- Features extracted from images
- These are then the targets

Reducing fibrosis in myocardial infarction

- High content screen using a library of 640

  FDA approved drugs (ENZO)
- Identify drugs to reduce fibrosis in myocardial infarction
- Screen used murine cardiac fibroblasts which differentiate into myofibroblasts in culture, expressing increased alpha SMA-RFP and collagen-alpha1-EGFP
- Targets: Intensity of
  - alphaSMA
  - Collagen
- Attributes
  - Fingerprints



Testing the predictions

- Some domain-specific knowledge / constraints applied: Predicted compounds filtered for FDA approved drugs that are not corticosteroids
- SMILE strings used in Chemmine to identify substances with structural similarity to non commercial compounds with high predicted values
- Three related compounds identified which are described in literature to have an anti-fibrotic effect
- Four related compounds identified which were not previously described to have an anti-fibrotic effect
- Tested in the wet-lab and one works really well ☺

Experiences within the project

- Analyzed several different screens
  - Compound screens
  - miRNA screens

- Found we can use models learned from miRNA screens to perform virtual compound screening

- Excellent collaboration with ICGEB

- Attractive virtual reality representation