Artificial Intelligence for Medicine and Pharma

Sašo Džeroski
Jozef Stefan Institute (JSI), Ljubljana, Slovenia
Jozef Stefan International Postgraduate School (JS IPS)

Medicine

Medicine is the science and practice of establishing

- diagnosis,
- prognosis,
- treatment, and
- prevention

of disease.

Encompasses a variety of health care practices to maintain and restore health by prevention/treatment.

Health care

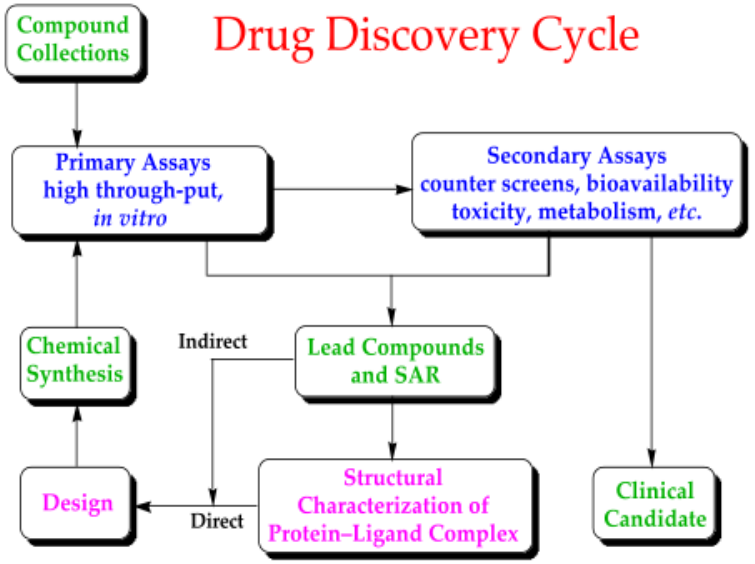
Health care is the maintenance or improvement of health via

- prevention,
- diagnosis,
- treatment,
- recovery, or
- cure

of disease, illness, injury, and other physical and mental impairments in people.

Drug discovery

- Drug discovery is the process by which new candidate medications are discovered.



- Compound screening: Applying compounds from a library (to a cellular assay) to determine their effect
- Genomic screening: Turning on/off of genes, to determine the effect, potential targets for compounds

Drug repositioning/repurposing

- Drug repositioning (aka drug repurposing) involves the investigation of existing drugs for new therapeutic purposes.
- The most famous/successful example of drug repurposing:

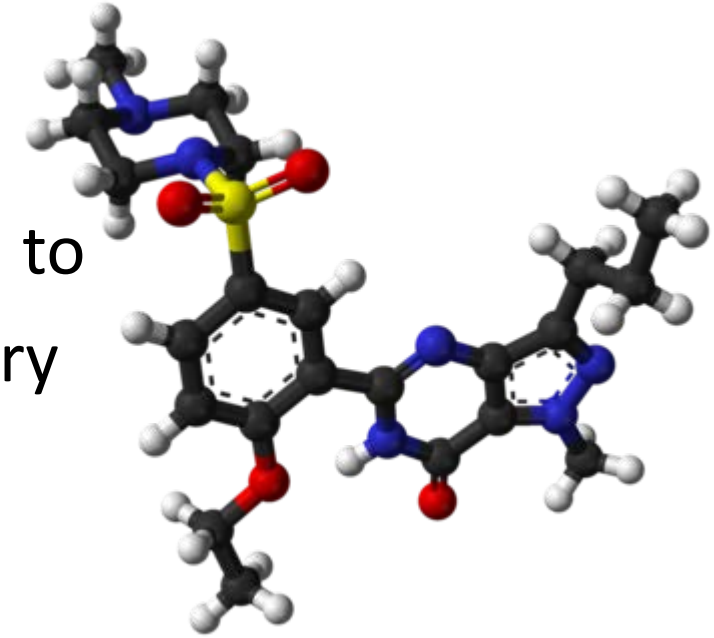
  sildenafil,

  originally used to

  treat pulmonary

  arterial

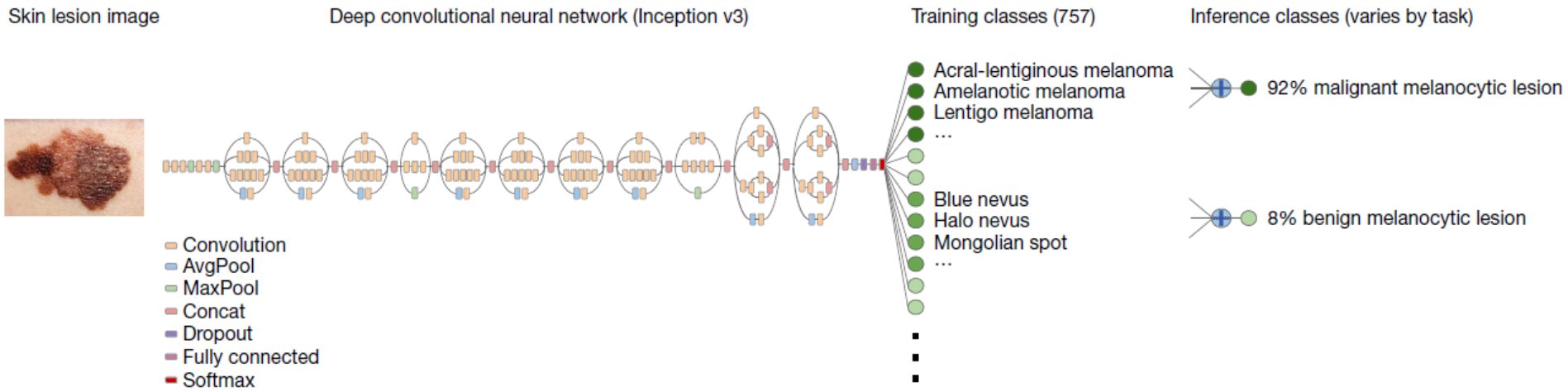
  hypertension



A taxonomy of AI approaches

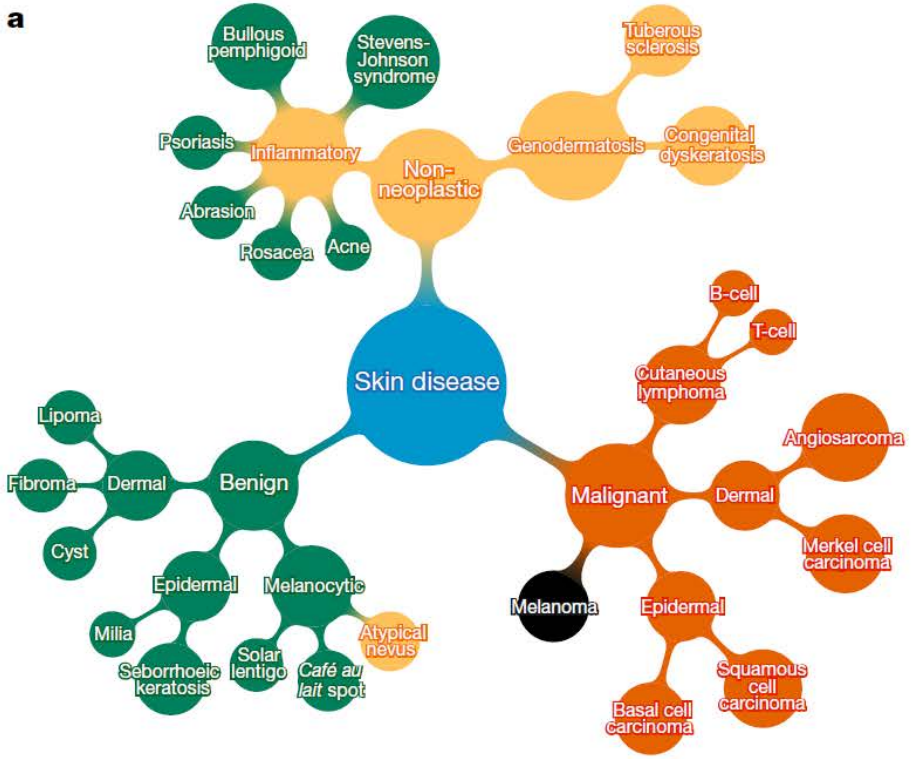
- Knowledge representation
- Knowledge engineering
- Reasoning & Planning
- Natural language processing
- Computer vision
- Machine learning (from big data)
  - Neural networks
    - Deep neural networks (DNNs)
  - Learning understandable/ explainable models
    - Trees & tree ensembles
    - Rules & rule ensembles

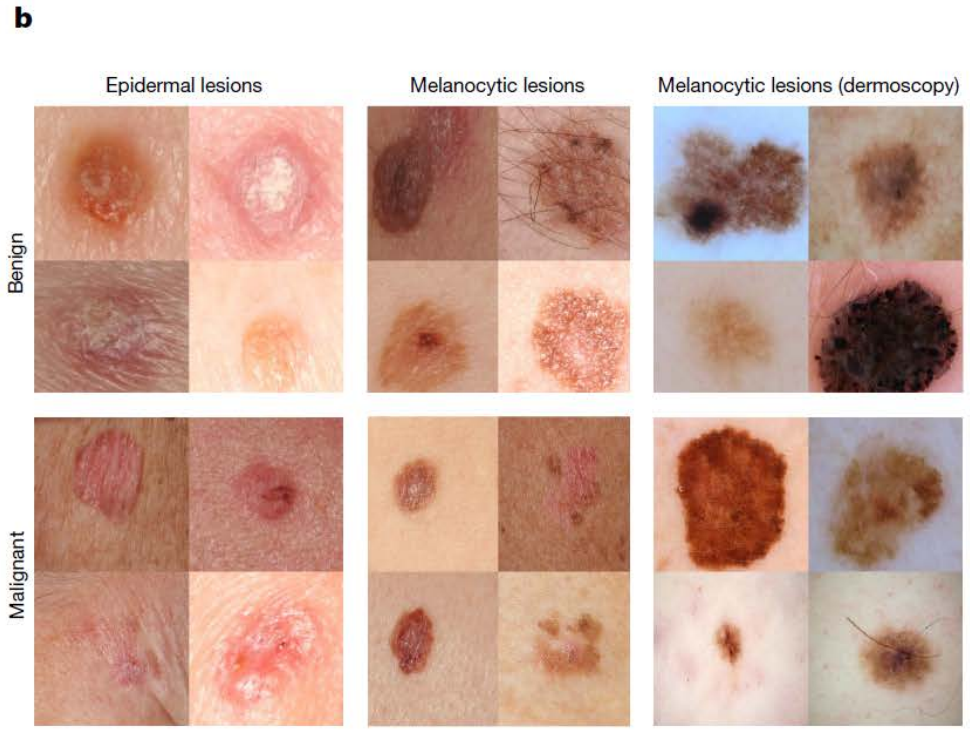
DNNs for Image-based Diagnosis: Classification of skin cancer



- DNN pretrained on ImageNet
- Fine tuned on 129450 images of skin lesions
- 757 training classes defined according to a novel taxonomy of skin disease

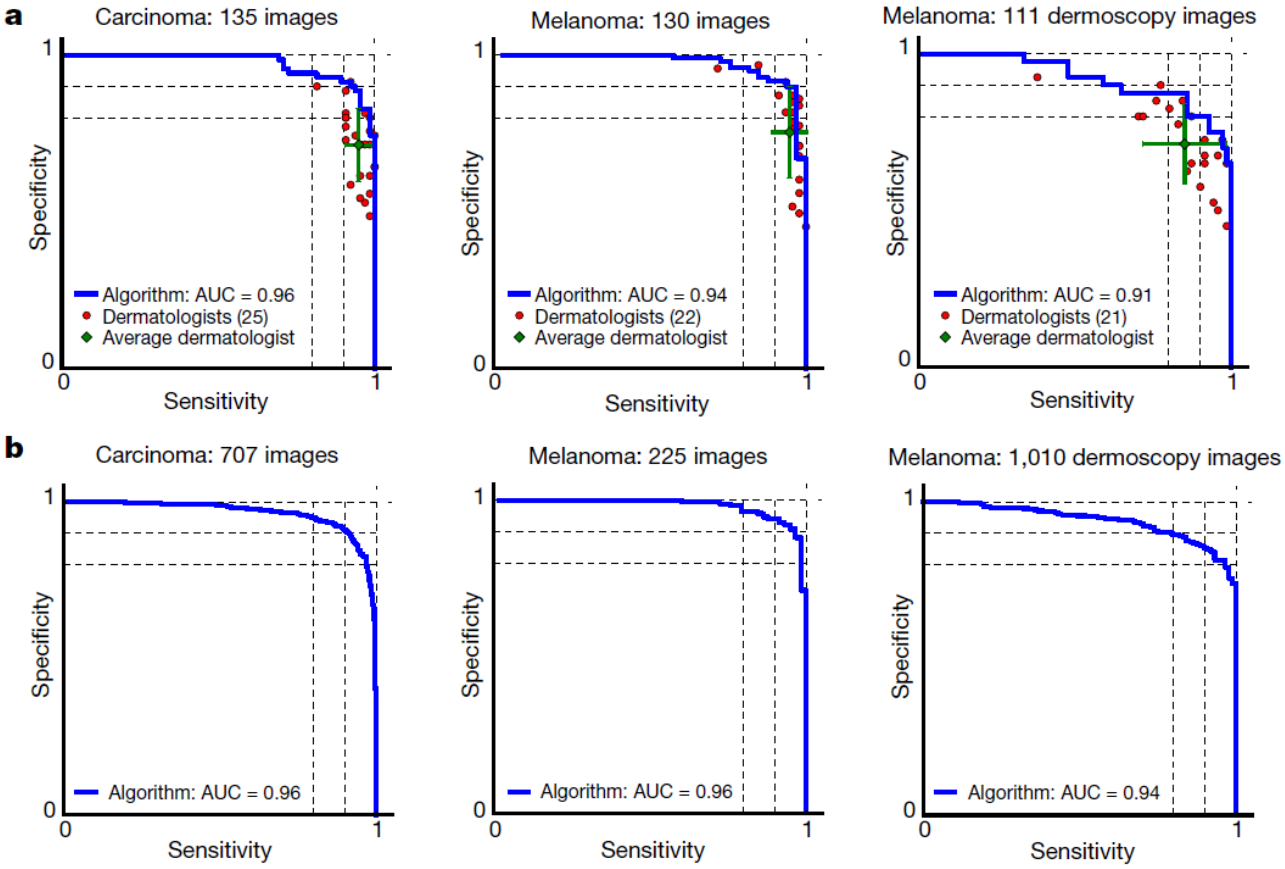
Disease taxonomy & test images





- A subset of the top of the tree-structured taxonomy of skin disease.
- A set of testing images (photos & dermoscopy images)

Performance: Comparison to Dermatologists



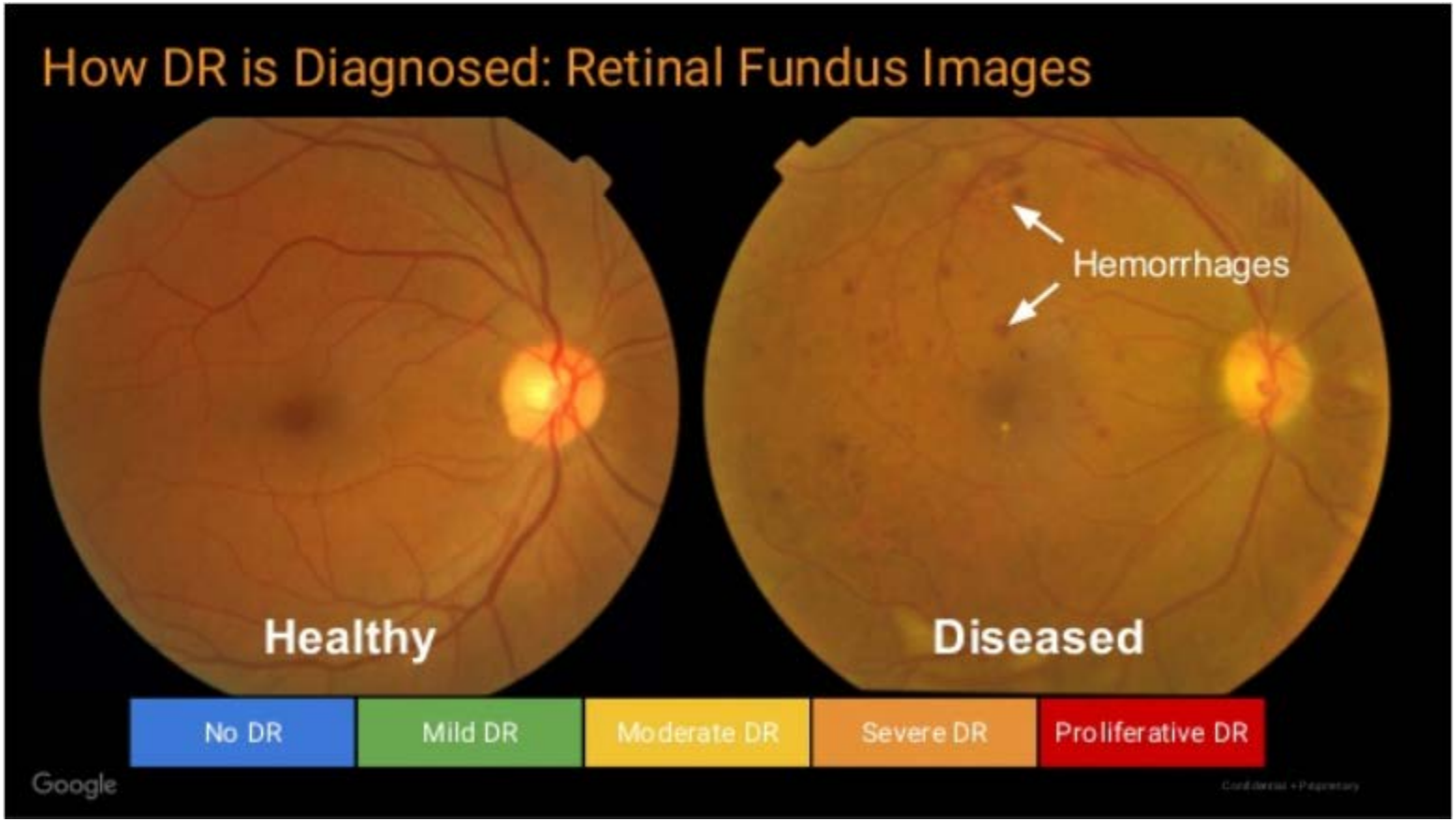
$$\text{sensitivity} = \frac{\text{true positive}}{\text{positive}}$$

$$\text{specificity} = \frac{\text{true negative}}{\text{negative}}$$

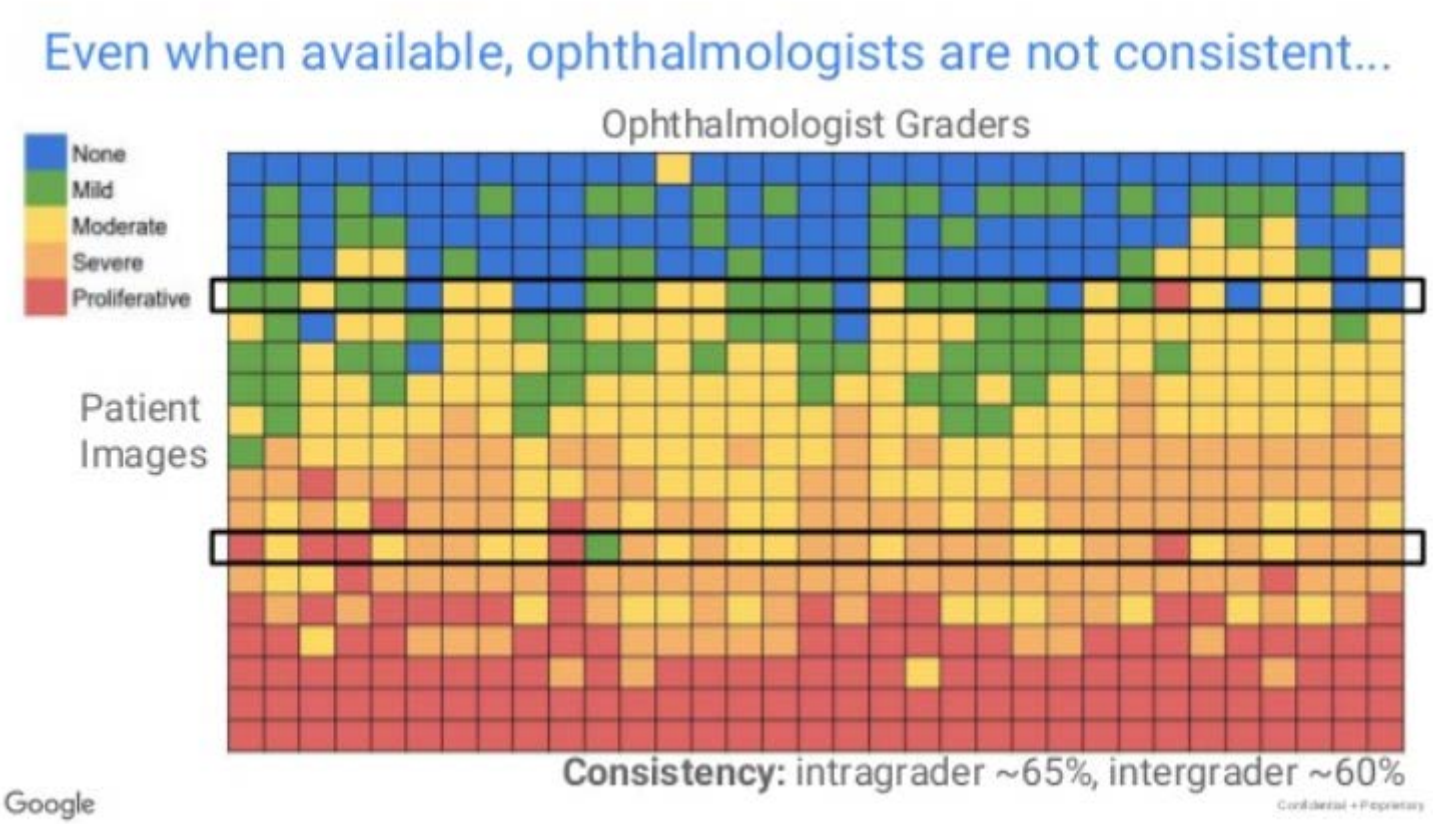
Diabetes causes blindness

- Fastest growing cause of blindness as
- A significant proportion of the population (5-10%) is diabetic
- Should be checked/ screened annually for diabetic retinopathy
-
- There is shortage of personnel to check/ grade images
- Grading is highly technical

Diagnosing Diabetic Retinopathy via Retinal Fundus Images

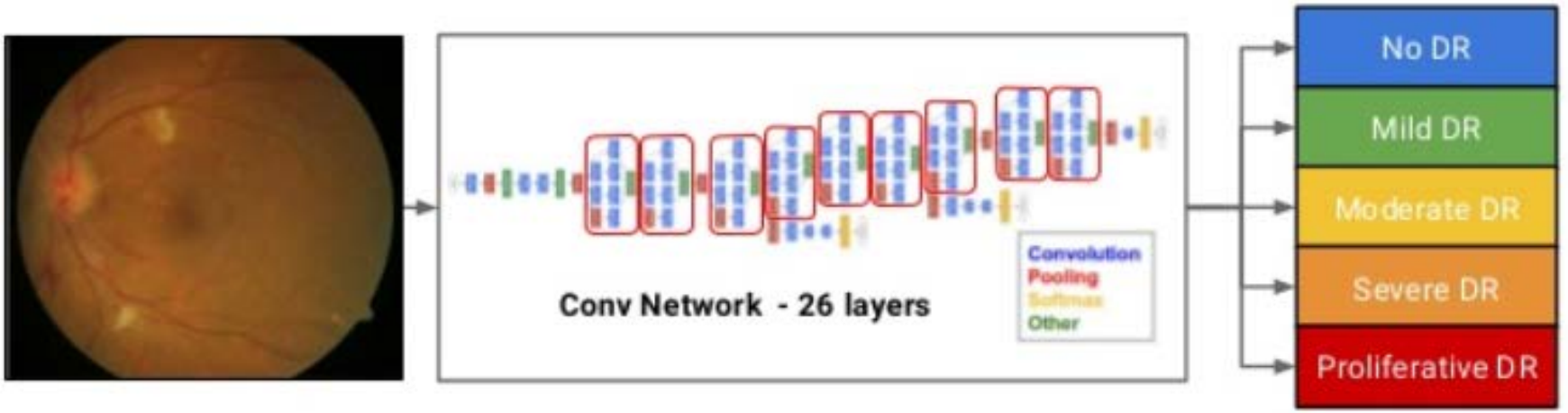


Additional problems with grading



Training a DNN for diagnosing DR

Adapt deep neural network to read fundus images

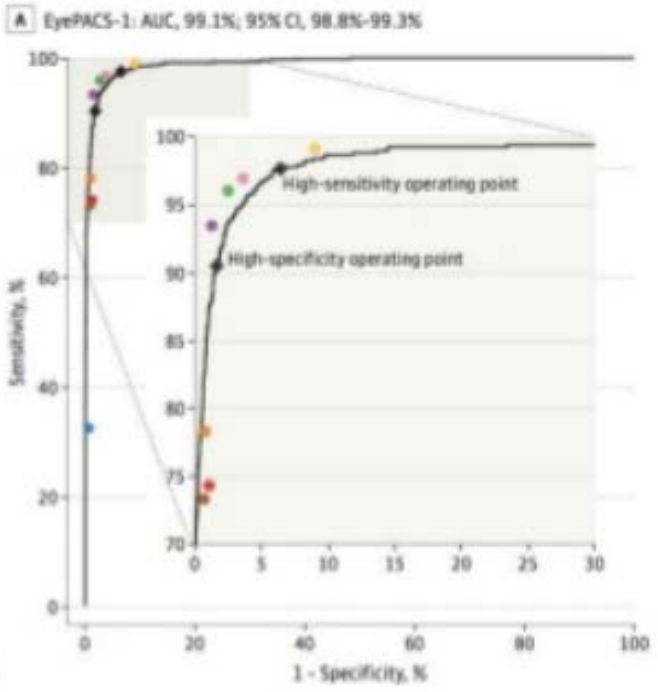


Labeling tool
54 ophthalmologists
130k images

Diagnostic performance as compared to ophtalmologists

JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs



F-score

0.95	0.91
Algorithm	Ophthalmologist (median)

"The study by Gulshan and colleagues **truly represents the brave new world in medicine.**"

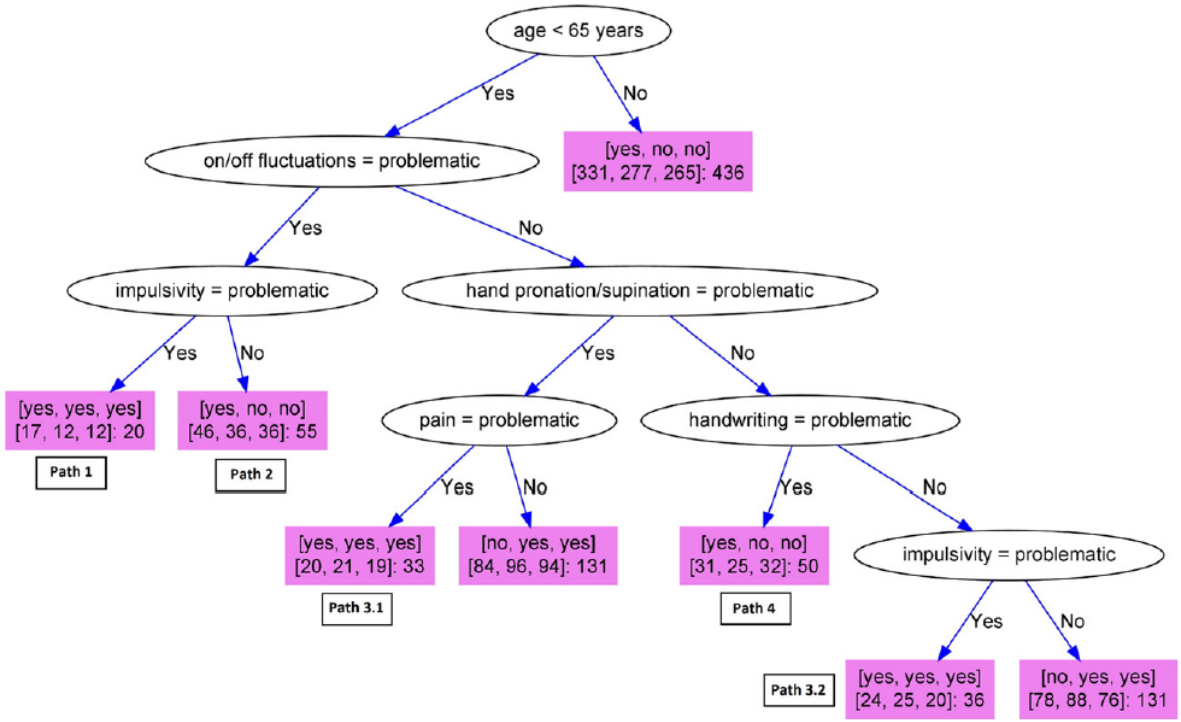
Dr. Andrew Beam, Dr. Isaac Kohane
Harvard Medical School

"Google just published this paper in JAMA (impact factor 37) [...] **It actually lives up to the hype.**"

Dr. Luke Oakden-Rayner
University of Adelaide

Understandable ML for therapy

- Indicating change of therapy for Parkinson's patients
- From patient's symptoms at visit, predict whether physician will change each of three groups of meds: Levodopa, dopamine agonists, MAO-B inhibitors



Drug discovery/repurposing

- Perform compound screening with a relatively small compound library to collect data
- From the collected data, learn a predictive QSAR model that relates compound structure to activity

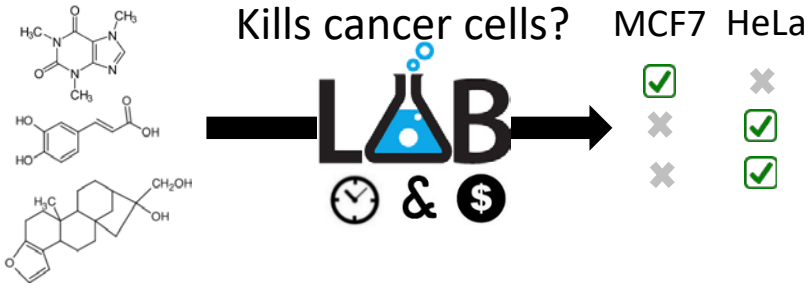
- Apply the learned model to perform virtual compound screening on a large set of compounds
- Find candidate compounds for new drugs

Real compound screening: Collecting data

- Testing compounds from libraries on cellular assays

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
...	...				...	...



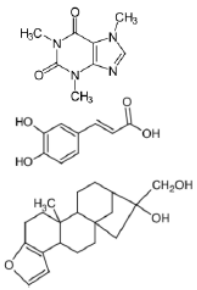
Learning QSAR models for Virtual Compound Screening

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	?	?
...	...				...	...



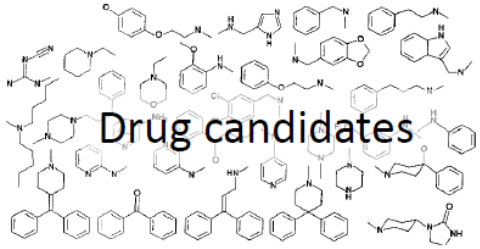
Kills cancer cells?

MCF7 HeLa

✓ ✗

✗ ✓

✗ ✓



?

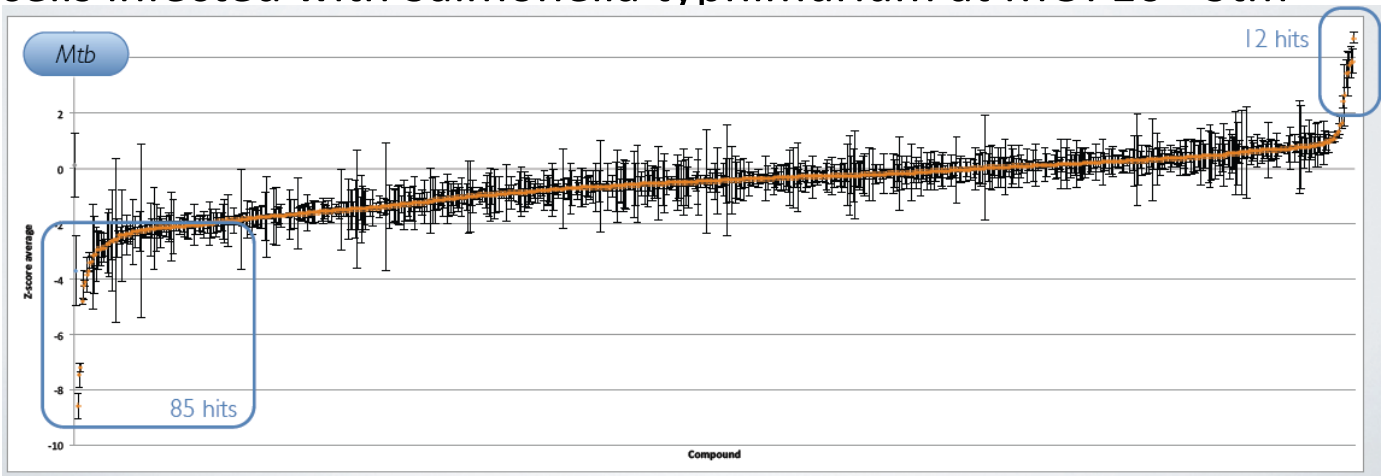
Learn → Supervised model → Predict

Virtual compound screening

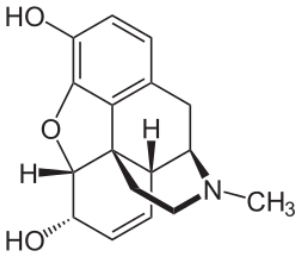
- Descriptive variables refer to compound structure
  - Functional groups
  - Fingerprints
  - Bulk properties
- May also describe the compound in terms of the proteins it targets (e.g. from PubChem)
  - Their functional annotations
  - Pathways they are involved in
  - Proteins that the targets interact with (and/or their functional annotations, pathways they are involved in)
- Target variables describe compound activity and toxicity

Host-targeted Drugs for MTB (Tuberculosis) and STM (Salmonella)

- Library of compounds
  - LOPAC library - Library Of Pharmacologically Active Compounds
    - 1260 compounds
  - Well-characterized compounds, many already applied in clinical practice for a range of conditions
- Flow cytometry (FACS) - measured reduction in bacterial load
  - MelJuSo cells infected with Mycobacterium tuberculosis at MOI 10 – Mtb
  - HeLa cells infected with Salmonella typhimurium at MOI 10 - Stm



MTB&STM: Host-targeted Drugs



Morphine

- Given SDF files, find PubChemID
- PubChem repository
  - Retrieve the proteins that were found to be active in bio-assays with human cells
- Dataset
  - 964 compounds were found active on human protein targets
  - 711 distinct protein targets were identified
- Each compound is described with
  - the respective protein targets
  - functional annotations of the respective protein targets
  - functional annotations of both the respective protein targets and the proteins they interact with

Model excerpts

- MTR models (trees, ensembles) were built that predict the effect of a compound on
  - Bacterial load
  - Host cell
- Example rule from a tree:

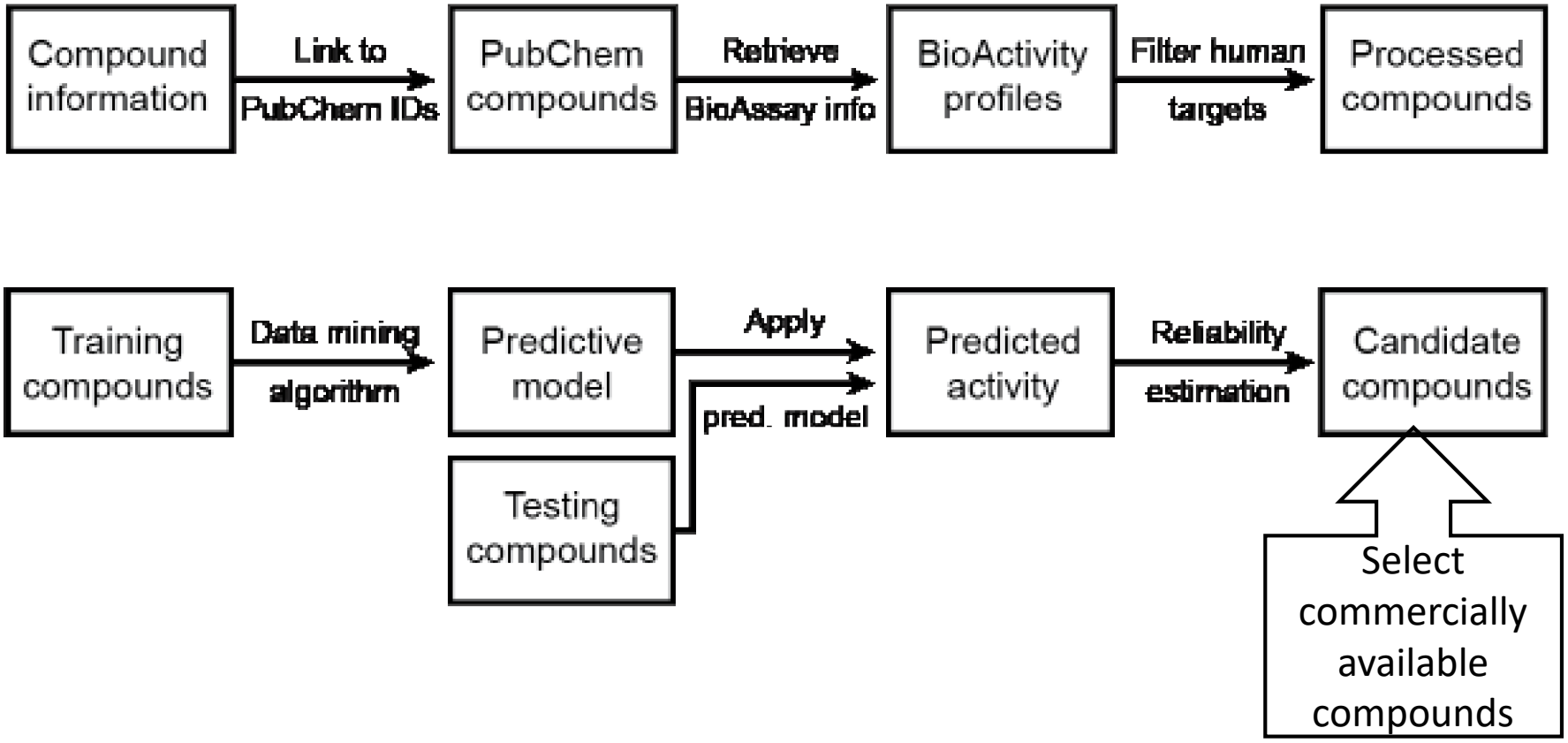
  IF compound targets the protein AAL 06595

  THEN bacterial load effect = -5.269 &

  host cell effect = 0.0475
- Functional profiles of targeted proteins

  IF a protein with function GO0002637 (regulation of immunoglobulin production) is targeted THEN …

MTB&STM: Host-targeted Drugs The Data Analysis Workflow



MTB&STM: Host-targeted Drugs Results

- Greatly increased proportions of hit compounds
  - 5 out of 9 (55.6%) for Mtb and
  - LOPAC primary screen (90 out of 1260 (7.1%) for *Mtb*
- The *in silico* predictive model successfully identified active compounds *de novo*

Abbr.	*Compound name*	*Alternative name(s)*	*Primary sceen z-score*	*Rescreen z-score*	*Activity*
Mycobacterium tuberculosis					
SU	SU 6656	2,3-Dihydro-N,N-dimethyl-2-oxo-3-[(4,5,6,7-tetrahydro-1H-indol-2-yl)methylene]-1H-indole-5-sulfonamide	**-5.79**	**-10.51**	Src family kinase inhibitor
Q	Quinacrine dihydrochloride		**-5.25**	**-9.90**	MAO inhibitor
SB	SB 216763	3-(2,4-Dichlorophenyl)-4-(1-methyl-1H-indol-3-yl)-1H-pyrrole-2,5-dione	**-6.02**	**-8.29**	GSK-3 kinase inhibitor
G	GW5074	3-(3, 5-Dibromo-4-hydroxybenzylidine-5-iodo-1,3-dihydro-indol-2-one)	**-4.86**	**-6.98**	Raf1 kinase inhibitor
T494	Tyrphostin AG 494	N-Phenyl-3,4-dihydroxybenzylidenecyanoacetamide	**-3.83**	**-6.93**	EGFR kinase inhibitor
L	3',4'-Dichlorobenzamil hydrochloride	L-594,881	**-3.87**	-5.13	$Na^+/Ca2^+$ exchanger inhibitor
H	Haloperidol		**-3.77**	-2.96	D2/D1 dopamine receptor antagonist

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

HTS: Modulating fibroblast to myofibroblast transition

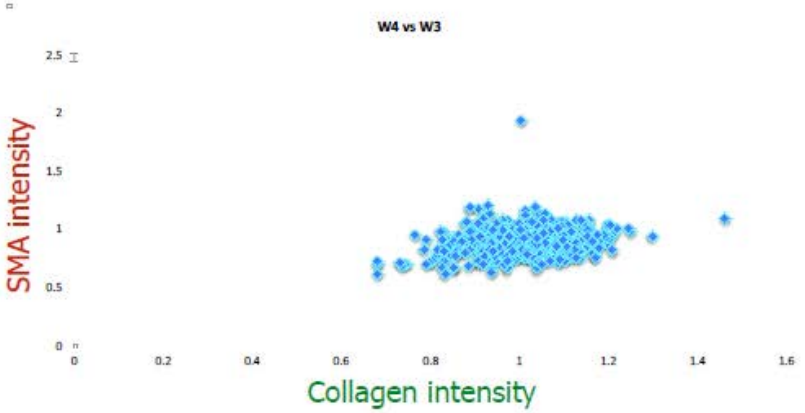
cardiac fibroblasts from a-SMA-RFP/
Coll α1(I)-EGFP mice

0h

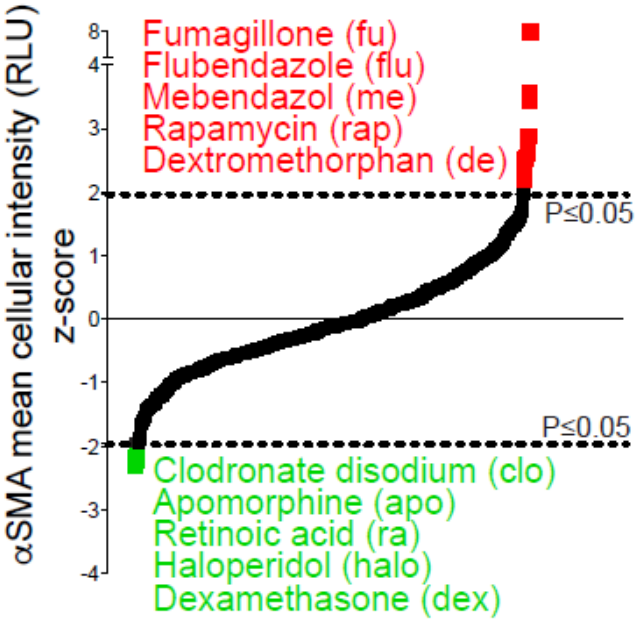
FDA-approved drugs (640 compounds)

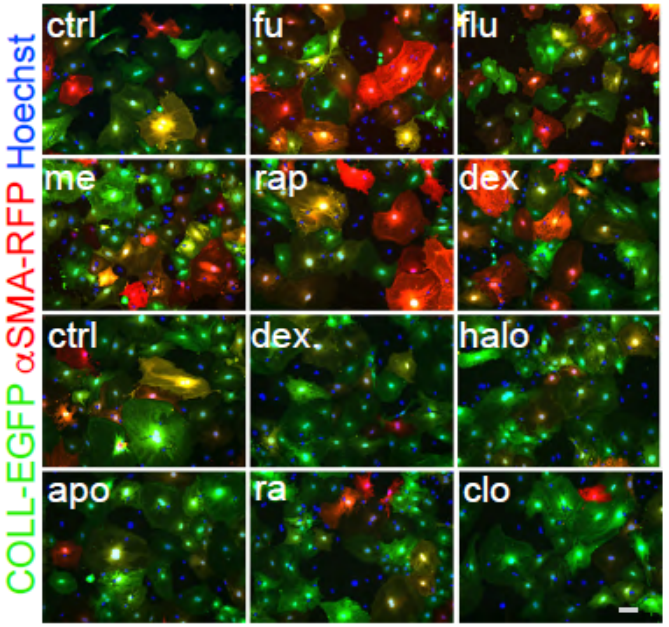
72h

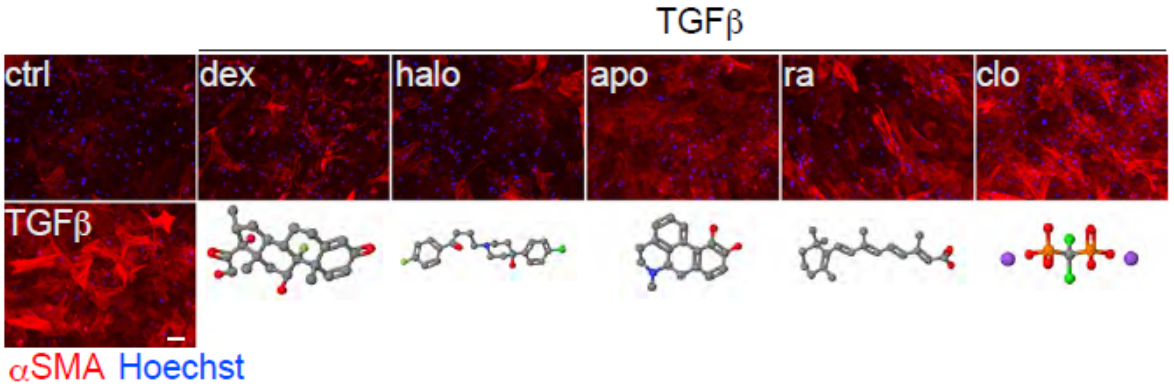
cell fixation, image acquisition and elaboration

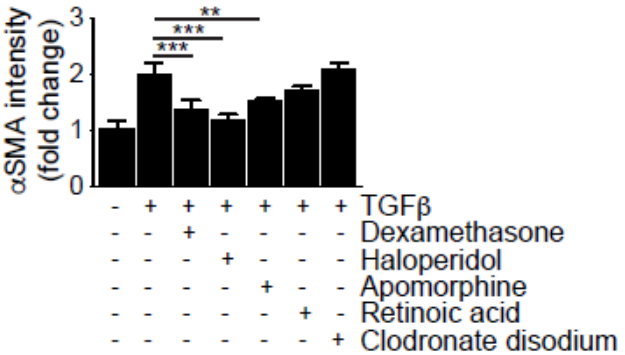


Hits in the HTS screen



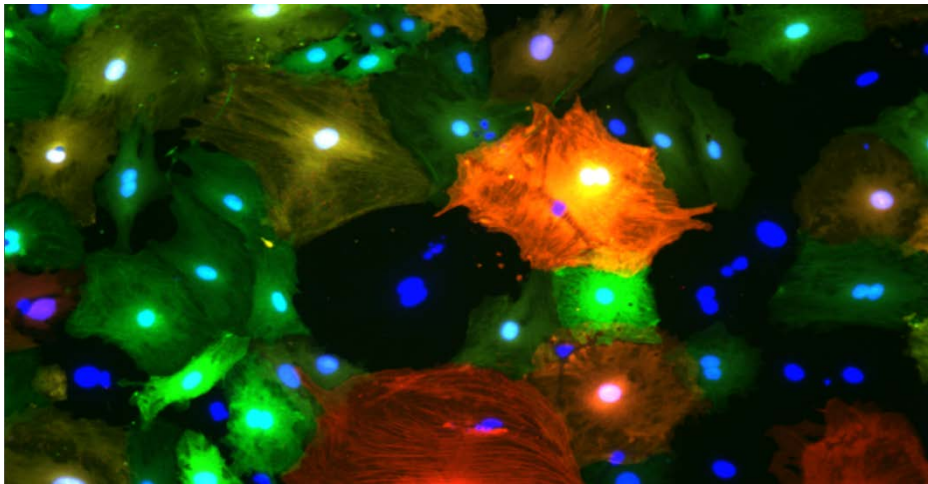






Reducing fibrosis in myocardial infarction

- 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



New candidate drugs to help recovery after heart attack

- 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
  - Melatonin * and Indomethacin *
  - **Acyclovir**
- Four related compounds identified which were not previously described to have an anti-fibrotic effect
  - **Dopamine**
  - Amiodarone * and Progesterone *
  - **Zanamivir**

AI in Medicine and Pharma

- Many different tasks to use AI for, from pharma, medicine, healthcare
- Many different AI methods to use, e.g., also decision support systems to avoid hospital infections

- Important issues unique to uses in medicine/healthcare
  - Explainability
  - Regulating the use of AI in medicine (FDA approvals)