ASCPT

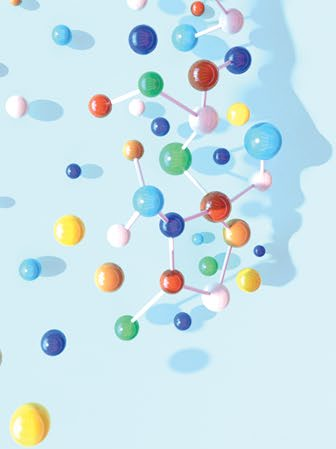
FROM MOLECULE TO PATIENT

ASCPT 2019 ANNUAL MEETING

FROM MOLECULE TO PATIENT

An Integrative Deep Learning Approach for De Novo Drug Discovery

Joel Dudley PhD
Icahn School of Medicine at Mount Sinai



FROM MOLECULE TO PATIENT

Conflicts of Interest Statement

Scientific founder or co-founder

- Onegevity Health
- OOVA
- Ontomics
- NuMedii

Scientific advisory board member

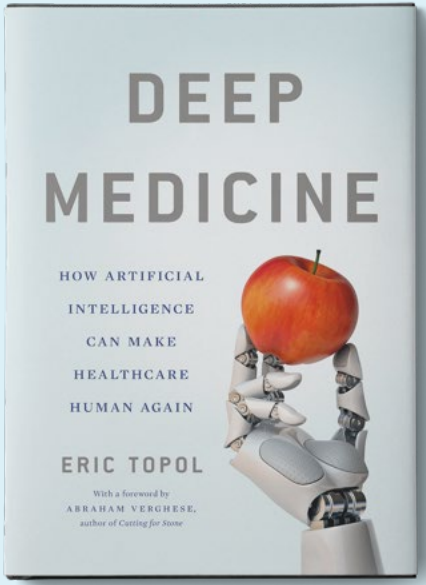
- Ayasdi
- LAM Therapeutics
- Solve Bio
- Hoy Health
- Biotia

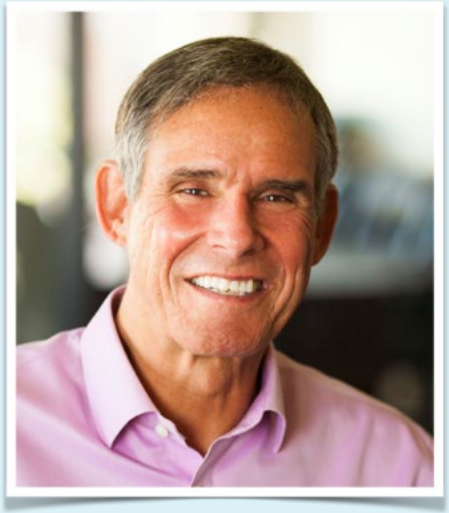
Past or present consultancy

- Janssen Pharmaceuticals
- Thorne Research
- Allergan
- AstraZeneca
- LEO Pharma
- Speaking honoraria
- Celgene
- Illumina
- Roche
- Takeda
- Lundbeck

Deep Learning hits the mainstream in biomedicine

FROM MOLECULE TO PATIENT

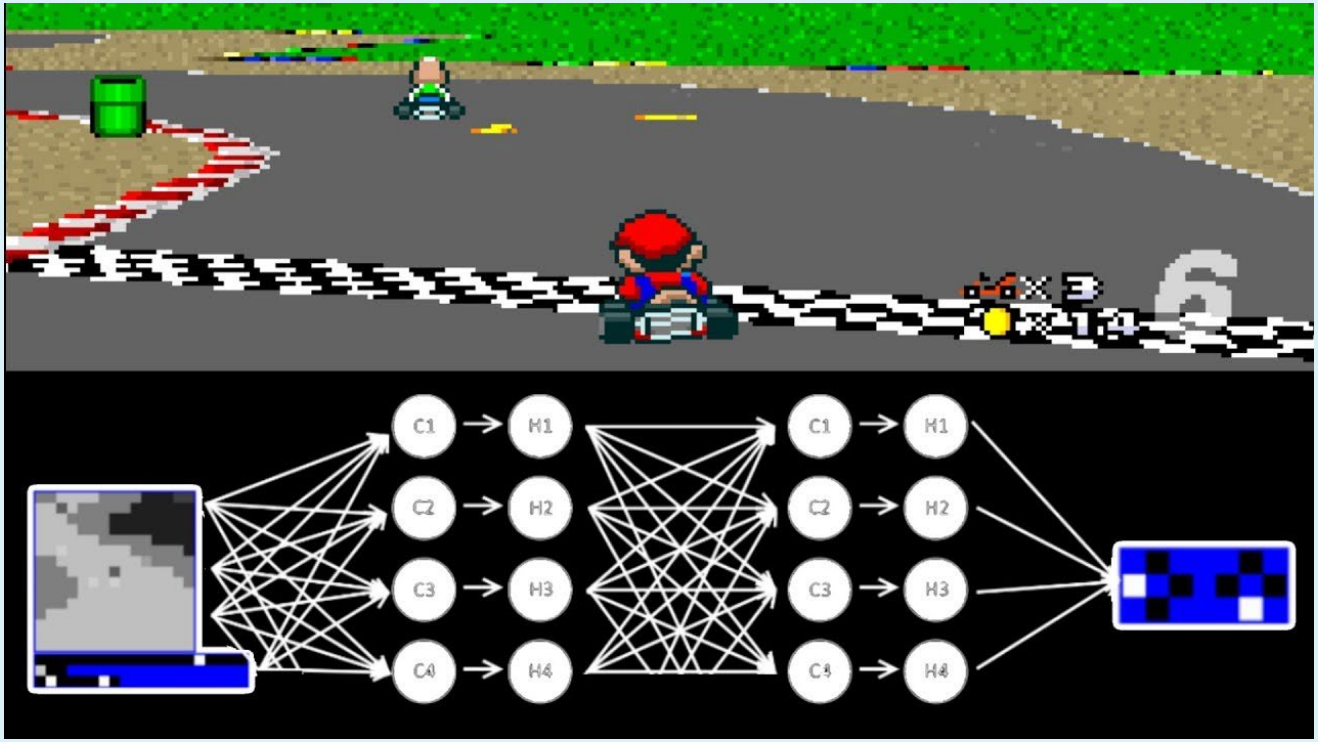




Dr. Eric Topol
Scripps Translational Research Institute

Deep Learning – what's the big deal?

Deep Learning – what's the big deal?



Deep Learning – what's the big deal?

Google computer works out how to spot cats

A Google research team has trained a network of 1,000 computers wired up like a brain to recognise cats.

The team built a neural network, which mimics the working of a biological brain, that worked out how to spot pictures of cats in just three days.

The cat-spotting computer was created as part of a larger project to investigate machine learning.

Millions of images were used to train the neural network

Google is planning to use the learning system to help with its indexing systems and with language translation.

Related Stories

http://www.bbc.com/news/technology-18595351

Deep Learning w/ EHR data

FROM MOLECULE TO PATIENT

CNBC

MIT Technology Review

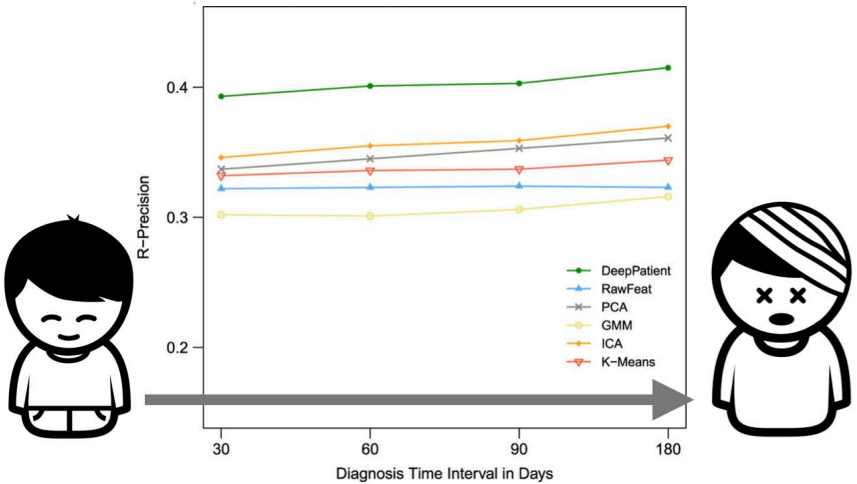
SCIENTIFIC AMERICAN

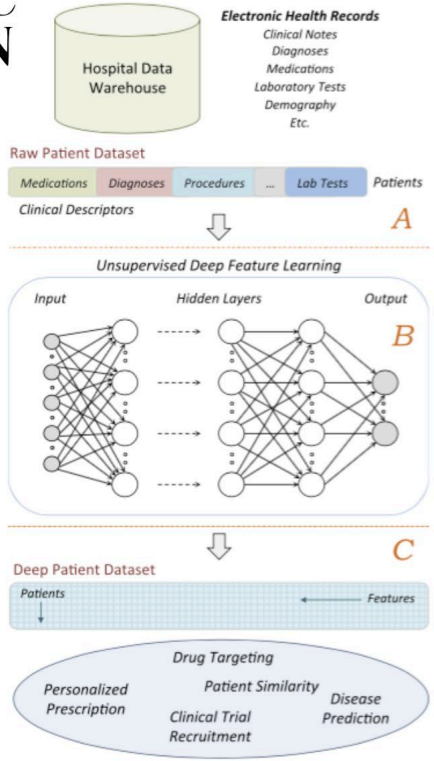
OPEN

Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records

Received: 28 January 2016
Accepted: 27 April 2016
Published: 17 May 2016

Riccardo Miotto[1,2,3], Li Li[1,2,3], Brian A. Kidd[1,2,3], Joel T. Dudley[1,2,3]





Miotto R, Li L, Kidd BA, Dudley JT. Scientific Reports 6 (2016)

Deep Learning w/ EHR data

FROM MOLECULE TO PATIENT

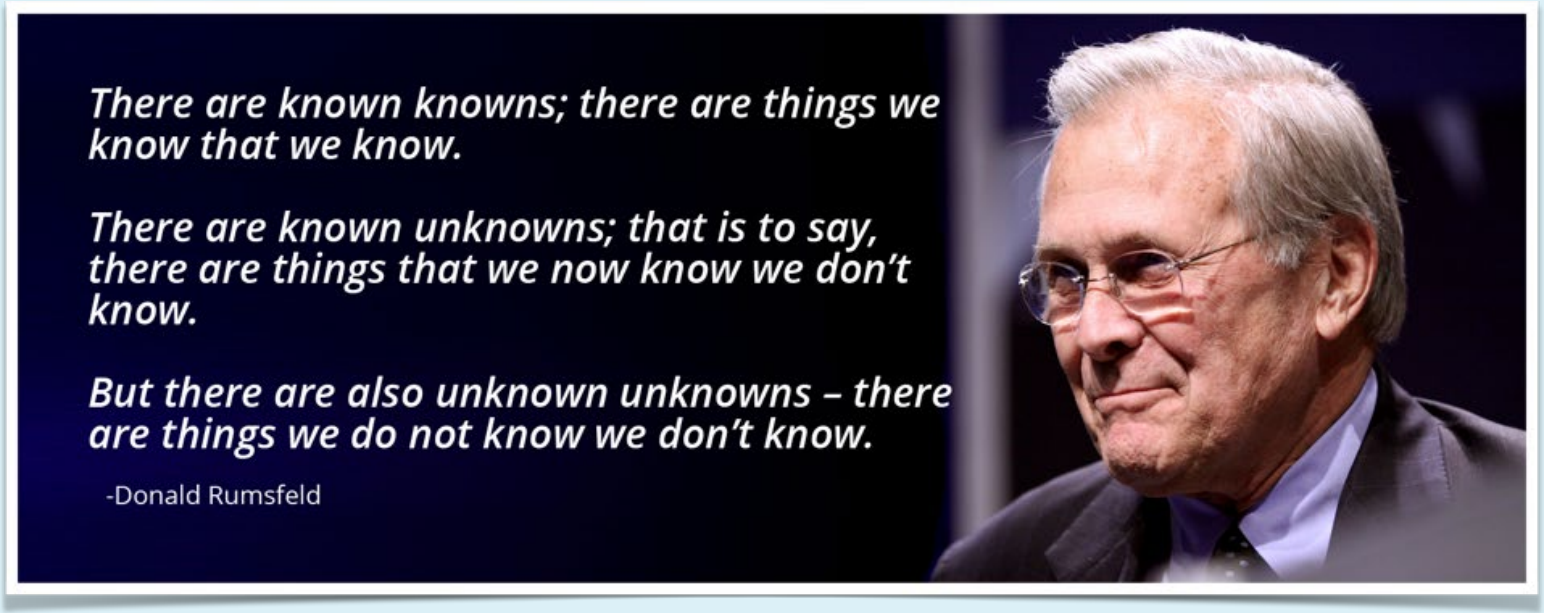
Deep Logistic Regression Network

Disease	AUC-ROC
Cancer of Liver	0.93
Regional Enteritis and Ulcerative Colitis	0.91
Type 2 Diabetes Mellitus	0.91
Congestive Heart Failure	0.90
Chronic Kidney Disease	0.89
Personality Disorders	0.89
Schizophrenia	0.88
Multiple Myeloma	0.87
Delirium and Dementia	0.85
Coronary Atherosclerosis	0.84

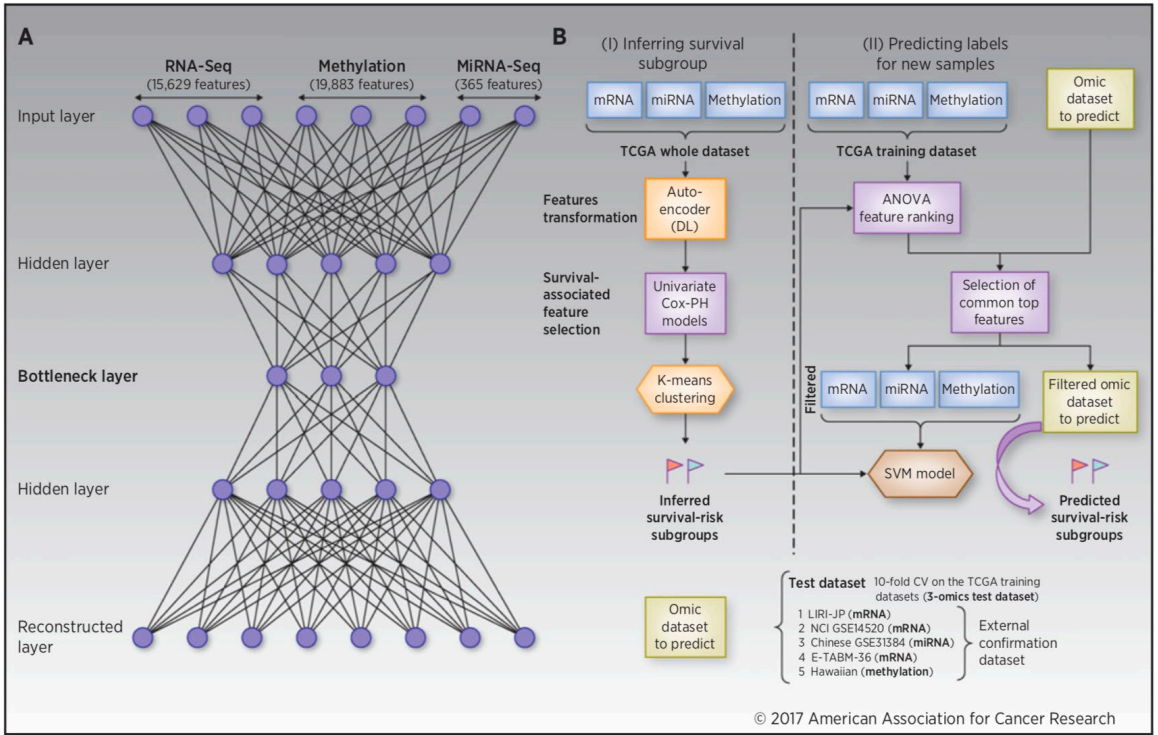
Miotto R, Li L, Kidd BA, Dudley JT. Scientific Reports 6 (2016)

FROM MOLECULE TO PATIENT

Deep Learning and Donald Rumsfeld. Bet you didn't expect that.



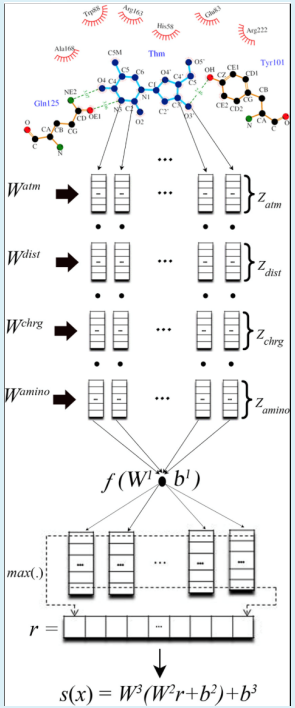
Learning latent factors and compressed representations

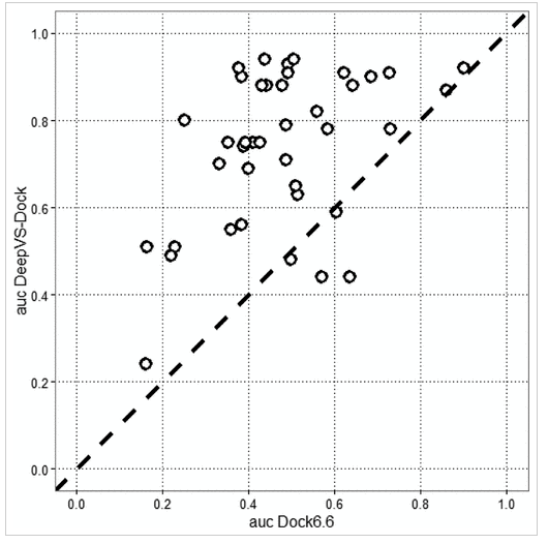


Chaudhary et al. CCR 2018

Boosting Docking-Based Virtual Screening with Deep Learning

FROM MOLECULE TO PATIENT



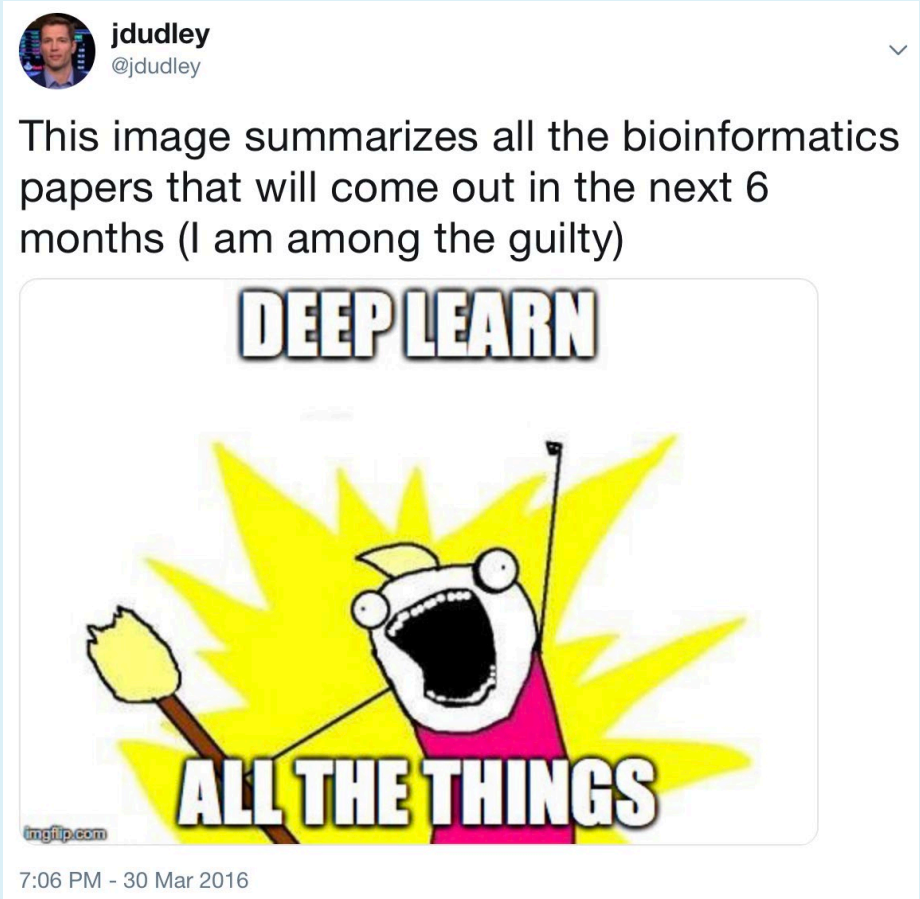


- DeepVS uses convolutional neural networks to learn abstract features (i.e. compound atom type, atomic partial charged and distance between atoms) that can discriminate active ligands
- The deep learning method outperforms two well-established virtual screening methods on a benchmarking dataset

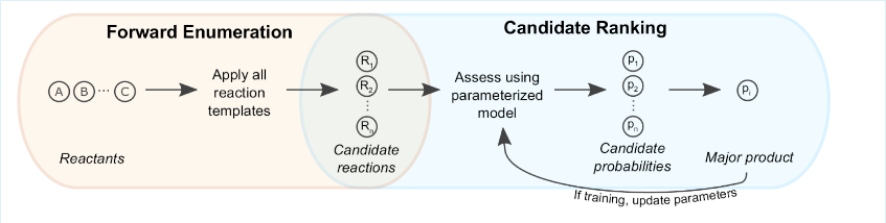
Janaina Cruz Pereira, Ernesto Raúl Caffarena, and Cicero Nogueira dos Santos *J. Chem. Inf. Model., 2016*

FROM MOLECULE TO PATIENT

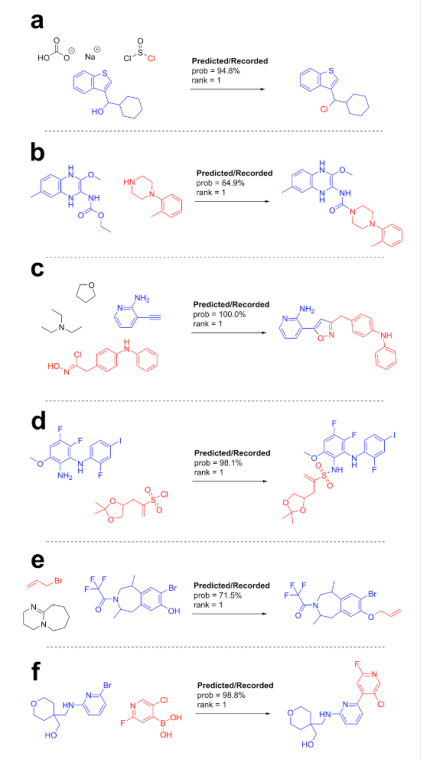
Prophecy and Deep Learning



Prediction of organic reaction outcomes using machine learning



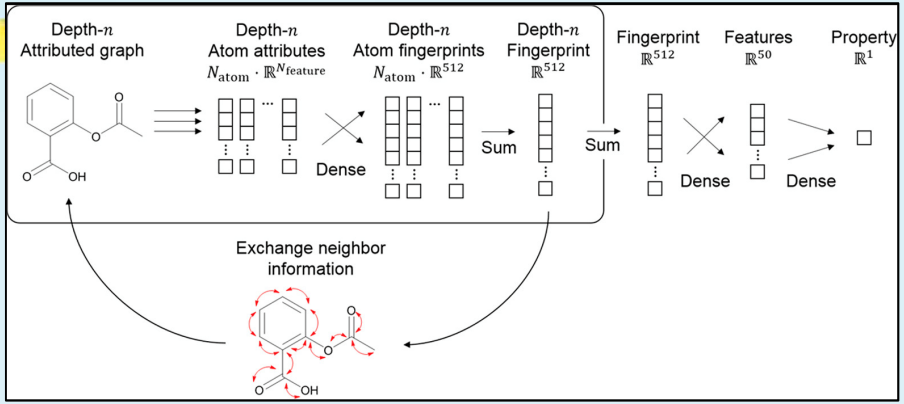
- Organic compound synthesis is difficult and requires extensive human planning.
- Existing computer-aided techniques essentially just iteratively apply known reaction templates
- Proposed reaction steps which work on paper often fail in the laboratory
- This manuscript uses published reactions from 15,000 US patents to generate lists of candidate products from reactants

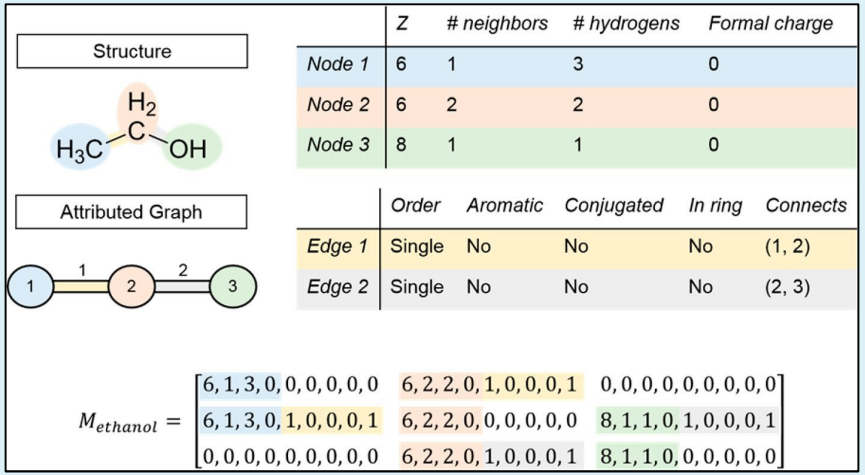


Connor Coley, Regina Barzilay, Tommi Jaakkola, William Green, Klavs Jensen *ACS Central Science, 2017*

Convolutional Embedding of Attributed Molecular Graphs for Physical Property Prediction

FROM MOLECULE TO PATIENT





	Z	# neighbors	# hydrogens	Formal charge
Node 1	6	1	3	0
Node 2	6	2	2	0
Node 3	8	1	1	0

	Order	Aromatic	Conjugated	In ring	Connects
Edge 1	Single	No	No	No	(1, 2)
Edge 2	Single	No	No	No	(2, 3)

$$M_{ethanol} = \begin{bmatrix} 6,1,3,0,0,0,0,0,0 & 6,2,2,0,1,0,0,0,1 & 0,0,0,0,0,0,0,0,0 \\ 6,1,3,0,1,0,0,0,1 & 6,2,2,0,0,0,0,0,0 & 8,1,1,0,1,0,0,0,1 \\ 0,0,0,0,0,0,0,0,0 & 6,2,2,0,1,0,0,0,1 & 8,1,1,0,0,0,0,0,0 \end{bmatrix}$$

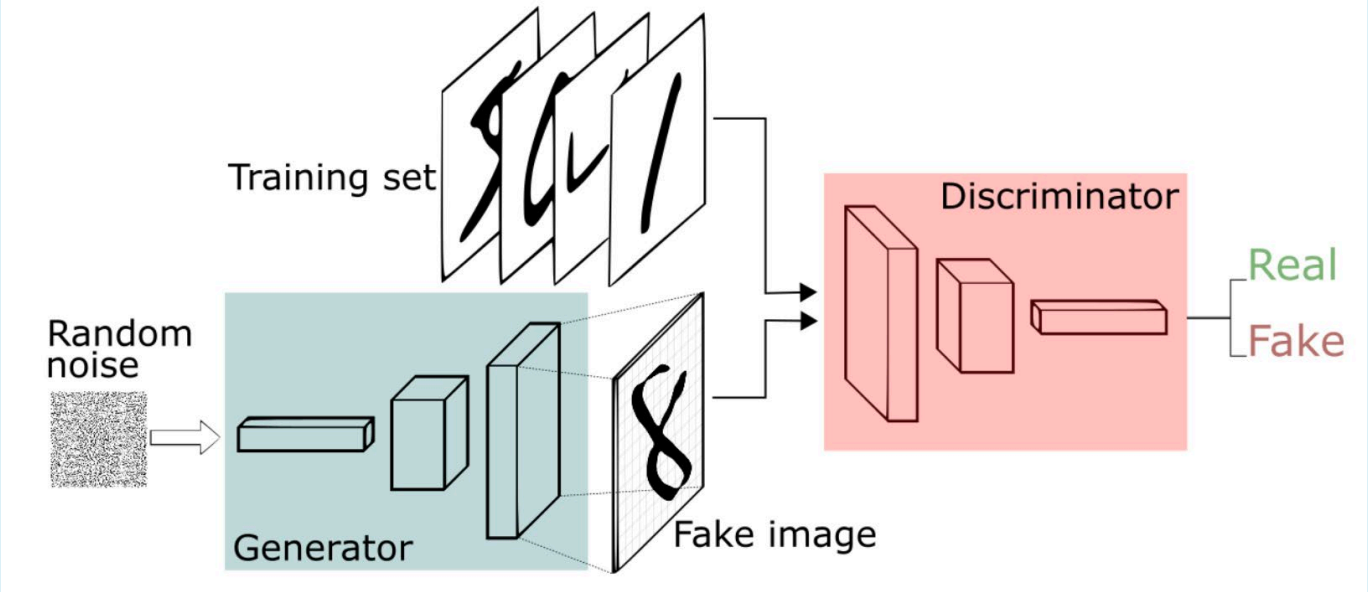
Connor Coley, Regina Barzilay, William Green, Tommi Jaakkola, and Klavs Jensen *J Chem Inf Model 2017*

Hello, I don’t exist!

https://thispersondoesnotexist.com/

StyleGAN – Kerras et al. 2018

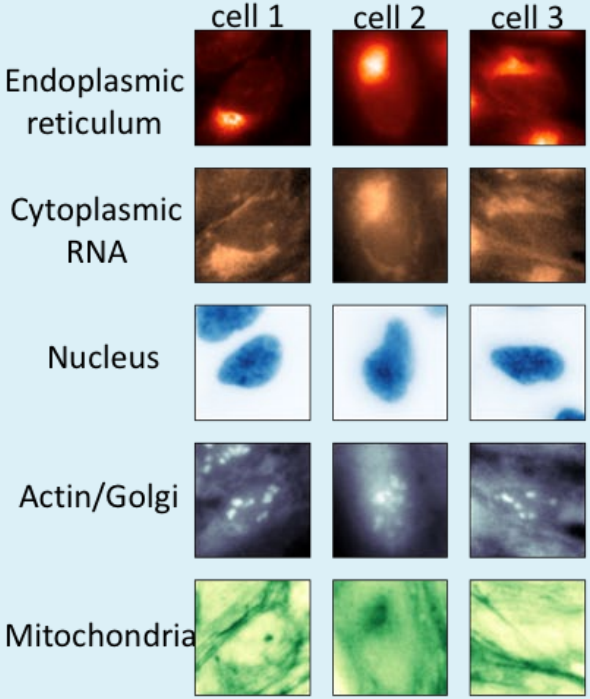
Generative Adversarial Networks (GAN)



GAN applied to phenotypic screening scenario

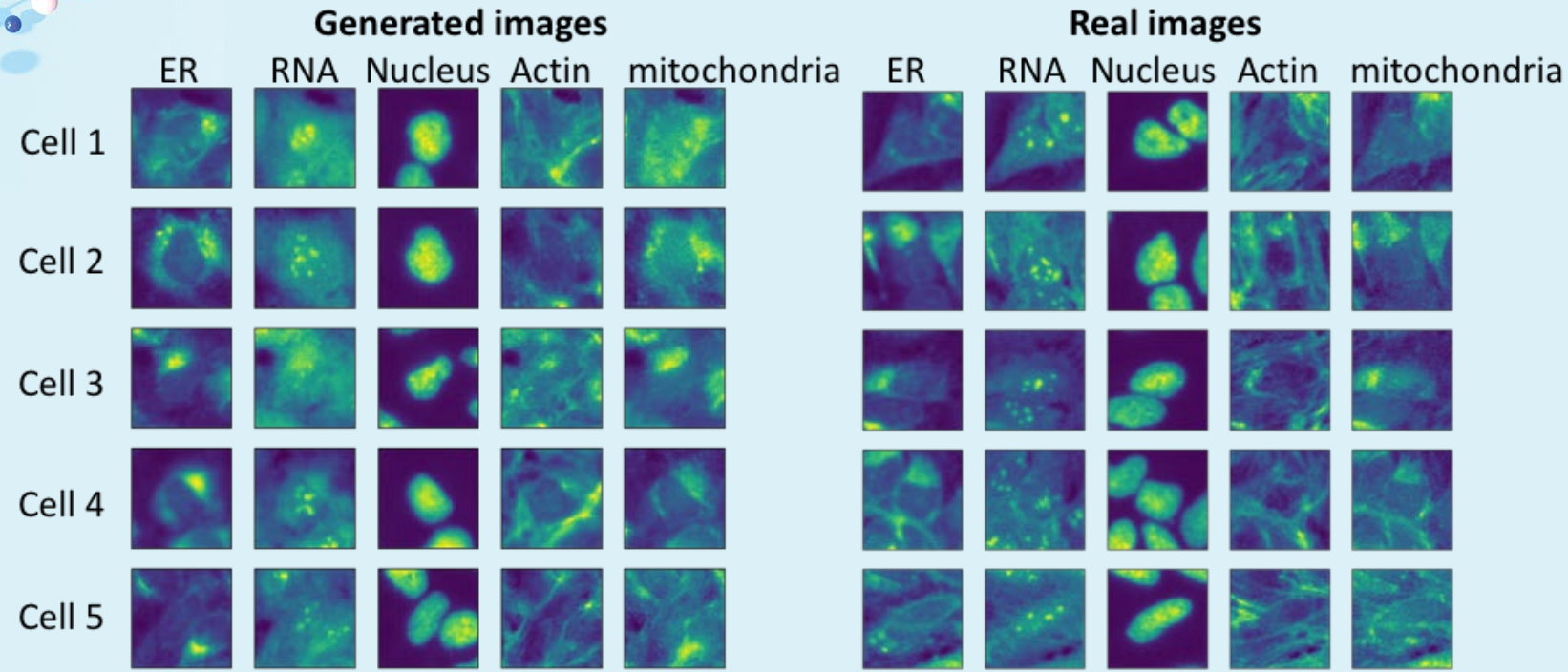
FROM MOLECULE TO PATIENT

- On average 4 repeats per treatment
- 6 images form each repeat
- 10~50 cells from 1 images
- 2 TB of image patches (64x64)



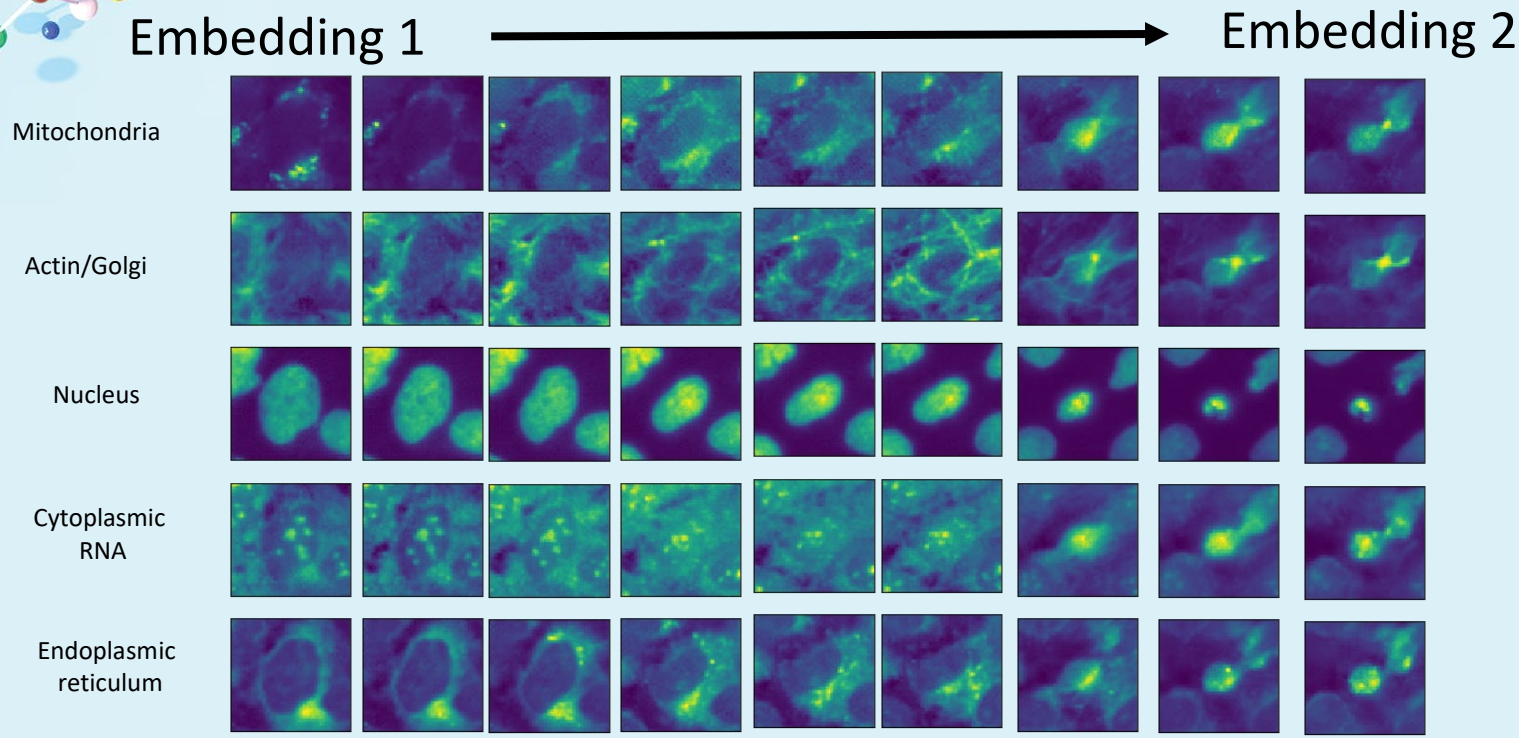
GAN applied to phenotypic screening scenario

FROM MOLECULE TO PATIENT

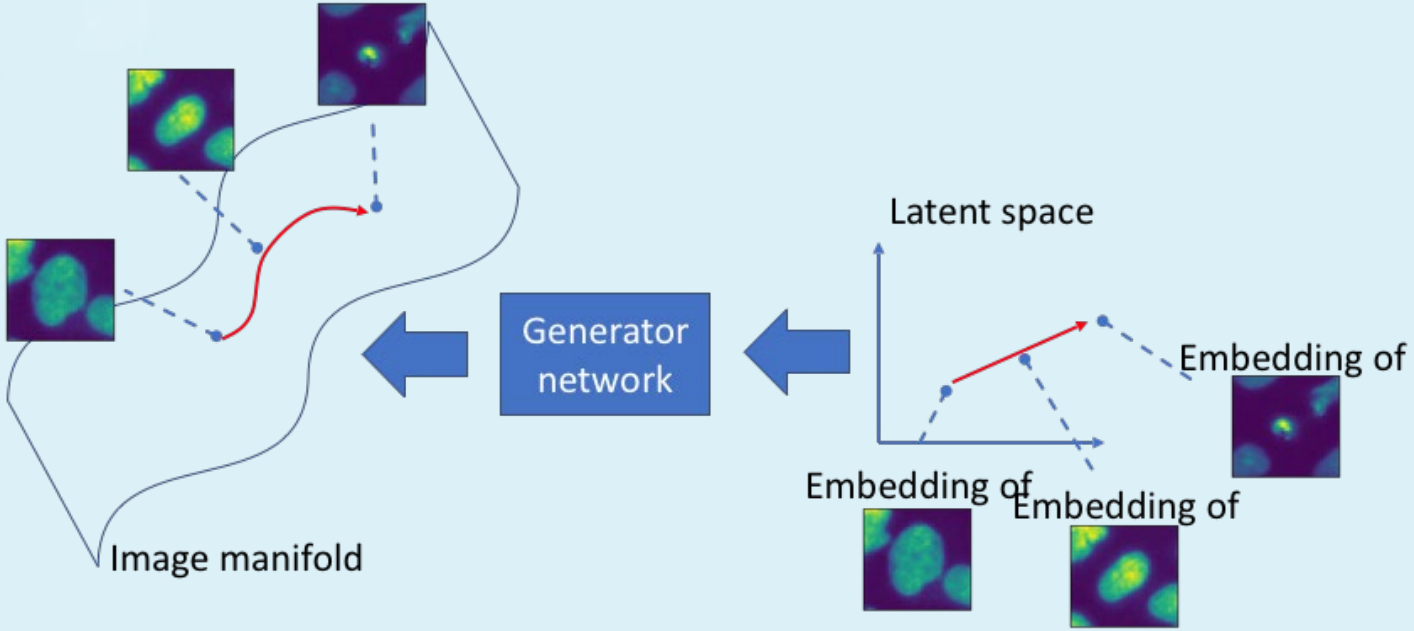


Moving from one image to another in the latent space

FROM MOLECULE TO PATIENT

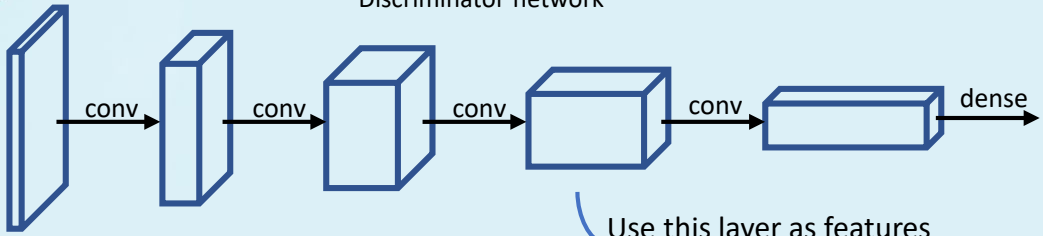


GANs encode “action” on the images



Using the discriminator to discover bioactive chemicals

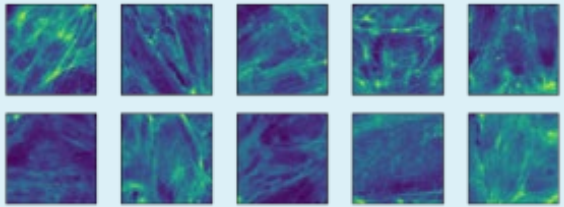
Discriminator network



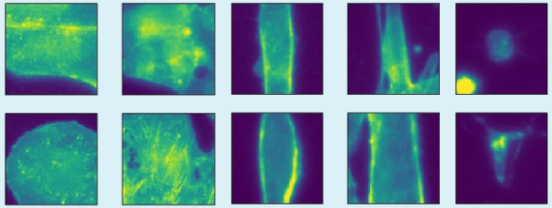
Use this layer as features

Novelty discovery → Phenotypes that are different from those of untreated cells → Classify if bioactive

Untreated cells (actin filament)



Novel phenotypes (actin filament)



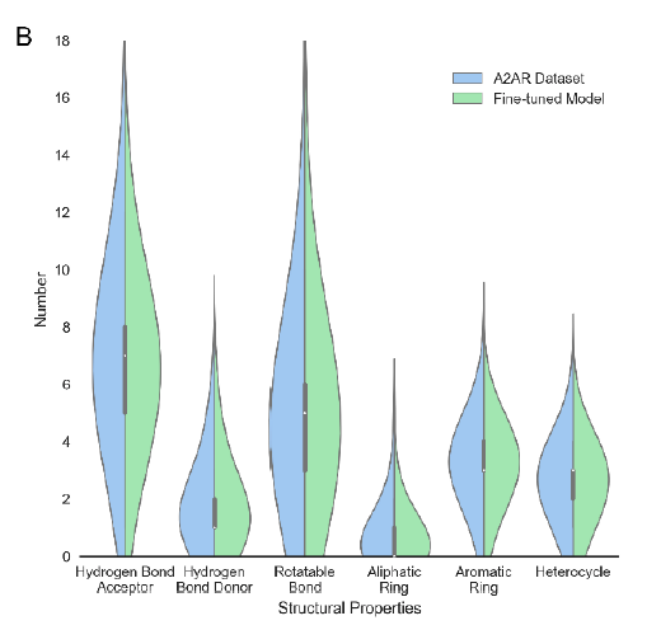
Benchmark on classifying 30 selected bioactive chemicals

	AUC-ROC	AUC-PR
GAN	**0.989±0.007**	**0.991±0.007**
Autoencoder	0.981±0.005	0.985±0.004
cell-profiler	0.967±0.020	0.970±0.019

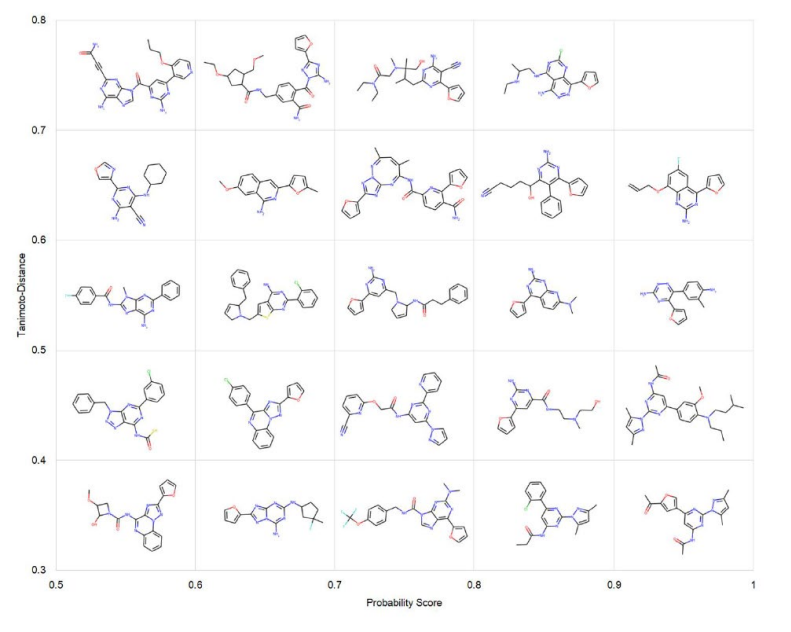
An exploration strategy improves the diversity of *de novo* ligands using deep reinforcement learning: a case for the adenosine A_{2A} receptor

FROM MOLECULE TO PATIENT

De Novo molecules similar to training molecules



Example generated molecules

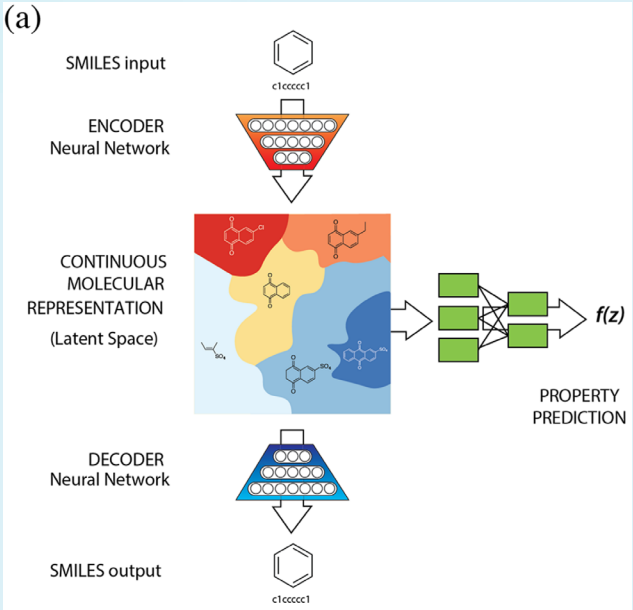


Xuhan Liu, Kai Ye, Herman W. T. van Vlijmen, Adriaan P. IJzerman, Gerard J. P. van Westen. ChemArxiv

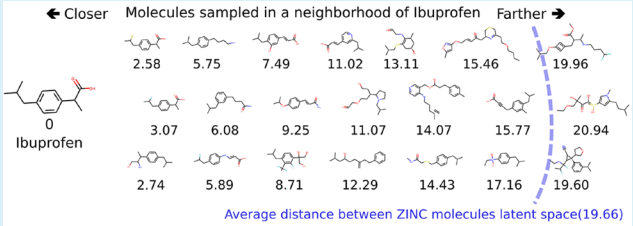
Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules

FROM MOLECULE TO PATIENT

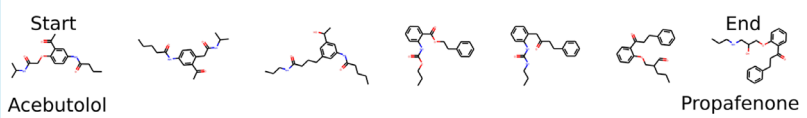
Estimate a continuous latent space for molecular structures by the variational autoencoder



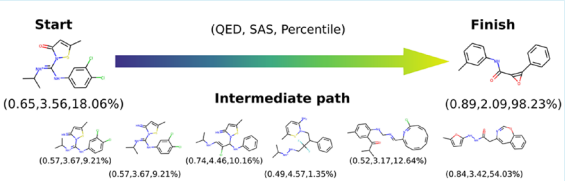
Search similar molecular structures in the latent space



Interpolate between two structures in the latent space



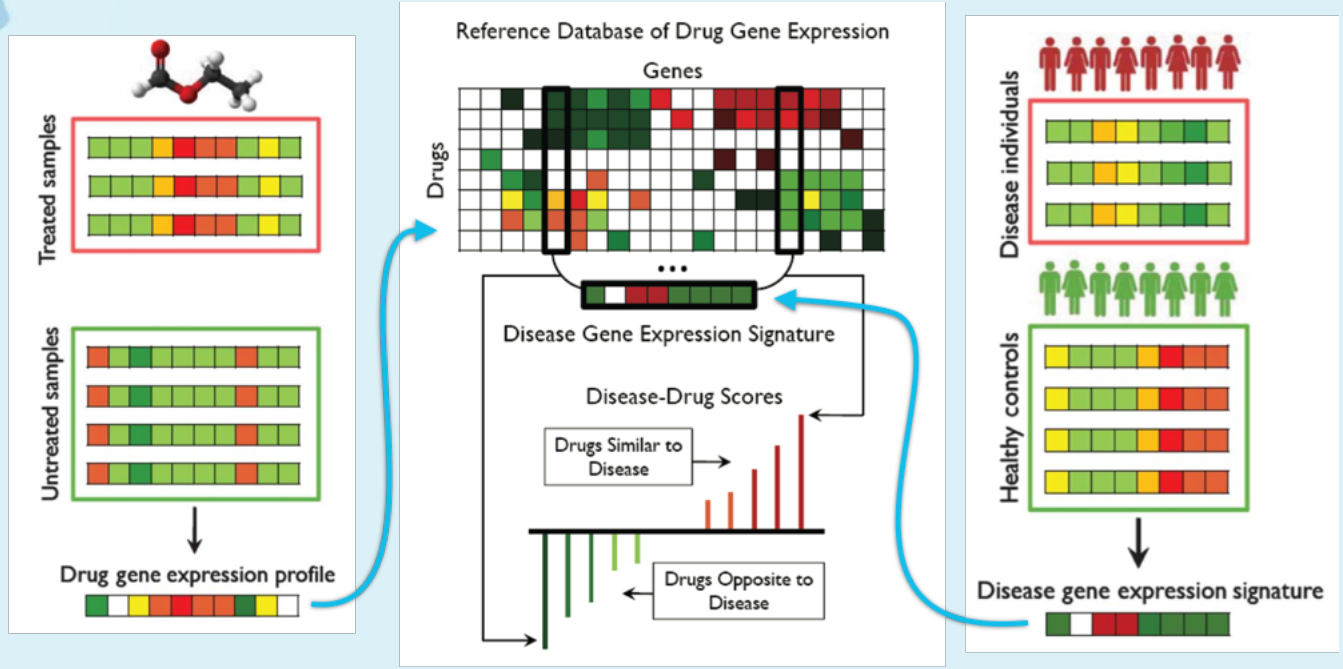
Generate a molecule that optimizes given properties



QED: Quantitative Estimation of Drug-likeness
SAS: Synthetic Accessibility score

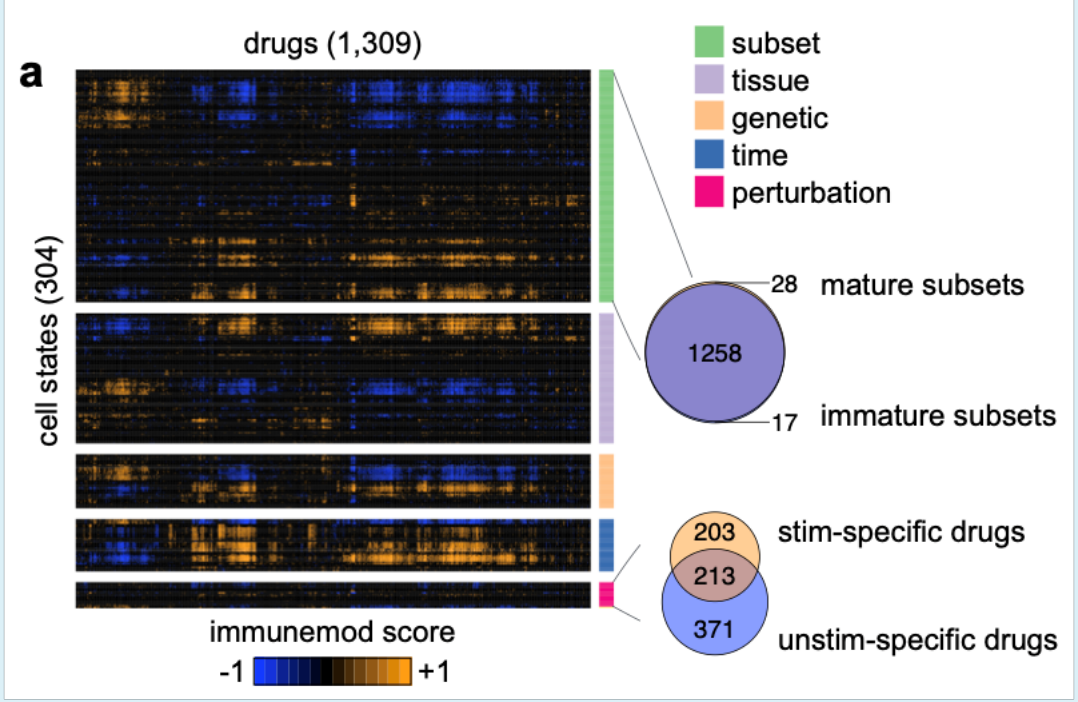
Gómez-Bombarelli, Rafael, et al., *ACS central science*, 2018

Starting with chemogenomic drug-disease relationships



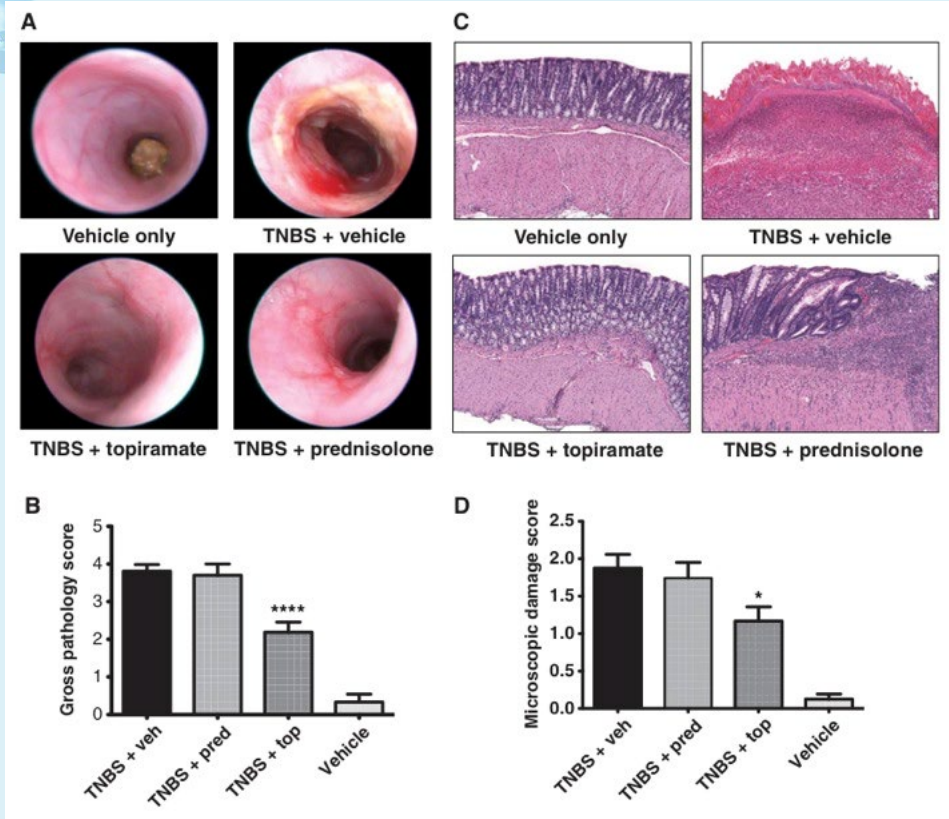
Sirota, M., Dudley, J. T., et al. (2011). Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data. *Science Translational Medicine*, 3(96).

Starting with chemogenomic drug-disease relationships



Kidd BA, Wroblewska A, Boland MR, Agudo J, Merad M, Tatonetti NP, Brown BD, Dudley JT. Mapping the effects of drugs on the immune system. ***Nature Biotechnology*** 34, 47–54 (2016)

Starting with chemogenomic drug-disease relationships

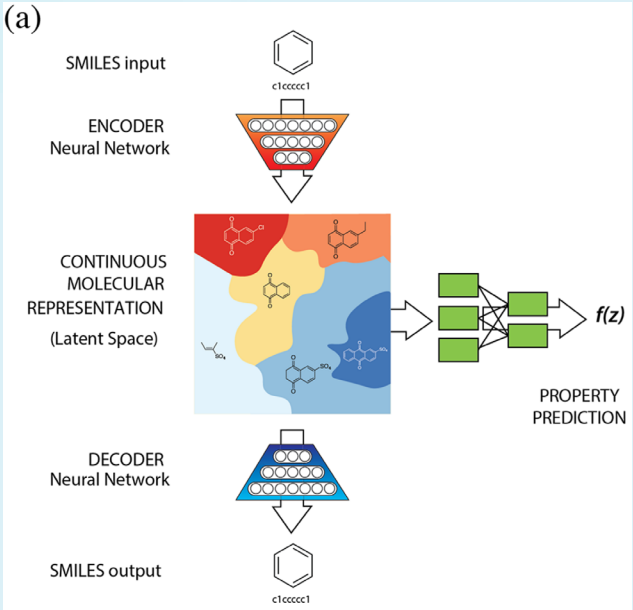


Dudley, J. T., Sirota, M., et al. (2011). Computational Repositioning of the Anticonvulsant Topiramate for Inflammatory Bowel Disease. Science Translational Medicine, 3(96).

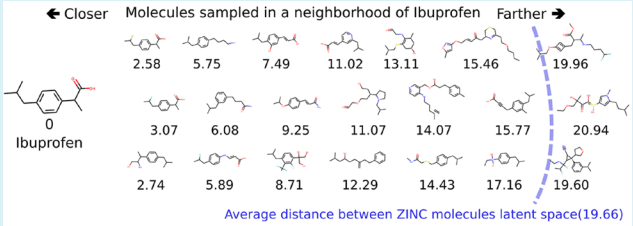
Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules

FROM MOLECULE TO PATIENT

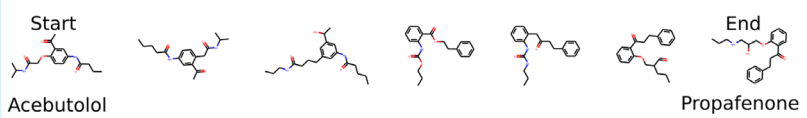
Estimate a continuous latent space for molecular structures by the variational autoencoder



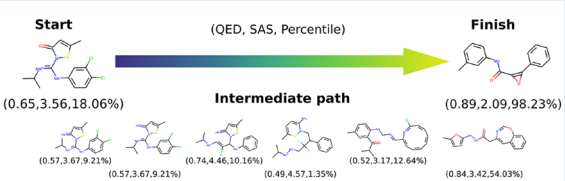
Search similar molecular structures in the latent space



Interpolate between two structures in the latent space



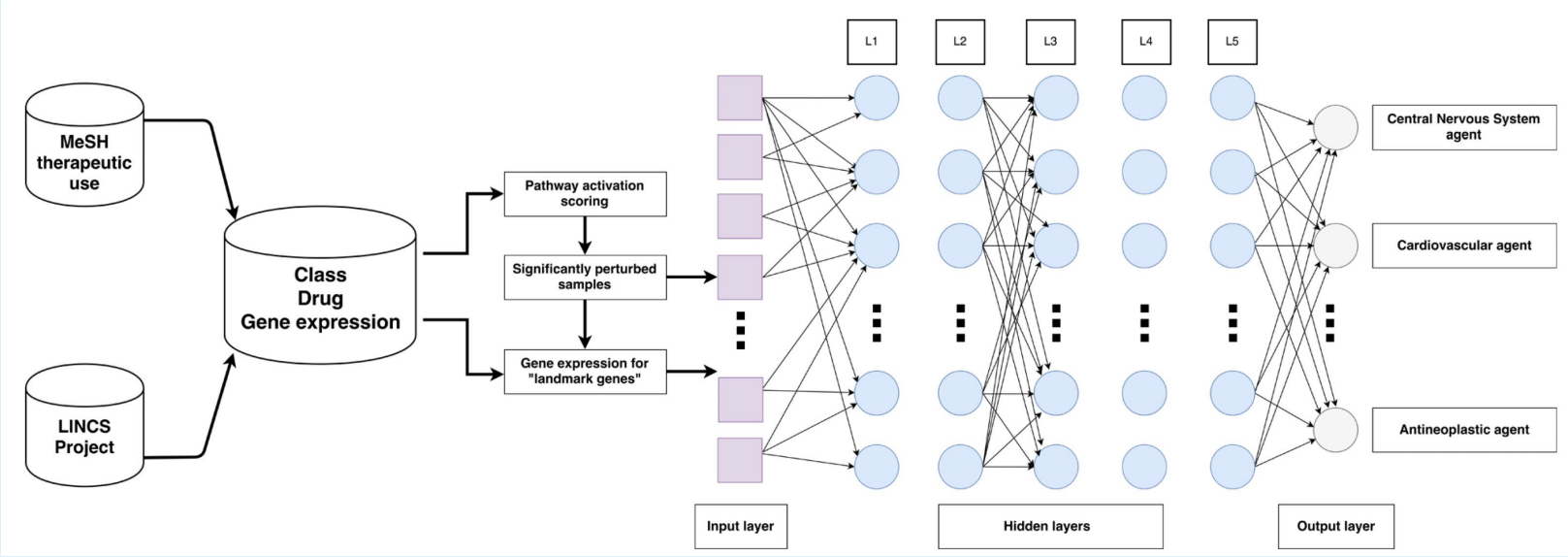
Generate a molecule that optimizes given properties



QED: Quantitative Estimation of Drug-likeness
SAS: Synthetic Accessibility score

Gómez-Bombarelli, Rafael, et al., *ACS central science*, 2018

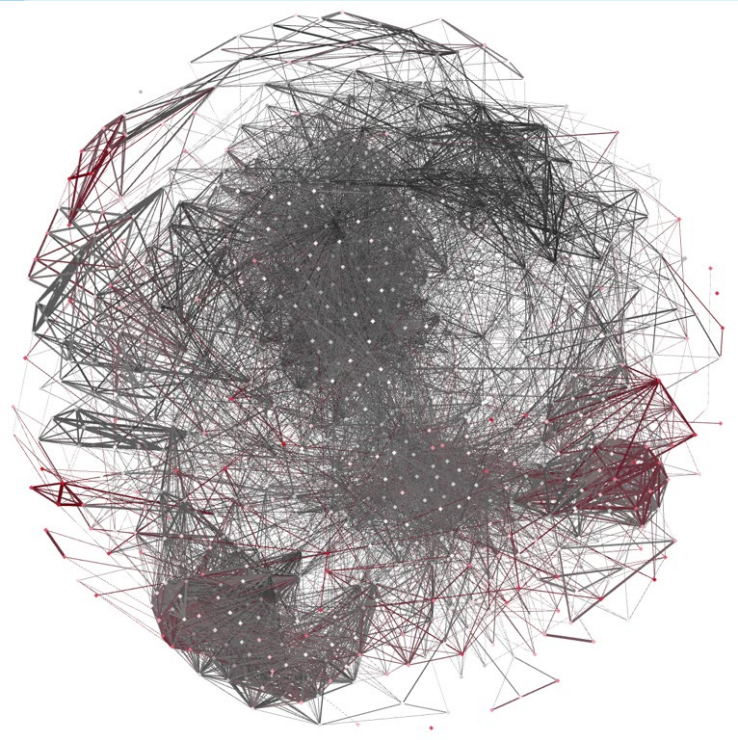
Multi-modal representation learning on chemogenomic data

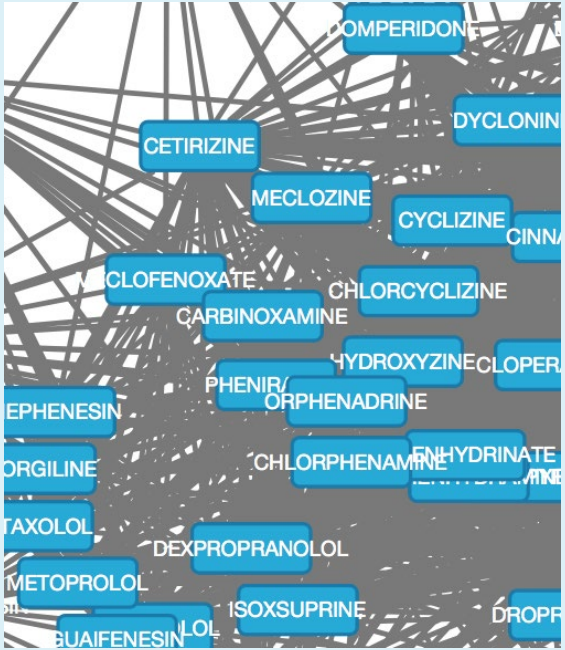


Aliper, Alexander, et al., *Molecular pharmaceutics*, 2016

Multi-modal representation learning on chemogenomic data

FROM MOLECULE TO PATIENT

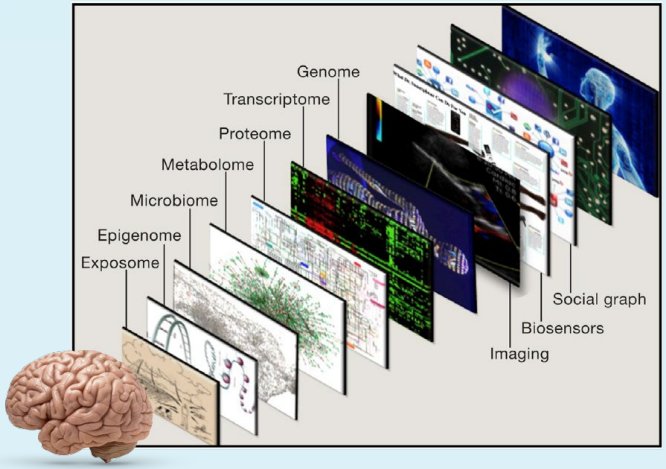


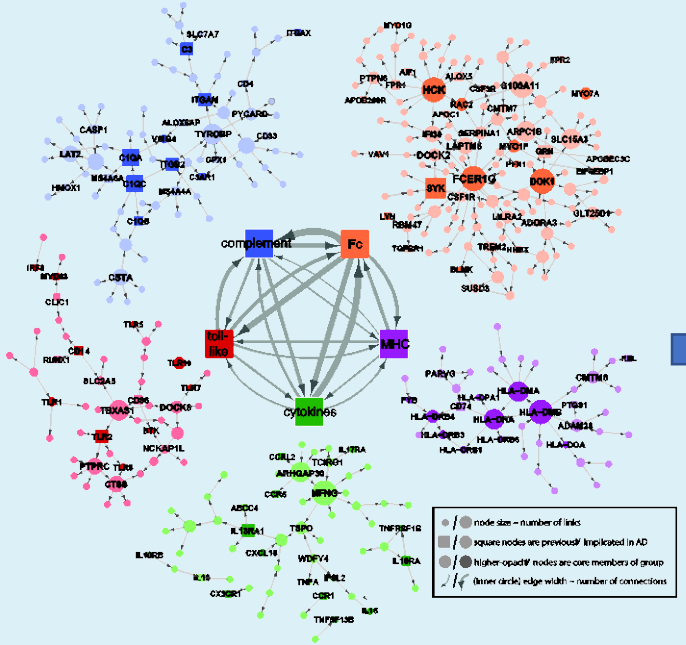


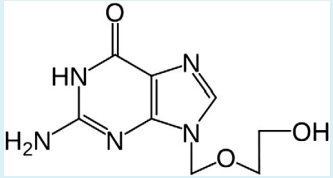
Unpublished work

The holy grail of generative deep learning for drug discovery

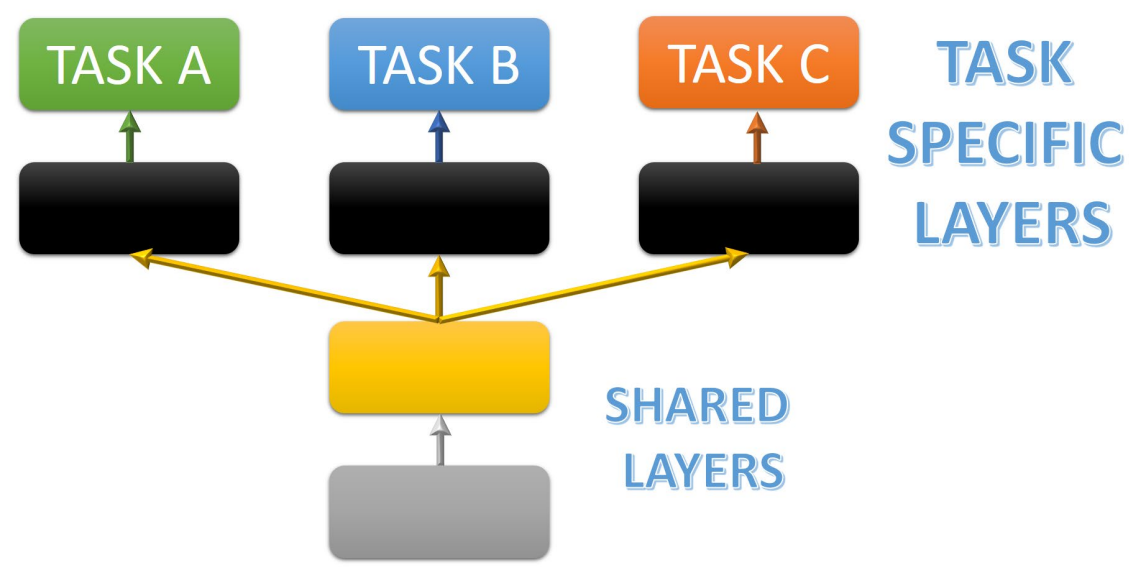
FROM MOLECULE TO PATIENT



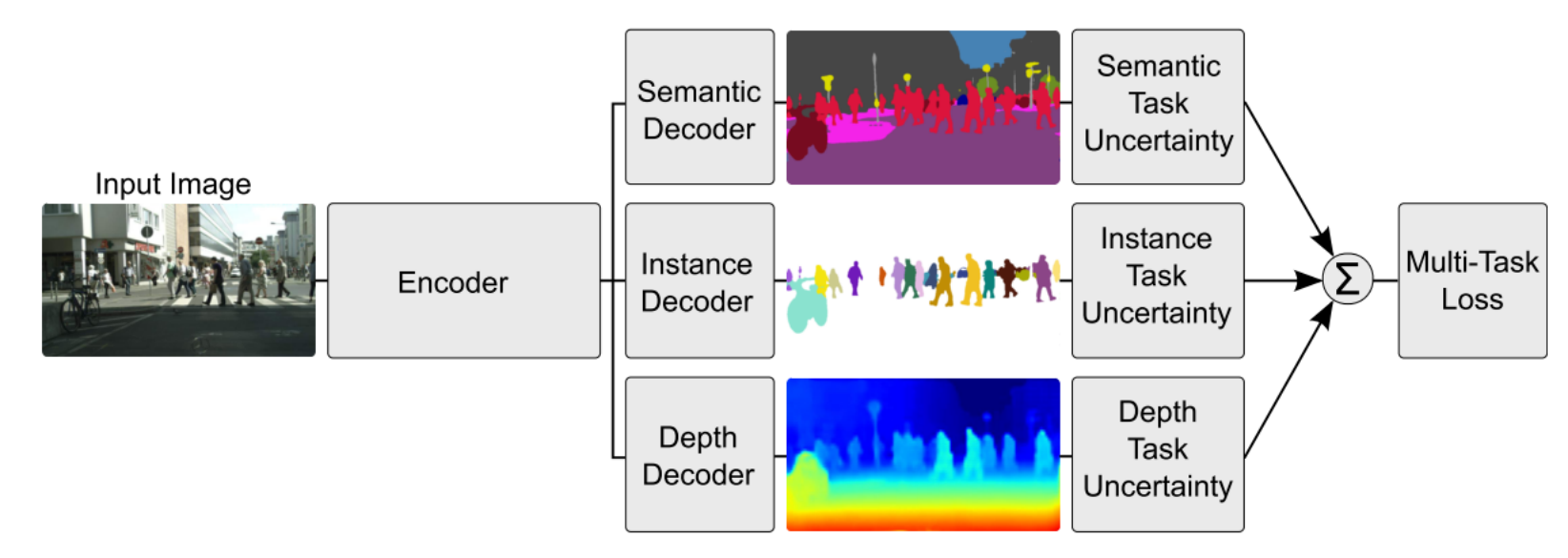




Multi-Task Learning



Multi-Task Learning



Source: http://ruder.io/multi-task/

Multi-Task Learning

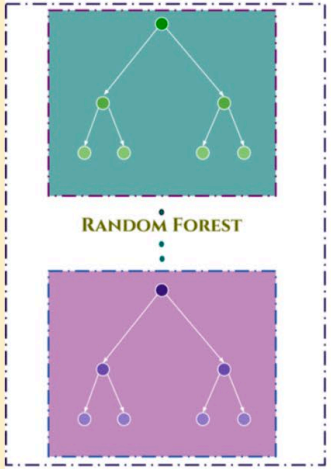
JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

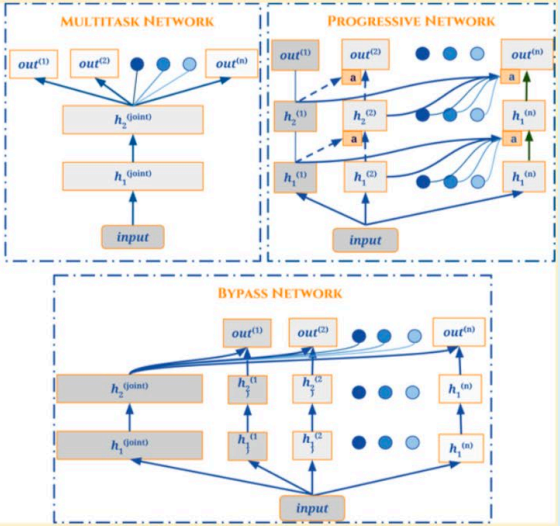
pubs.acs.org/jcim

Is Multitask Deep Learning Practical for Pharma?

Bharath Ramsundar,[†] Bowen Liu,[‡] Zhenqin Wu,[‡] Andreas Verras,[¶] Matthew Tudor,[§] Robert P. Sheridan,[¶] and Vijay Pande*,[‡]



`import deepchem as dc`	Import DeepChem
`model = dc.models.TensorflowMultiTaskRegressor(len(KAGGLE_tasks), n_features, layer_sizes=[2000, 1000, 1000], dropouts=[.25, .25, .25], weight_init_stddevs=[.02, .02, .02], bias_init_consts=[1., 1., 1.], learning_rate=.00003, penalty=.0004, penalty_type="l2", seed=123, optimizer="adam", batch_size=100)`	Define model
`model.fit(train_dataset, nb_epoch=10)`	Fit to data
`metric = dc.metrics.Metric( dc.metrics.pearson_r2_score, task_averager=np.mean) train_scores = model.evaluate(train_dataset, [metric]) valid_scores = model.evaluate(valid_dataset, [metric]) test_scores = model.evaluate(test_dataset, [metric])`	Evaluate model fit



Thank You!

FROM MOLECULE TO PATIENT

- Mount Sinai
  - Sam Gandy
  - Hao Chi
  - Riccardo Miotto
  - Kipp Johnson
  - Jessica DeFreitas
- ASU
  - Ben Readhead

Email: joel.dudley@mssm.edu
Twitter: @jdudley

NIH National Institute on Aging