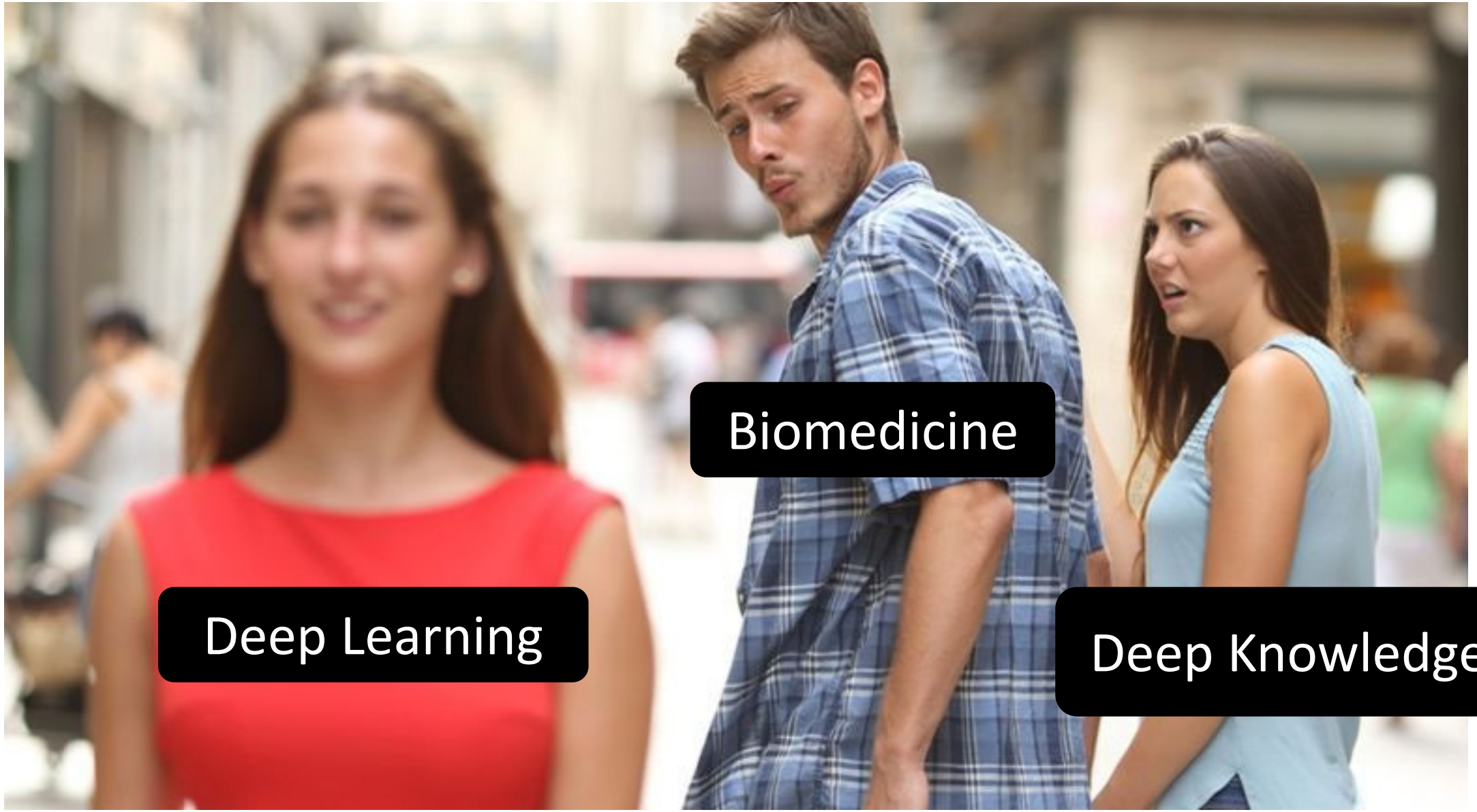
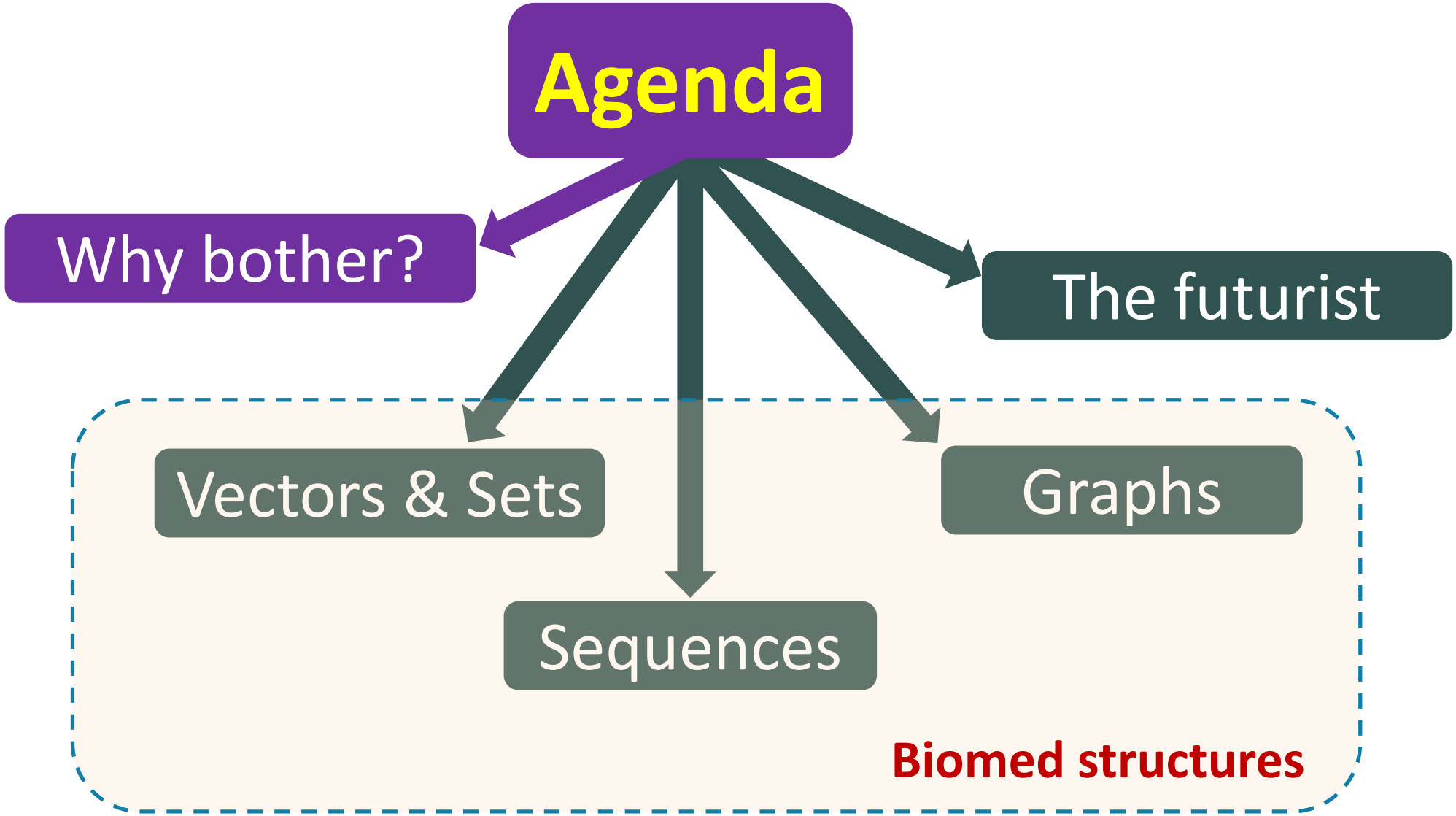
Deep Learning for **Biomedicine**

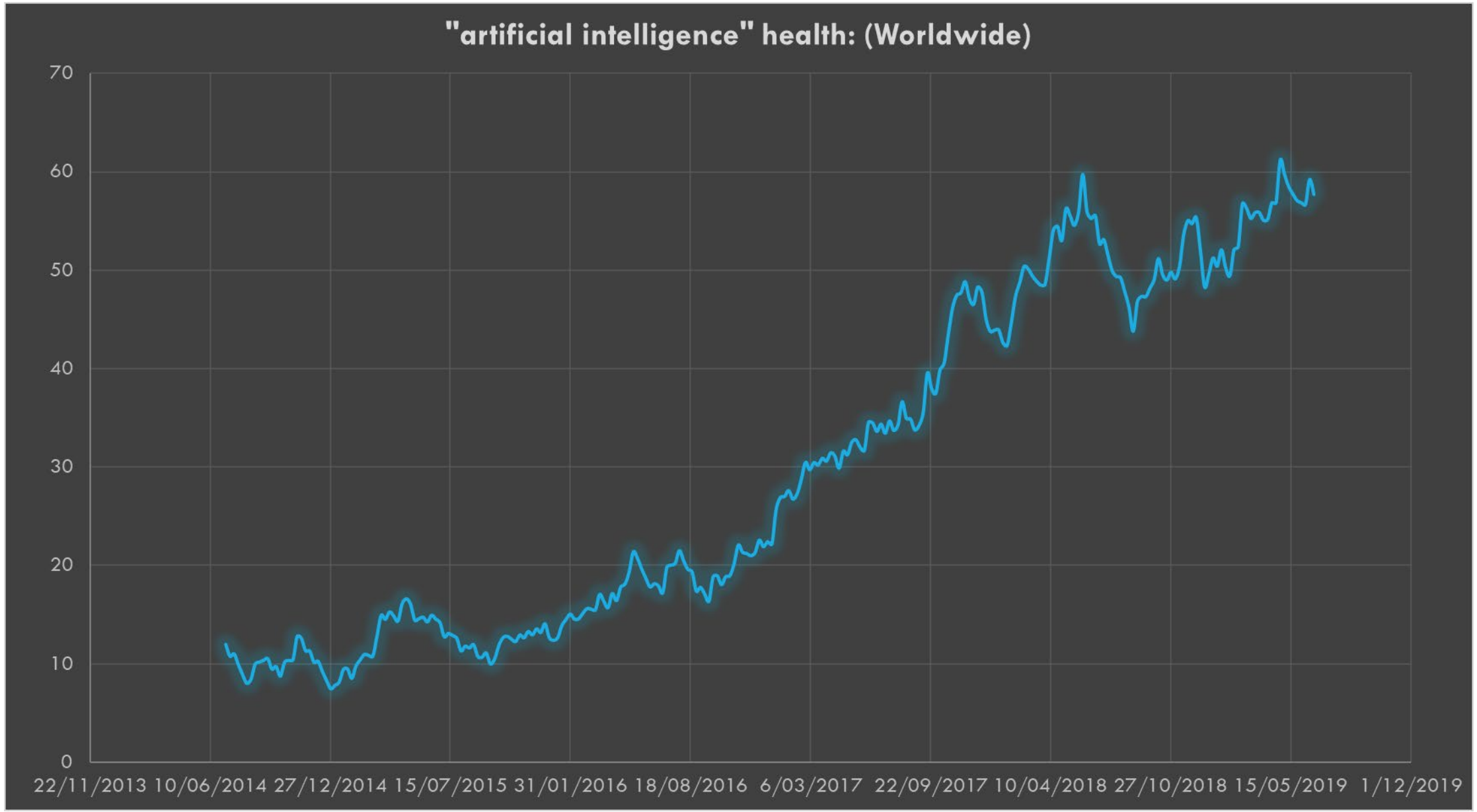
Truyen Tran
Deakin University

truyen.tran@deakin.edu.au
truyentran.github.io
@truyenoz
letdataspeak.blogspot.com
goo.gl/3jJ1O0

Jakarta, July 2019


Biomedicine
Deep Learning
Deep Knowledge


Agenda
Why bother?
The futurist
Vectors & Sets
Graphs
Sequences
Biomed structures


"artificial intelligence" health: (Worldwide)
70
60
50
40
30
20
10
0
22/11/2013
10/06/2014
27/12/2014
15/07/2015
31/01/2016
18/08/2016
6/03/2017
22/09/2017
10/04/2018
27/10/2018
15/05/2019
1/12/2019

Recent AI/ML/KDD activities

Conference on Machine Learning for Healthcare (MLHC), 2019

ICML/IJCAI/AAAI (2019)

- Health Intelligence
- Workshop on Computational Biology
- Knowledge Discovery in Healthcare III: Towards Learning Healthcare Systems (KDH)

KDD/SDM/ICDM (2018-2019)

- **Health Day at KDD'18**
- epiDAMIK: Epidemiology meets Data Mining and Knowledge discovery
- 17th International Workshop on Data Mining in Bioinformatics
- Workshop on Data Mining in Bioinformatics (BIOKDD 2019)
- [DsHealth 2019] 2019 KDD workshop on Applied data science in Healthcare: bridging the gap between data and knowledge

Why now?

High-impact & **data-intensive**.

- Andrew Ng's rule: impact on 100M+ people.
- Biomedicine is the only industry that will never shrink!

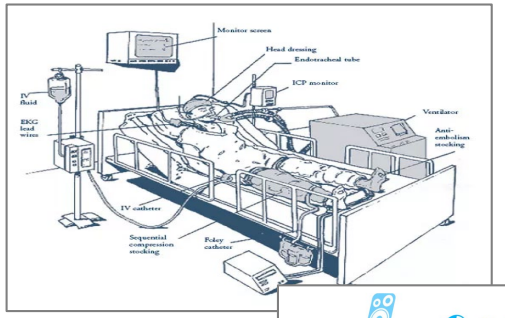
Ripe for innovations fuelled by deep learning techniques.

- Major recent advances and low hanging fruits are being picked.

Great **challenges**:

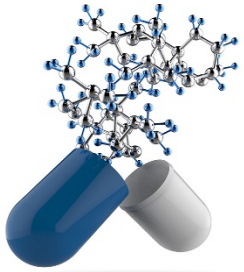
- High volume and high dimensional;
- Any modality: 2D-4D vision, time-series, 1D signals, sound, text, social network, graphs.
- Metric scale from nano-meter (atoms) to meters (human body and brain).
- Time scale from mini-seconds (ion channels) to 100 years.
- Complexity unimaginable (e.g., brain, DNA, cell networks).
- Great privacy concerns;

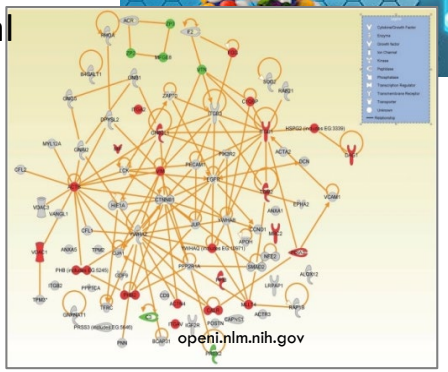
It is the right time to join force with biomedical scientists!



healthpages.org


ase.edu


pharmacy.umaryland.edu



marketingland.com

PubMed

Big Rooms in Biomedicine

Human genome

3 billion base-pairs (characters), 20K genes, 98% non-coding regions

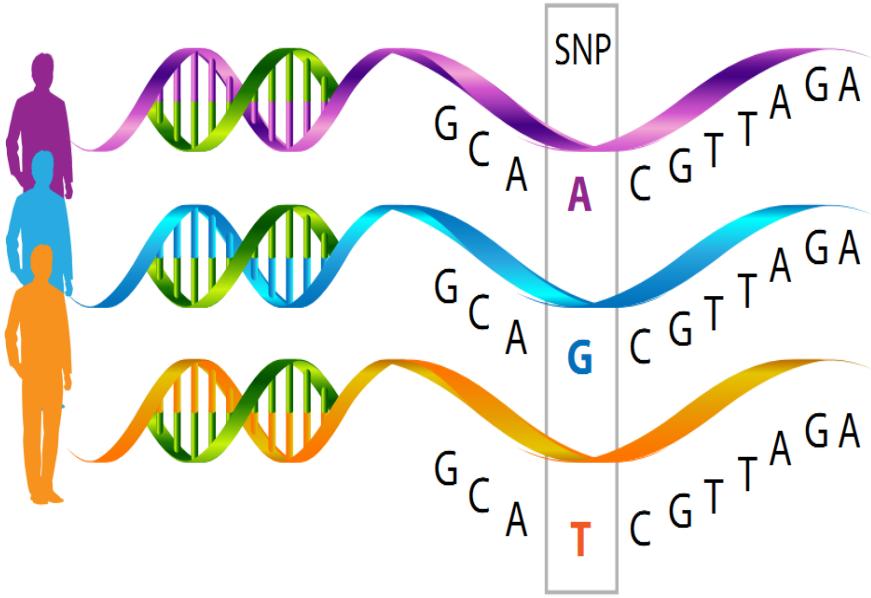
Any two random persons share 99.9% genome

The 0.1% difference is thought to account for all variations between us

- Appearance: Height (80% heritable), BMI, hair, skin colors
- IQ, education levels
- Genetic disorders such as cancers, bipolar, schizophrenia, autism, diabetes, etc.

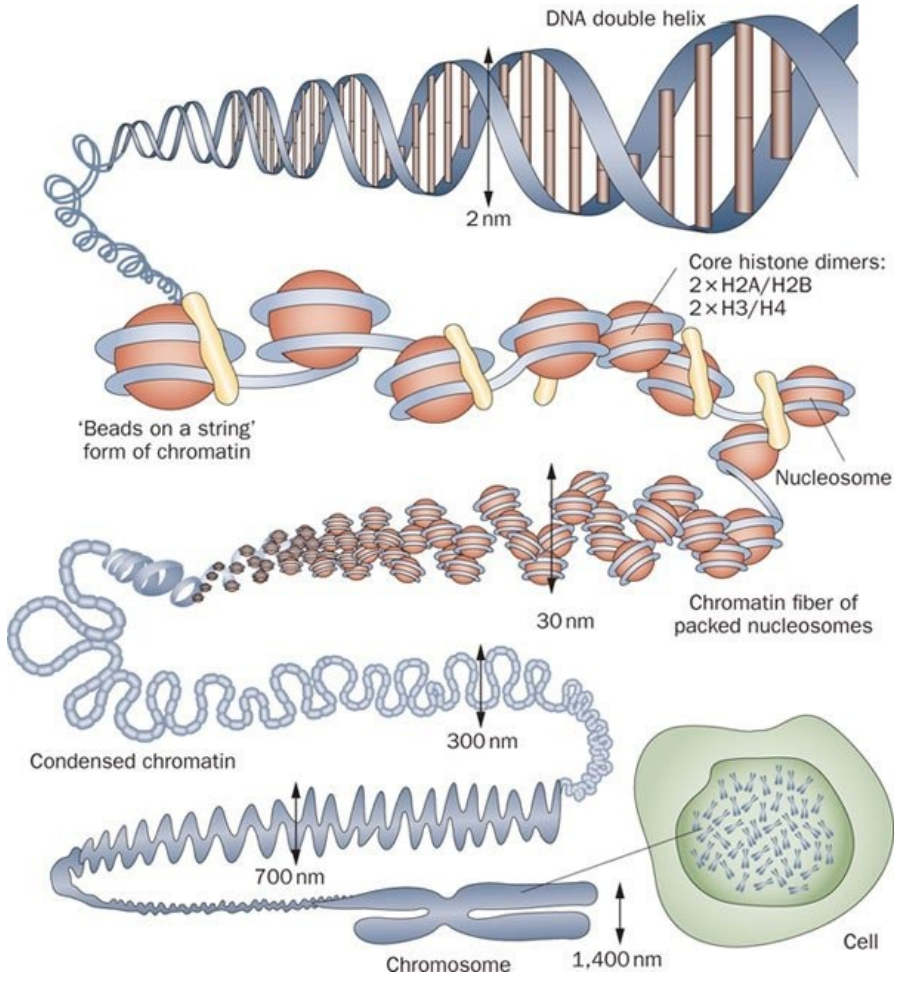
Any two random persons share about 60% variations (SNV/SNP)

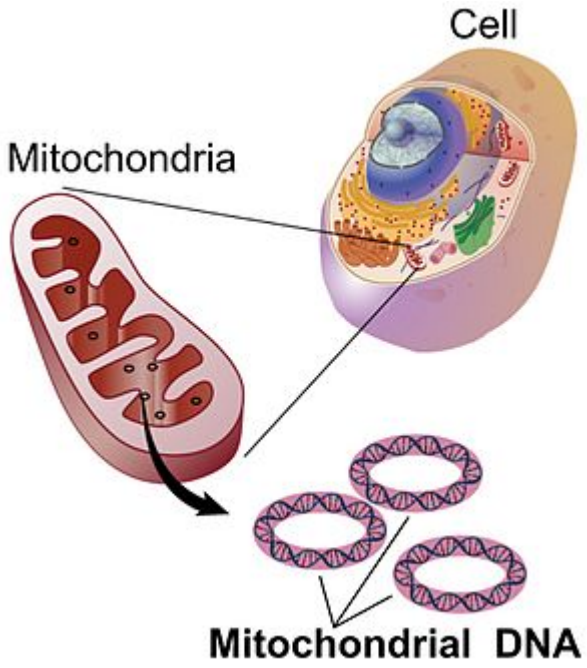
As we age, there are small mutations within our cells

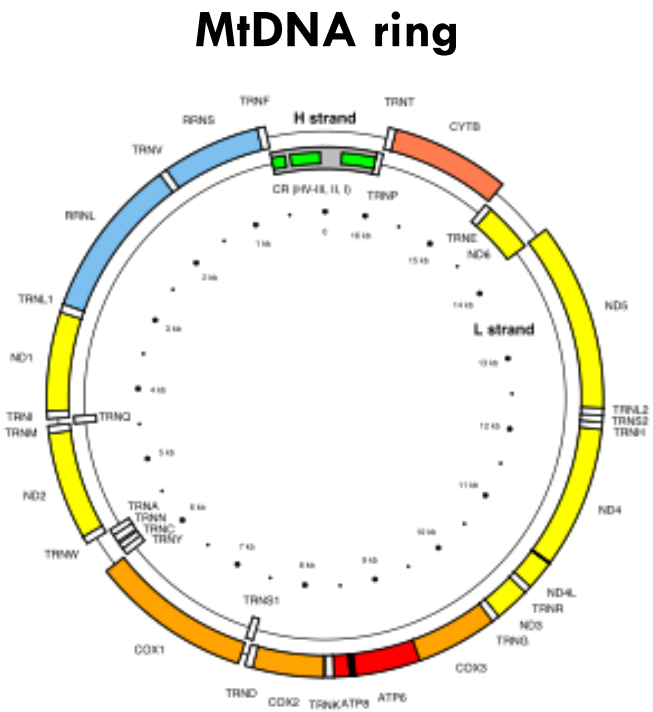


https://neuroendoimmune.files.wordpress.com

The cell, nuclear DNA & MtDNA







https://qph.ec.quoracdn.net/main-qimg-2c39fede406d71fb534bbae6cc9b8aad-c
https://en.wikipedia.org/wiki/Mitochondrial_DNA

Sequencing

The first step is to read (sequence) the DNA/MtDNA, and represent the information as string of characters (A,C,G,T) in computer.

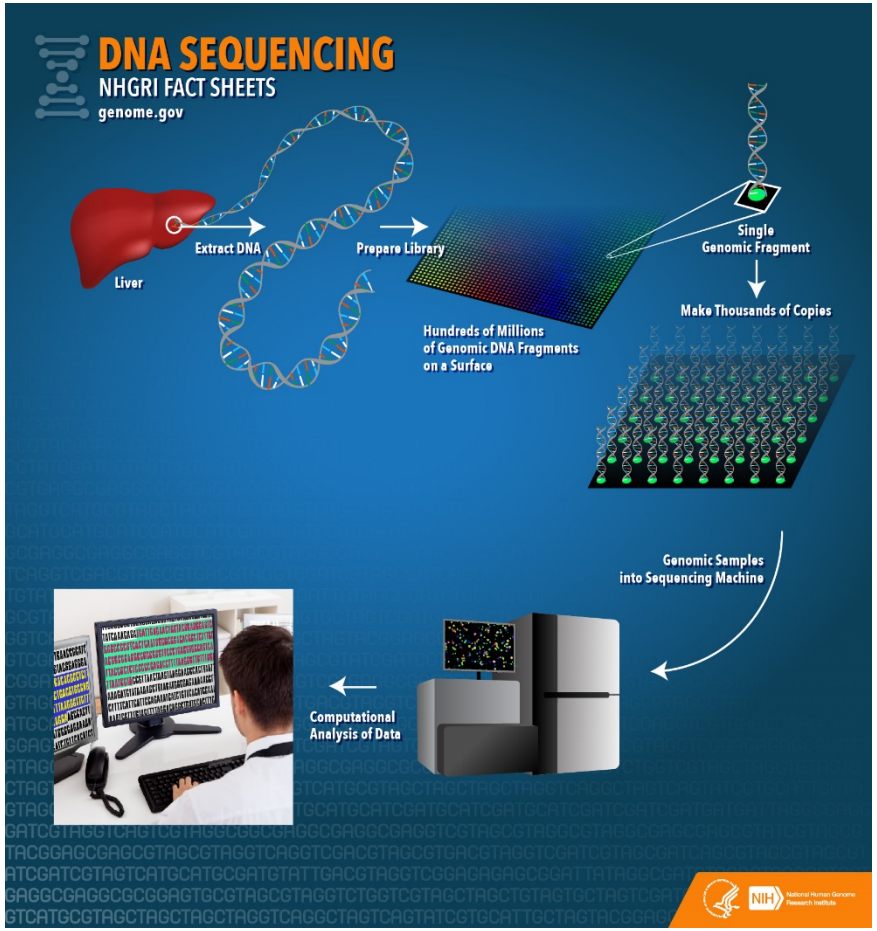
The most popular technique these days read short sequences (hundreds of characters), and align.

Each position is read typically at least 30 times to get enough confidence → Huge storage!!!

String alignment is then the key to final sequence → Need super-computer to do this fast.

A DNA sequence is compared against the reference genome. Only the difference (0.1%) need to be stored.

- This does not usually apply for MtDNA, as each cell has as many as 500 MtDNAs, they are slightly different! More different as we age.

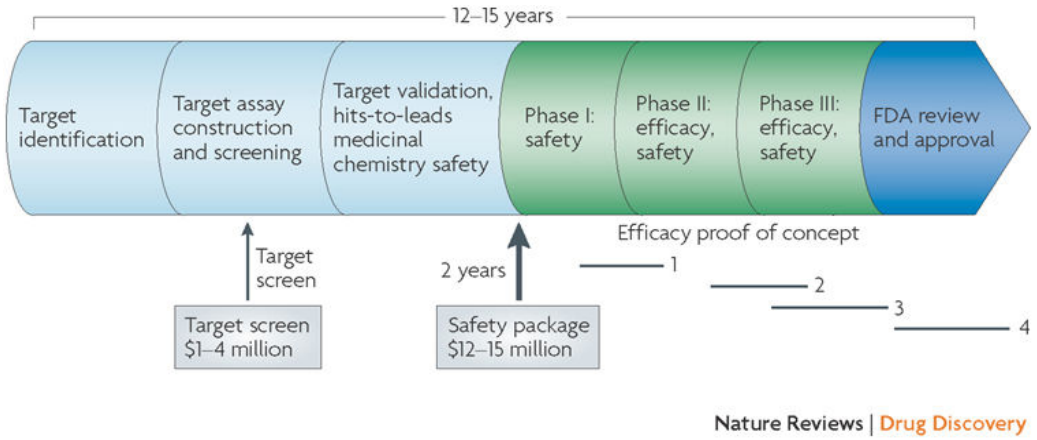


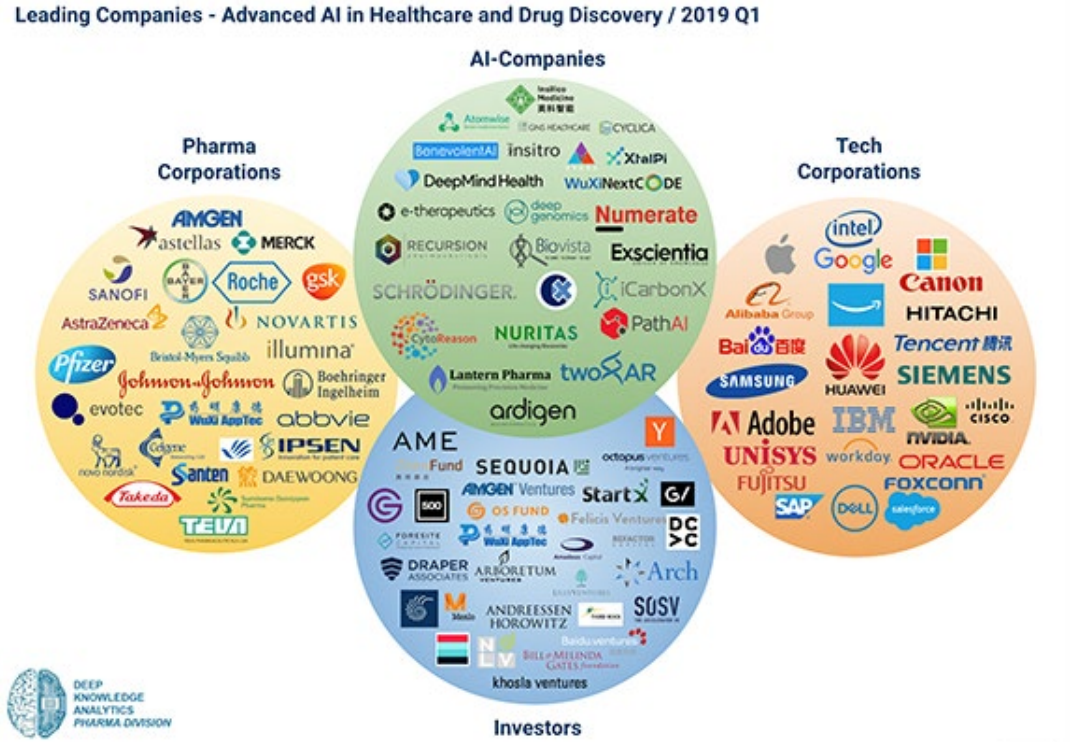
Source: https://www.genome.gov

The state of AI for drug design

This is not new. Since 1960s!

$500M - $2B





#REF: Roses, Allen D. "Pharmacogenetics in drug discovery and development: a translational perspective." *Nature reviews Drug discovery* 7.10 (2008): 807-817.

http://www.pharmexec.com/specialized-metrics-properly-assess-ai-pharma-startups

The three questions

Given a molecule, is this drug? Aka properties/targets/effects prediction.

- Druglikeness
- Targets it can modulate and how much
- Its dynamics/kinetics/effects if administered orally or via injection

Given a target, what are molecules?

- If the list of molecules is given, pick the good one. If evaluation is expensive, need to search, e.g., using BO.
- If no molecule is found, need to generate from scratch → generative models + BO, or RL.

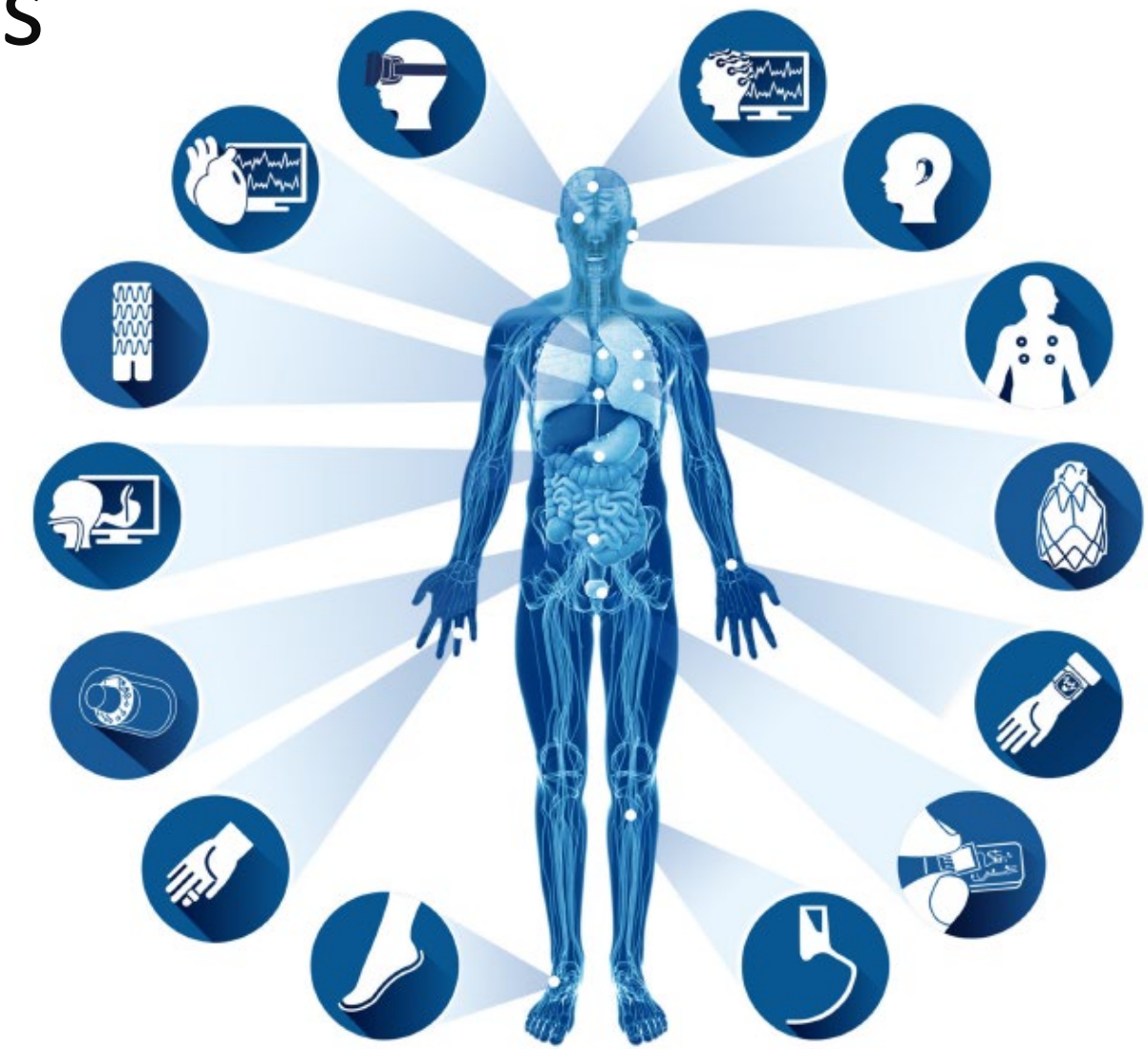
Given a molecular graph, what are the steps to make the molecule?

- Synthetic tractability
- Reaction planning, or retrosynthesis

Sensing technologies and data

Raw signals are ideal candidates for deep learning

Speech & vision techniques can be applied with minimal changes



#REF: Ravì, Daniele, et al. "Deep learning for health informatics." *IEEE journal of biomedical and health informatics* 21.1 (2017): 4-21.

Electronic medical records (EMR)

Need to model the healthcare processes, which are interactions of:

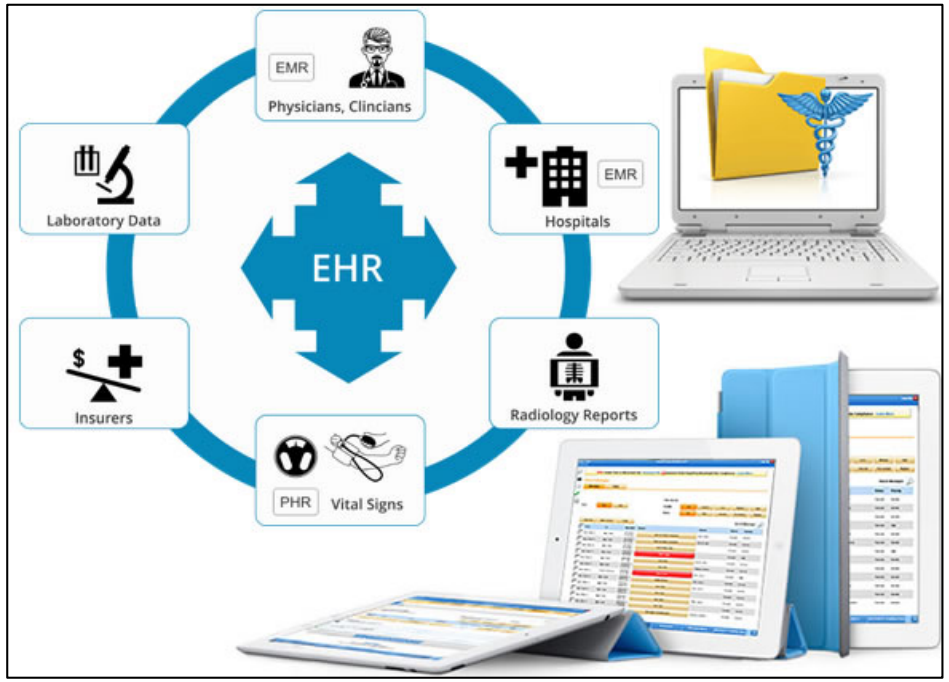
- Disease progression
- Interventions & care processes
- Recording processes (Electronic Medical/Health Records)

Irregular timing, event-based, sequence of (interacting) sets

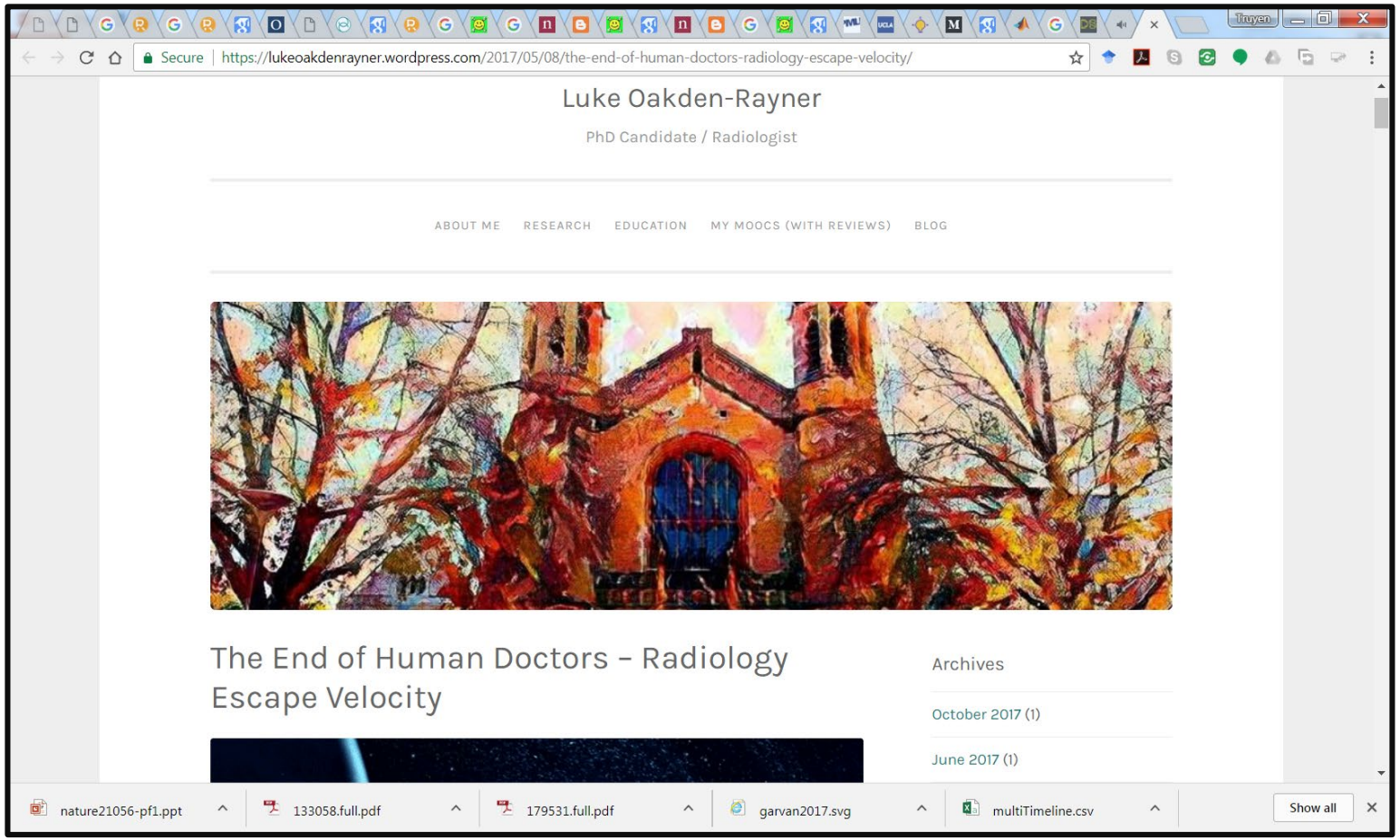
Multiple resolutions

Mixed modalities: biomarkers, code, text, social, wearables

Human-in-the-loop; negative/positive feedback



Source: medicalbillingcodings.org



“They should stop training radiologists now.”

Geoff Hinton (as of April 2017)

https://www.newyorker.com/magazine/2017/04/03/ai-versus-md

An art of modelling biomedicine: Analogy

Video as sequence of frame, but also a complex 3D graph of objects, actions and scenes

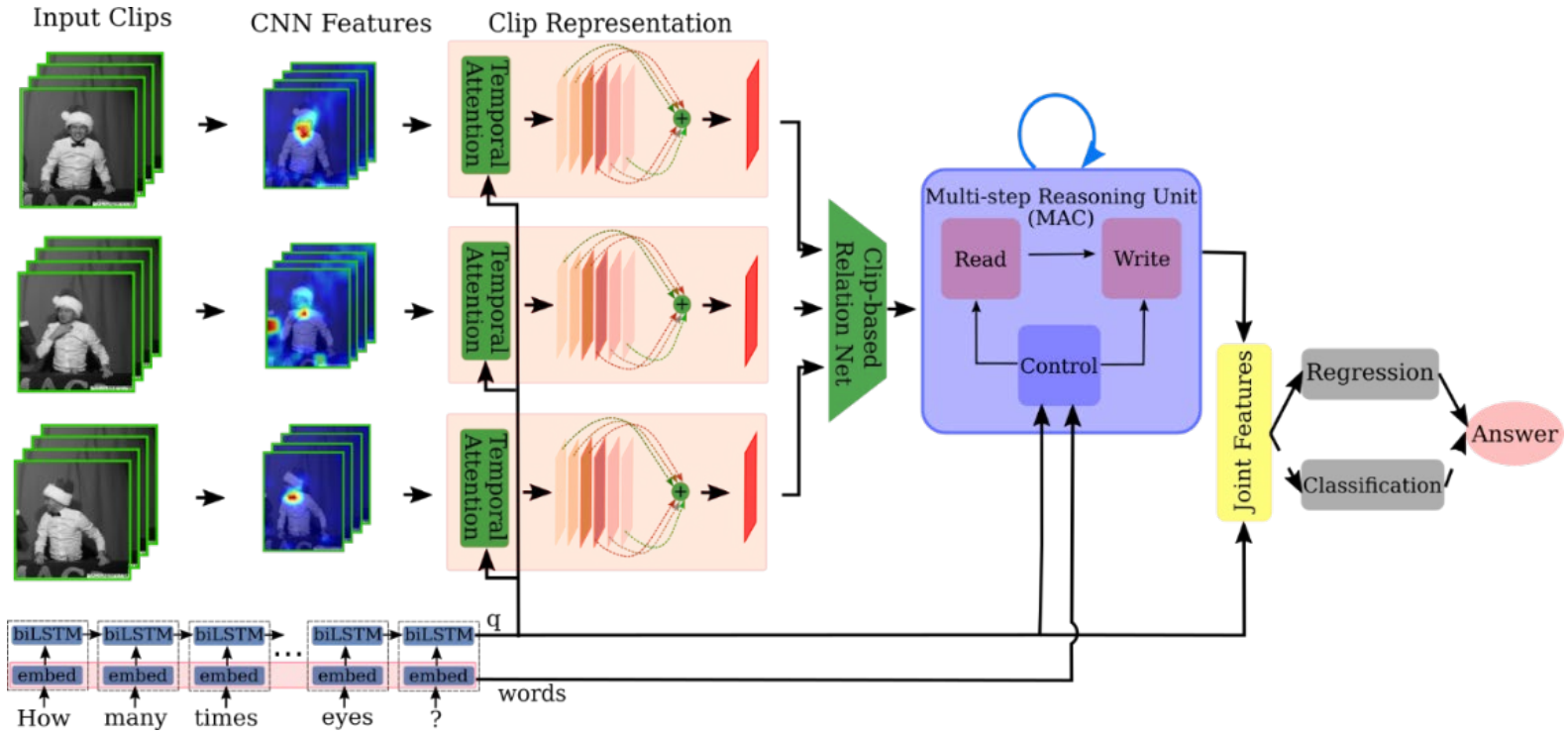
- → **Protein, RNA**

Question as sequence of words, but also a complex dependency graph of concepts

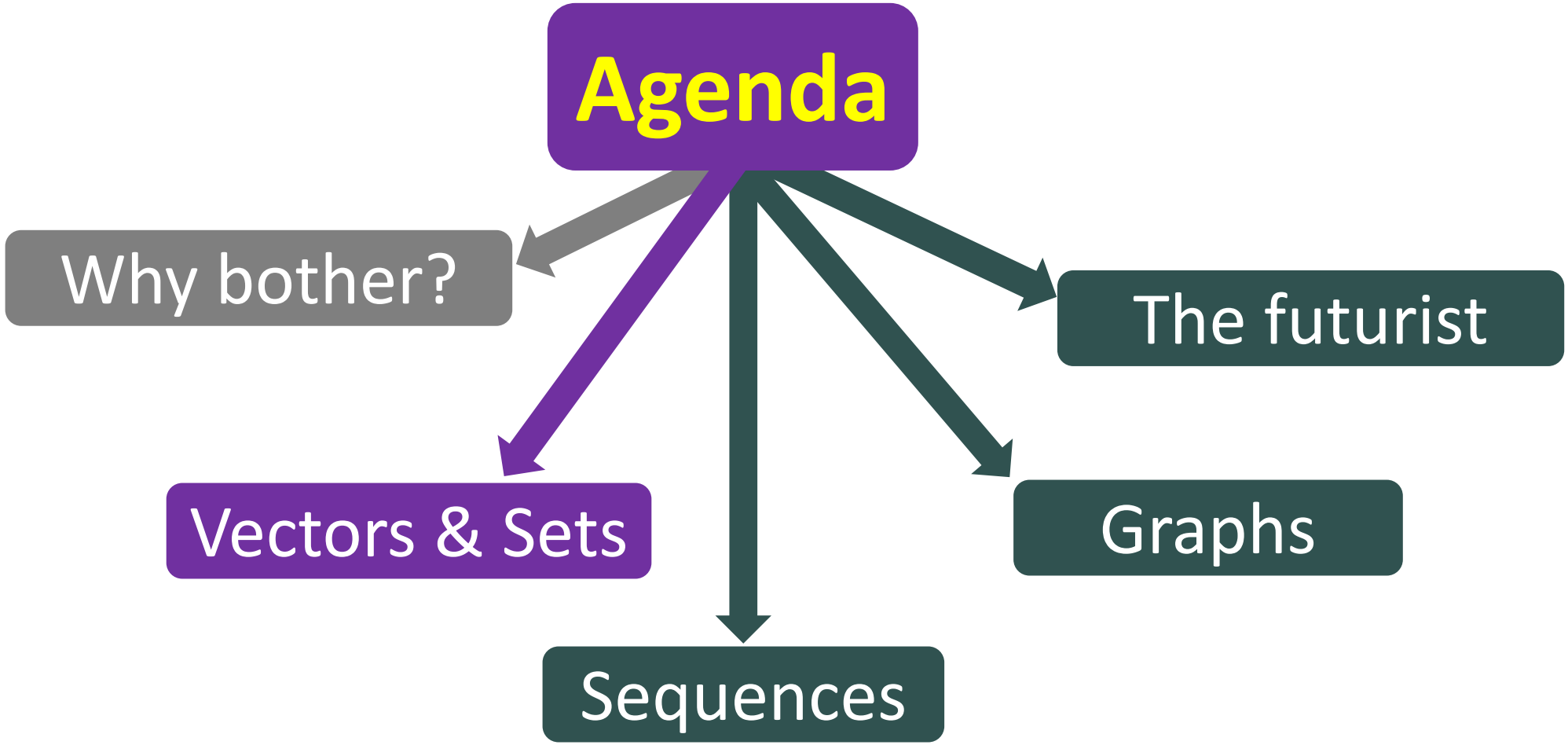
- → **Protein, drug**

Answer as facts (what and where) and deduced knowledge.

- → **Affinity, binding sites, modulation effect**

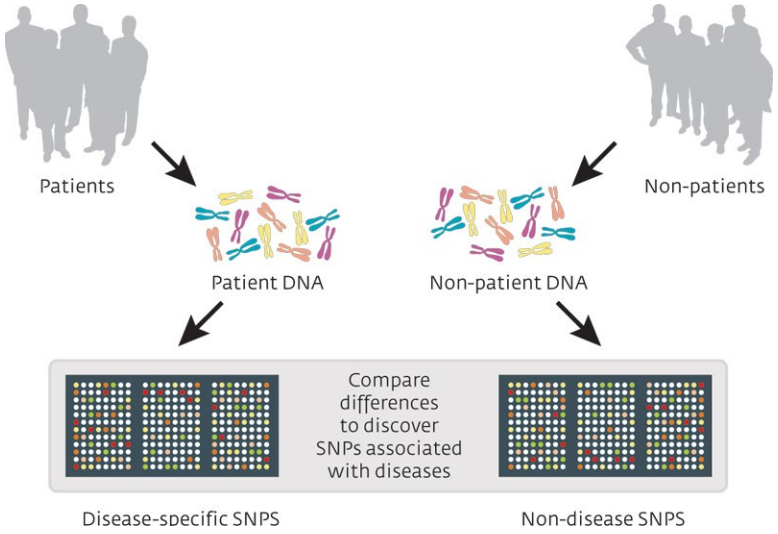


#Ref: Minh-Thao Le, Vuong Le, Truyen Tran, "Learning to Reason with Relational Video Representation for Question Answering", *In preparation 2019*.


Agenda
Why bother?
Vectors & Sets
Sequences
Graphs
The futurist

“Diet networks” for GWAS

#REF: Romero, Adriana, et al. "Diet Networks: Thin Parameters for Fat Genomic" *ICLR* (2017).



GWAS = Genome Wide Association Study

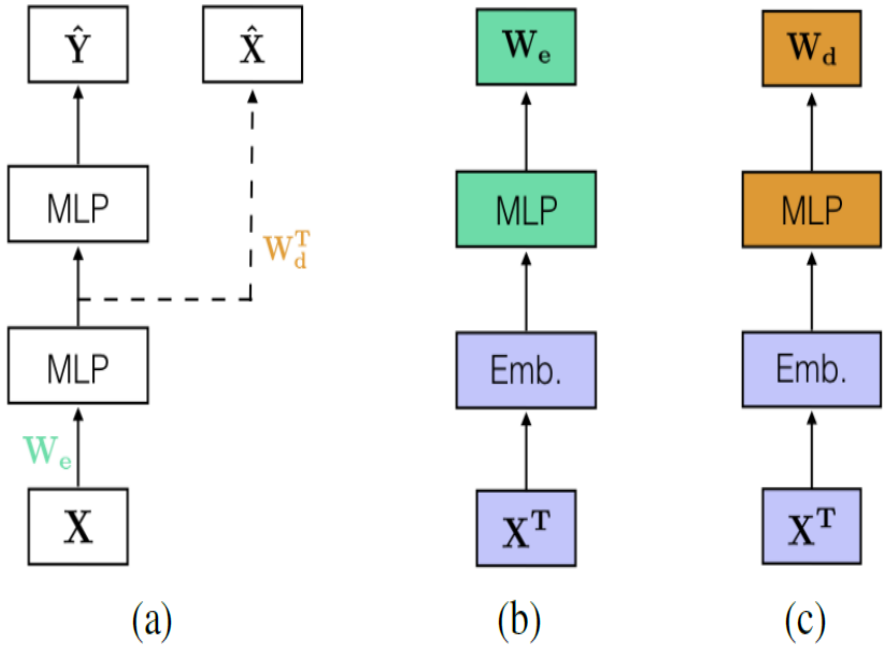
Diet Net uses a “hypernet” to generate the main net.

Features are embedded (not data instance).

Unsupervised autoencoder as regularizer.

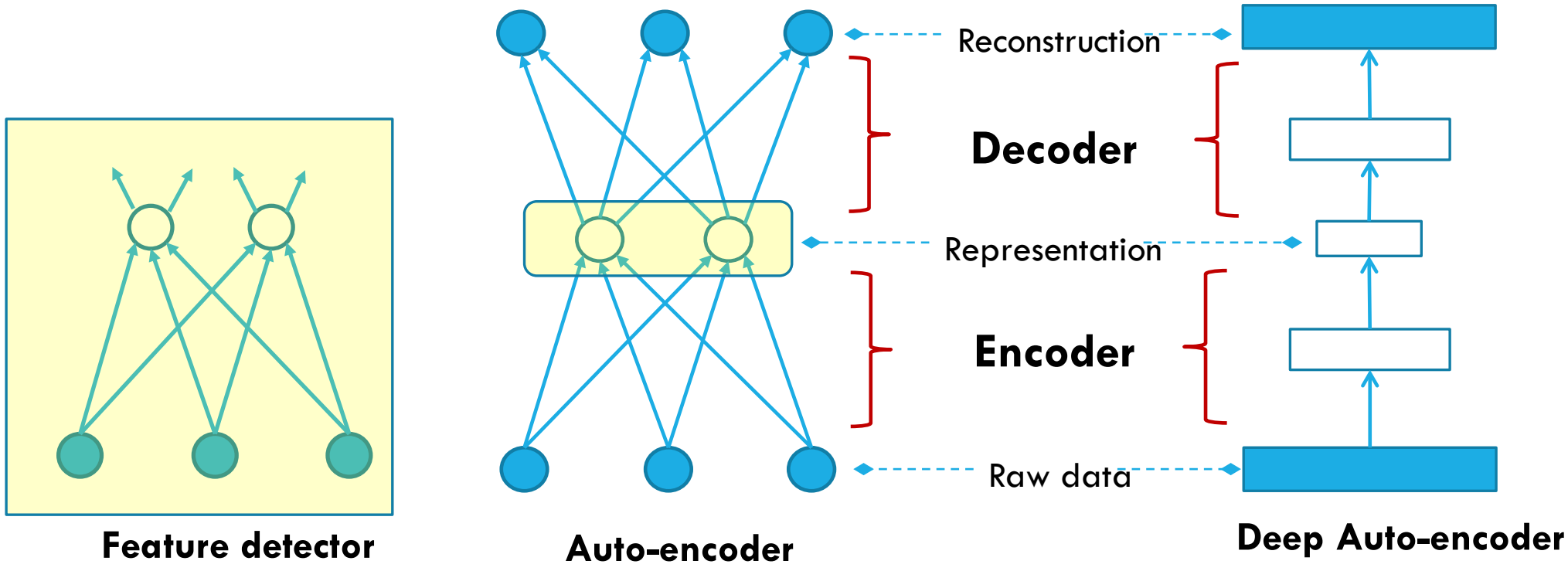
Works well on country prediction on the 1000 Genomes Project dataset.

- But this is a relatively easy problem. PCA, even random subspace can do quite well!



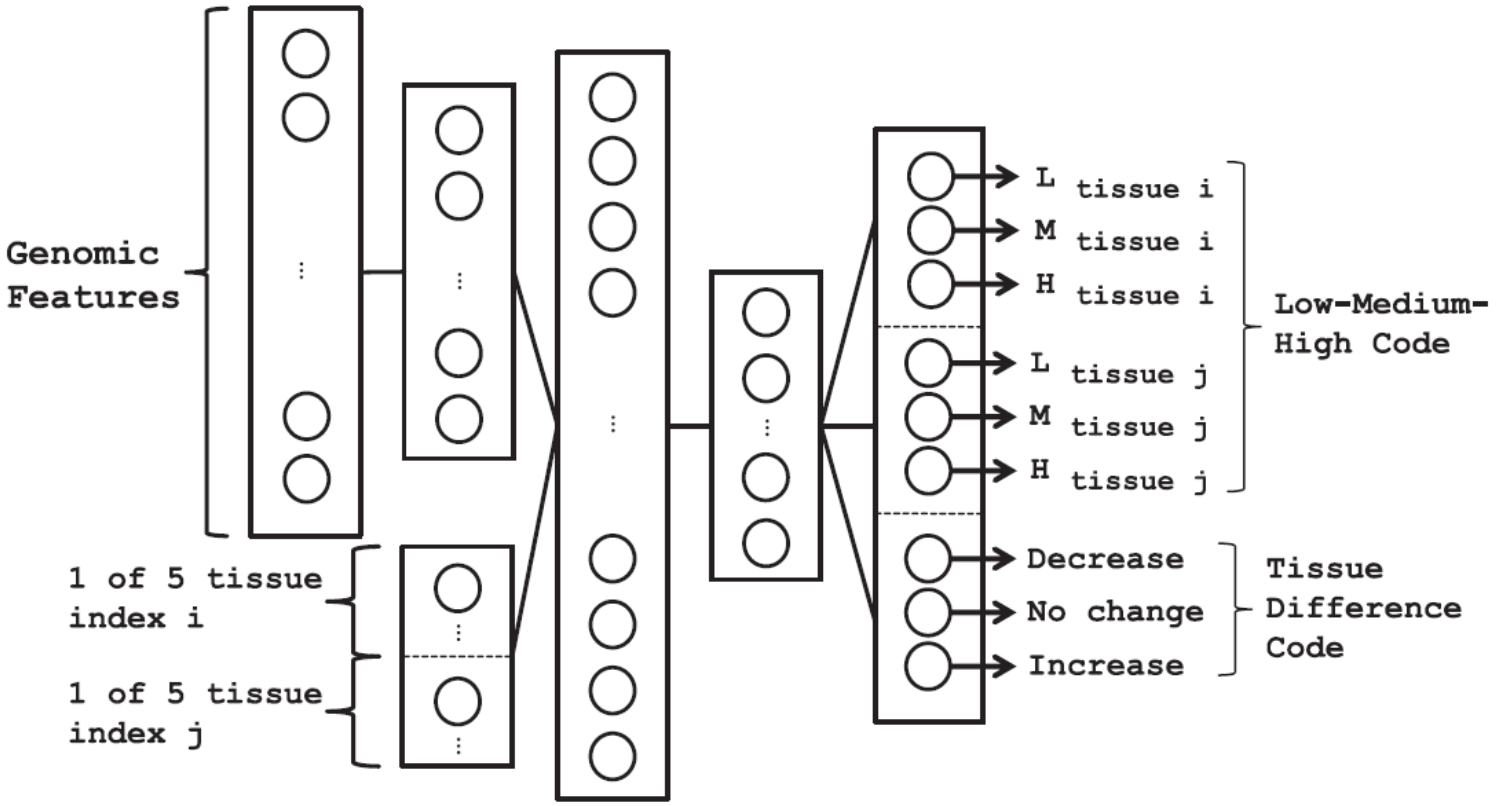
Images taken from the paper

DeepPatient: Representing medical records with Stacked Denoising Autoencoder



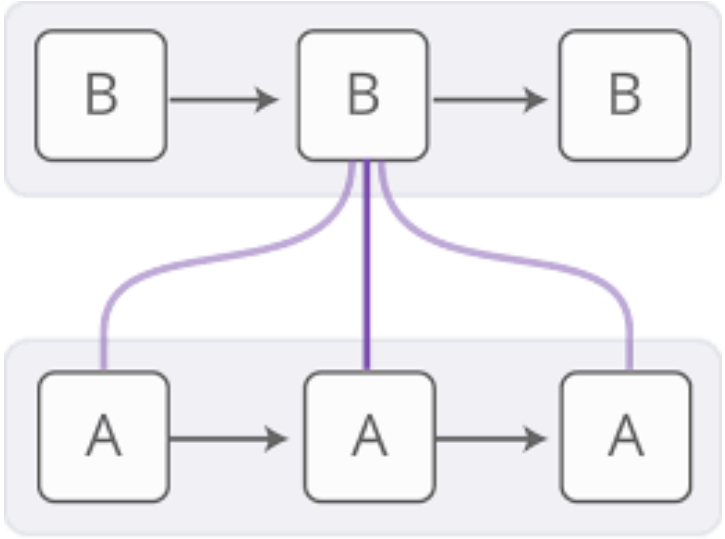
#Ref: Miotto, Riccardo, et al. "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records." *Scientific reports* 6 (2016): 26094.

Use of feedforward nets: Tissue-regulated splicing code



#REF: Leung, Michael KK, et al. "Deep learning of the tissue-regulated splicing code." *Bioinformatics* 30.12 (2014): i121-i129.

Operation on Gen Set



http://distill.pub/2016/augmented-rnns/

Attention mechanism

DeepTRIAGE: Interpretable and Individualised Biomarker Scores using Attention Mechanism for the Classification of Breast Cancer Sub-types

Adham Beykikhoshk[1,*], Thomas P. Quinn[1,*], Samuel C. Lee[1], Truyen Tran[1], and Svetha Venkatesh[1]

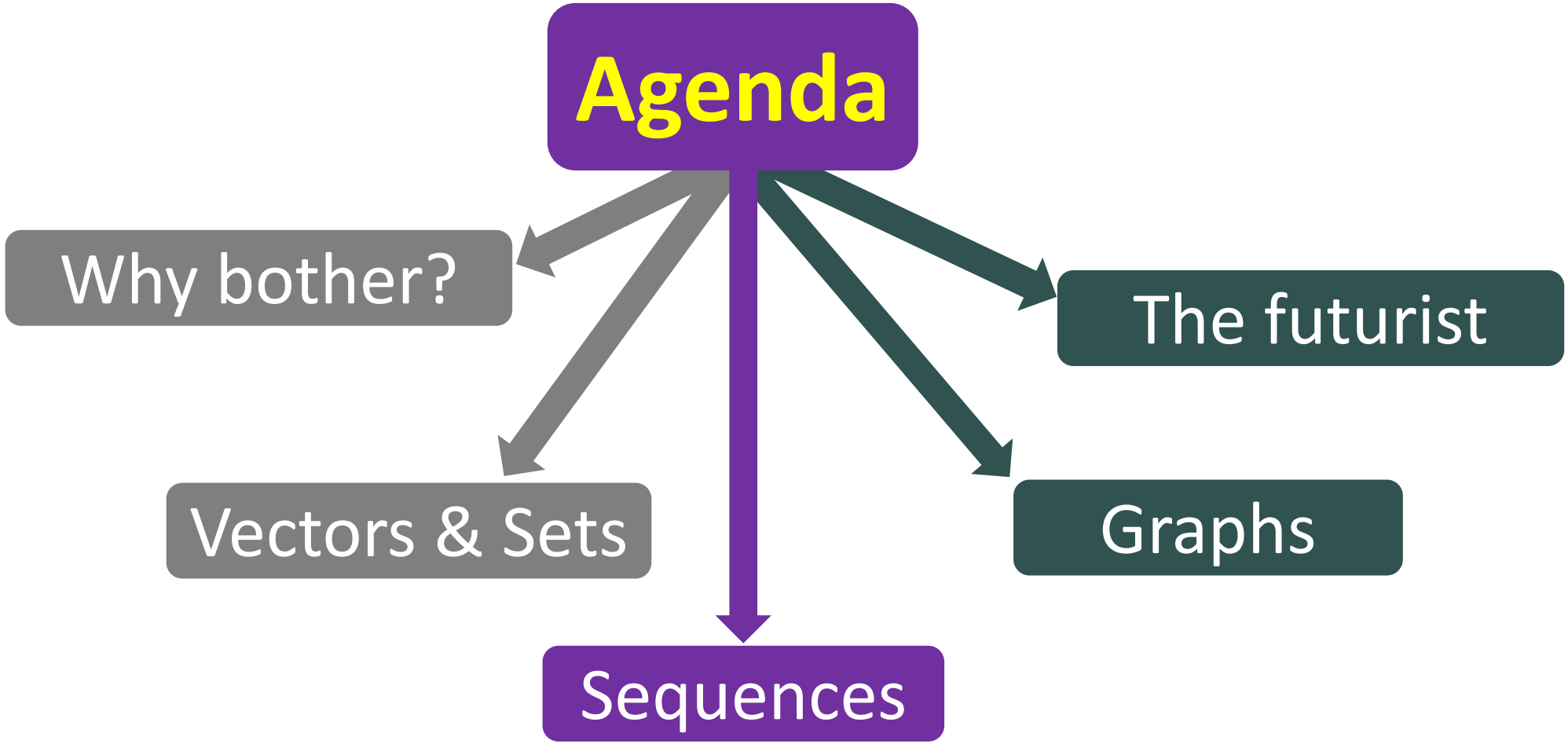
[1]Centre for Pattern Recognition and Data Analytics, Deakin University, Geelong, Australia
* adham.beyki@deakin.edu.au; contacttomquinn@gmail.com

Abstract

Motivation: Breast cancer is a collection of multiple tissue pathologies, each with a distinct molecular signature that correlates with patient prognosis and response to therapy. Accurately differentiating between breast cancer sub-types is an important part of clinical decision-making. Already, this problem has been addressed using machine learning methods that separate tissue samples into distinct groups. However, there remains unexplained heterogeneity within the established sub-types that cannot be resolved by the commonly used classification algorithms. In this paper, we propose a novel deep learning architecture, called DeepTRIAGE (Deep learning for the TRactable Individualised Analysis of Gene Expression), which not only classifies cancer sub-types with comparable accuracy, but simultaneously assigns each patient their own set of interpretable and individualised biomarker scores. These personalised scores describe how important each feature is in the classification of each patient, and can be analysed post-hoc to generate new hypotheses about intra-class heterogeneity.

Results: We apply the DeepTRIAGE framework to classify the gene expression signatures of luminal A and luminal B breast cancer sub-types, and illustrate its use for genes and gene set (i.e., GO and KEGG) features. Using DeepTRIAGE, we find that the GINS1 gene and the kinetochore organisation GO term are the most important features for luminal sub-type classification. Through classification, DeepTRIAGE simultaneously reveals heterogeneity within the luminal A biomarker scores that significantly associate with tumour stage, placing all luminal samples along a continuum of severity.

Availability and implementation: The proposed model is implemented in Python using PyTorch framework. The analysis is done in Python and R. All Methods and models are freely available from https://github.com/adham/BiomarkerAttend.


Agenda
Why bother?
Vectors & Sets
Sequences
Graphs
The futurist

Deep architectures for nanopore sequencing

Aimed at real time recognition

The setting is similar to speech recognition!

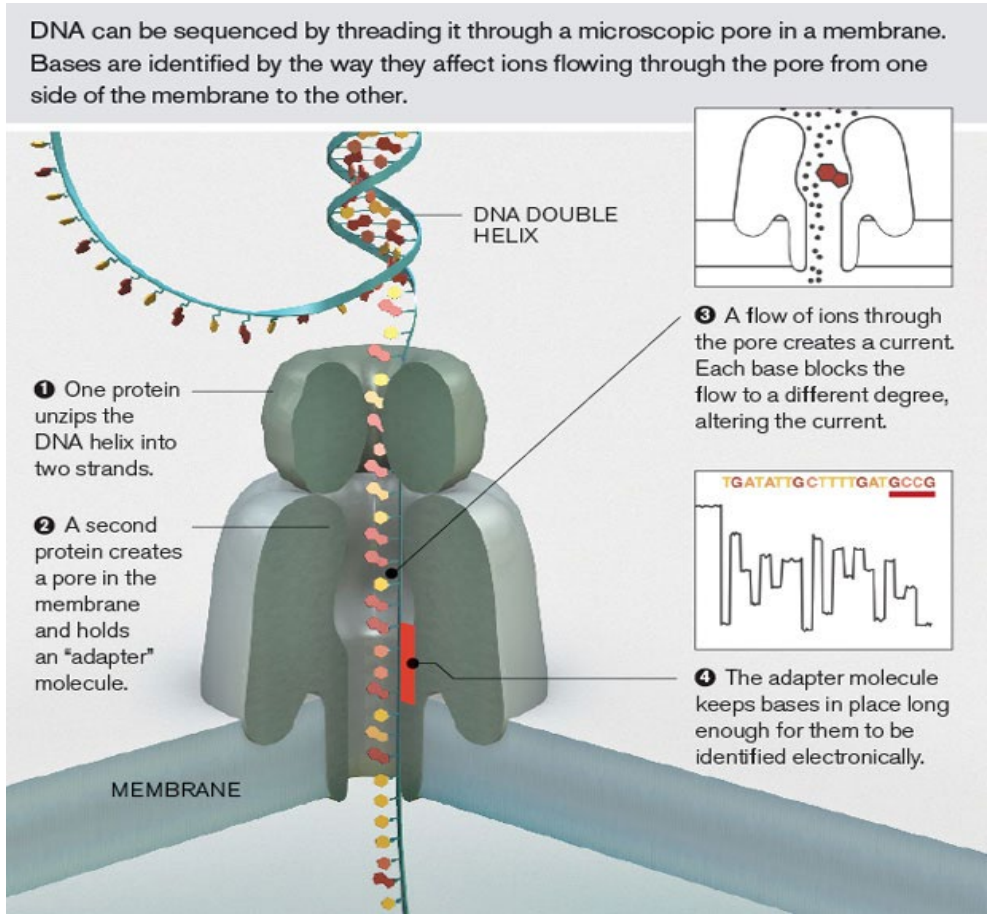
- → The early days used HMMs. Now LSTMs.

We will briefly review the latest:

- **Chiron** (Teng et al., May 2018, UQ, Australia)

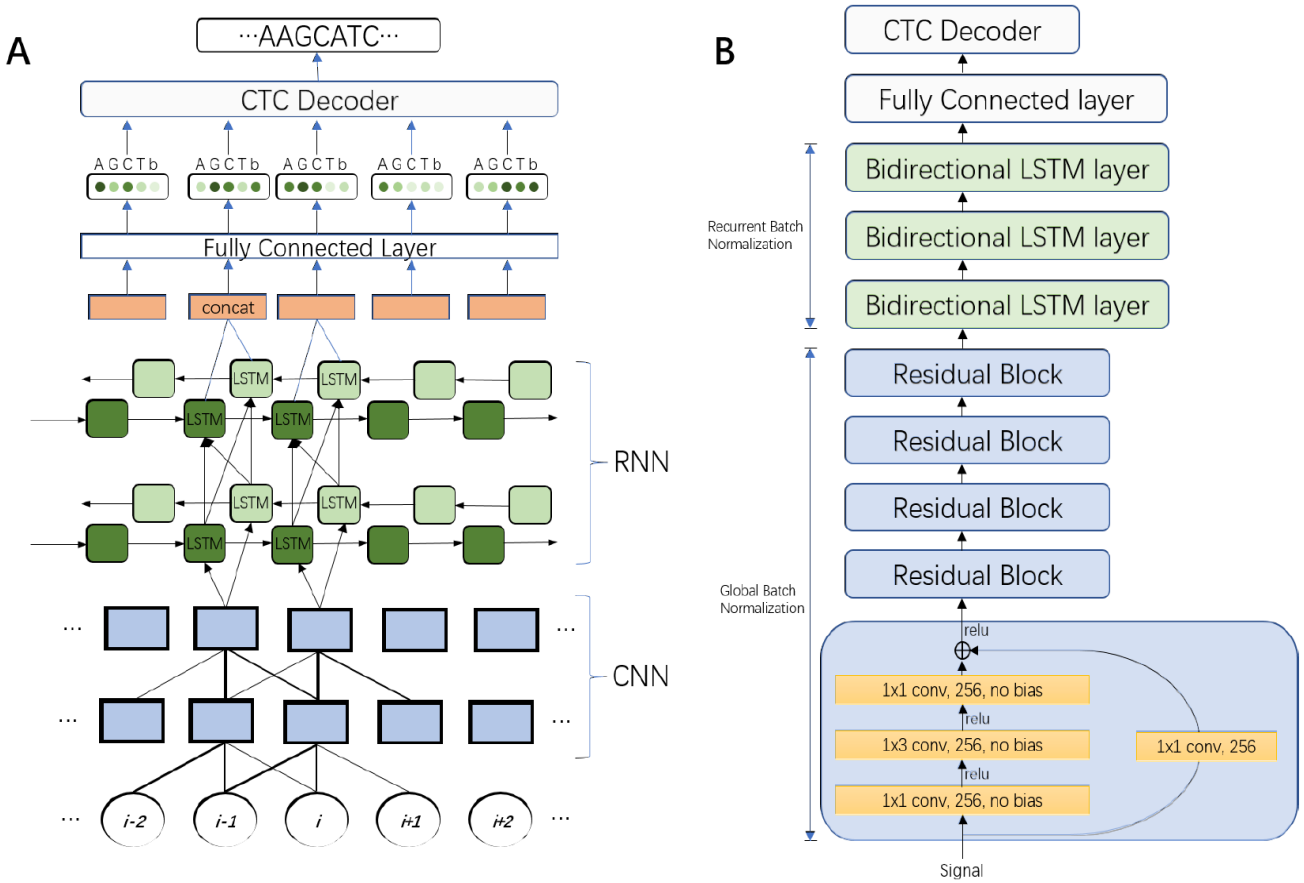
Other GRU/LSTM variants

- Nanonet (Oxford Nanopore Technologies, 2016)
- BasecRAWller (Stoiber & Brown, May 2017)
- **DeepNano** (Boza et al., June 2017, Comenius University in Bratislava, Slovakia)



Source: technologyreview.com

Chiron

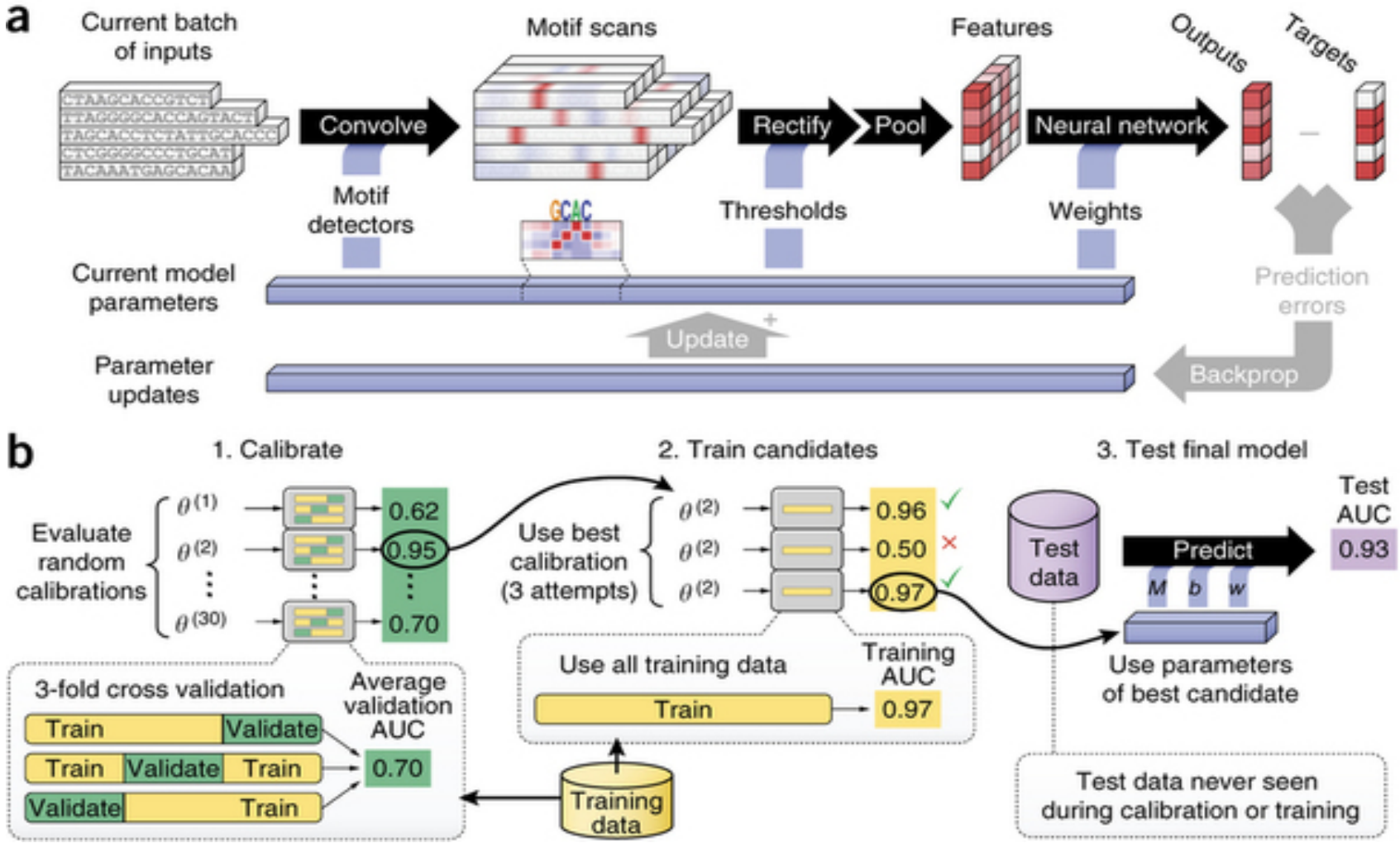


Dataset	Basecaller	**Identity Rate**
Lambda	Metrichor	0.8650 (-0.0246)
	Albacore	**0.8896**
	BasecRAWller	0.8154 (-0.0742)
	Chiron	0.8776 (-0.012)
E. coli	Metrichor	0.8864 (-0.0193)
	Albacore	0.901 (-0.0047)
	BasecRAWller	0.8254 (-0.0803)
	Chiron	**0.9057**
M. tuberculosis	Metrichor	0.8802 (-0.0117)
	Albacore	**0.8919**
	BasecRAWller	0.8241 (-0.0678)
	Chiron	0.8851 (-0.0068)
Human	Metrichor	0.794 (-0.0446)
	Albacore	**0.8386**
	BasecRAWller	0.8149 (-0.0237)
	Chiron	0.8154 (-0.0232)

#REF: Teng, Haotien, et al. "Chiron: Translating nanopore raw signal directly into nucleotide sequence using deep learning", GigaScience, Volume 7, Issue 5, 1 May 2018, giy037.

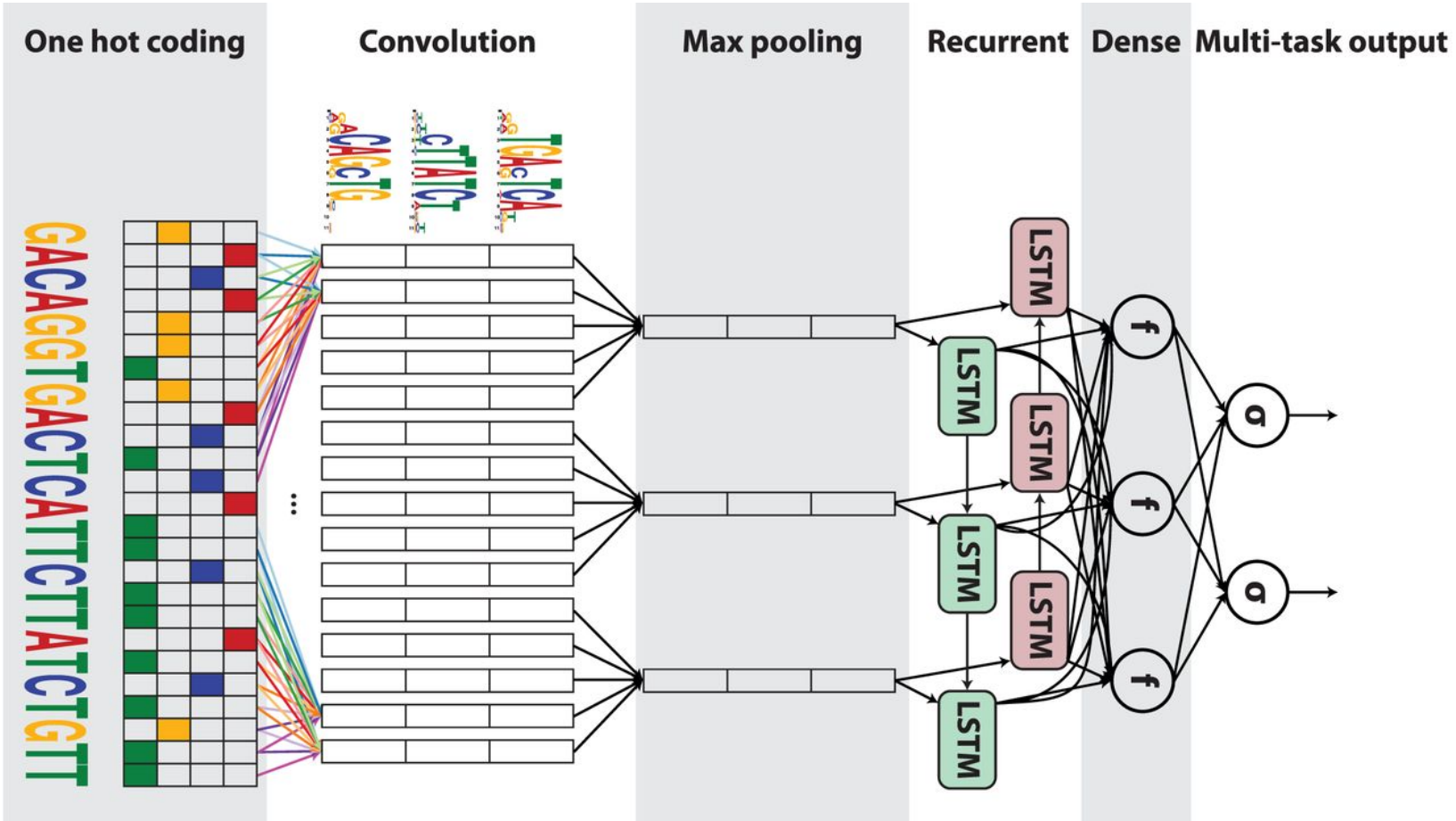
Identifying binding sites

DeepBind (Alipanahi et al, Nature Biotech 2015)



http://www.nature.com/nbt/journal/v33/n8/full/nbt.3300.html

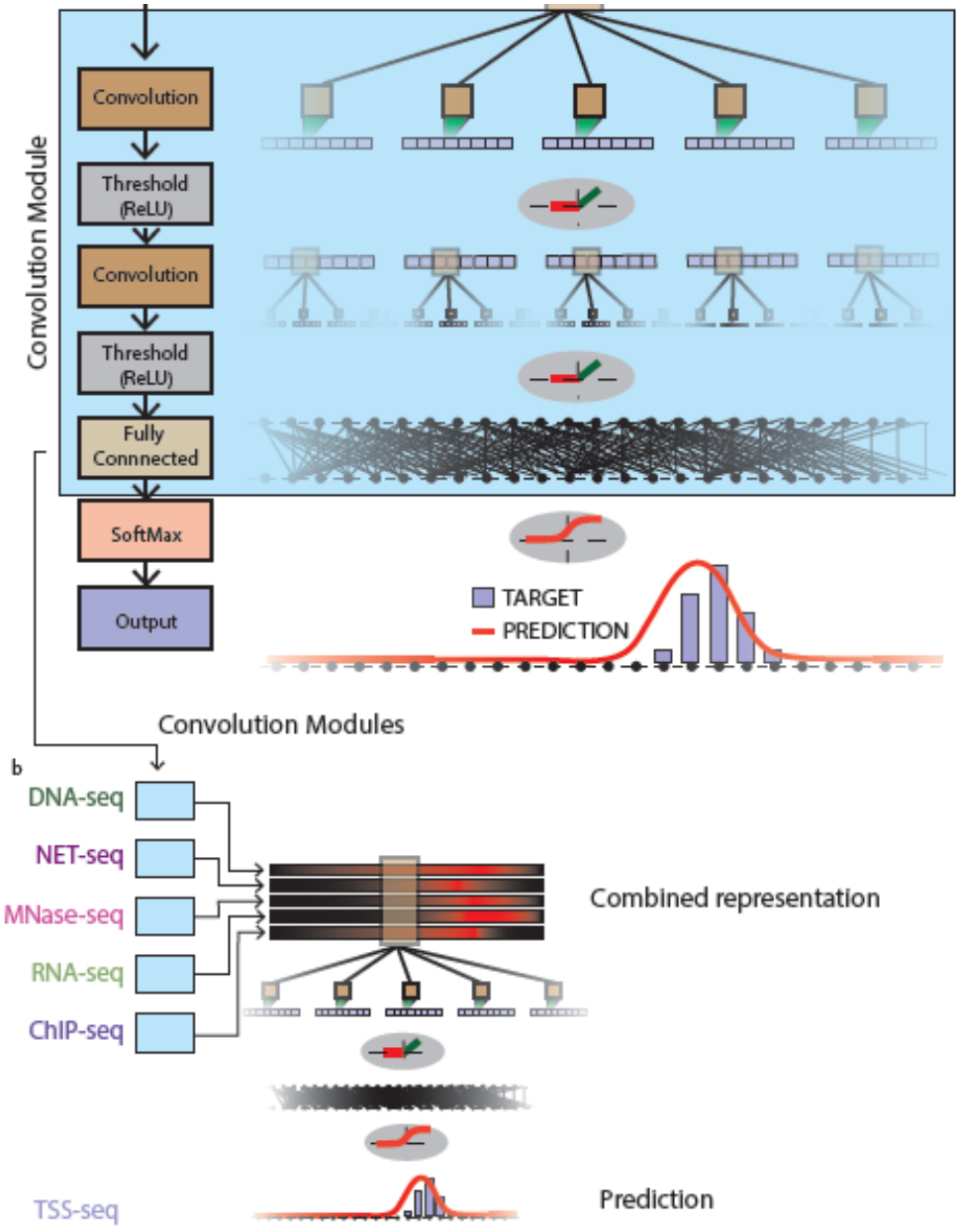
User of CNN+RNNs: DanQ

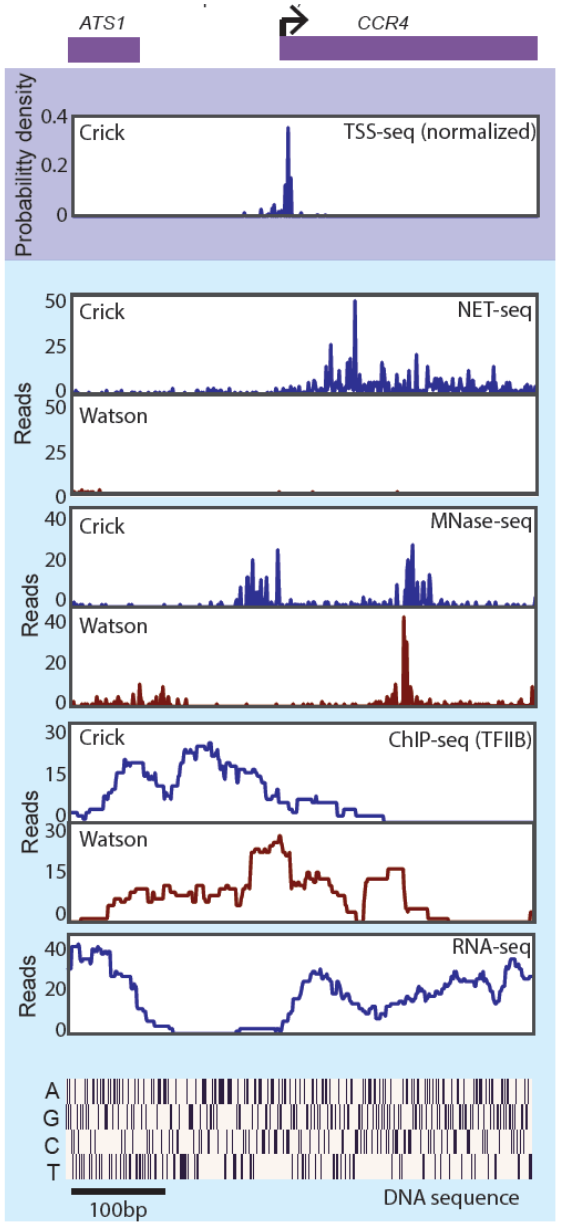


#REF: Quang, Daniel, and Xiaohui Xie. "DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences." *Nucleic acids research* 44.11 (2016): e107-e107.

Multiple modalities

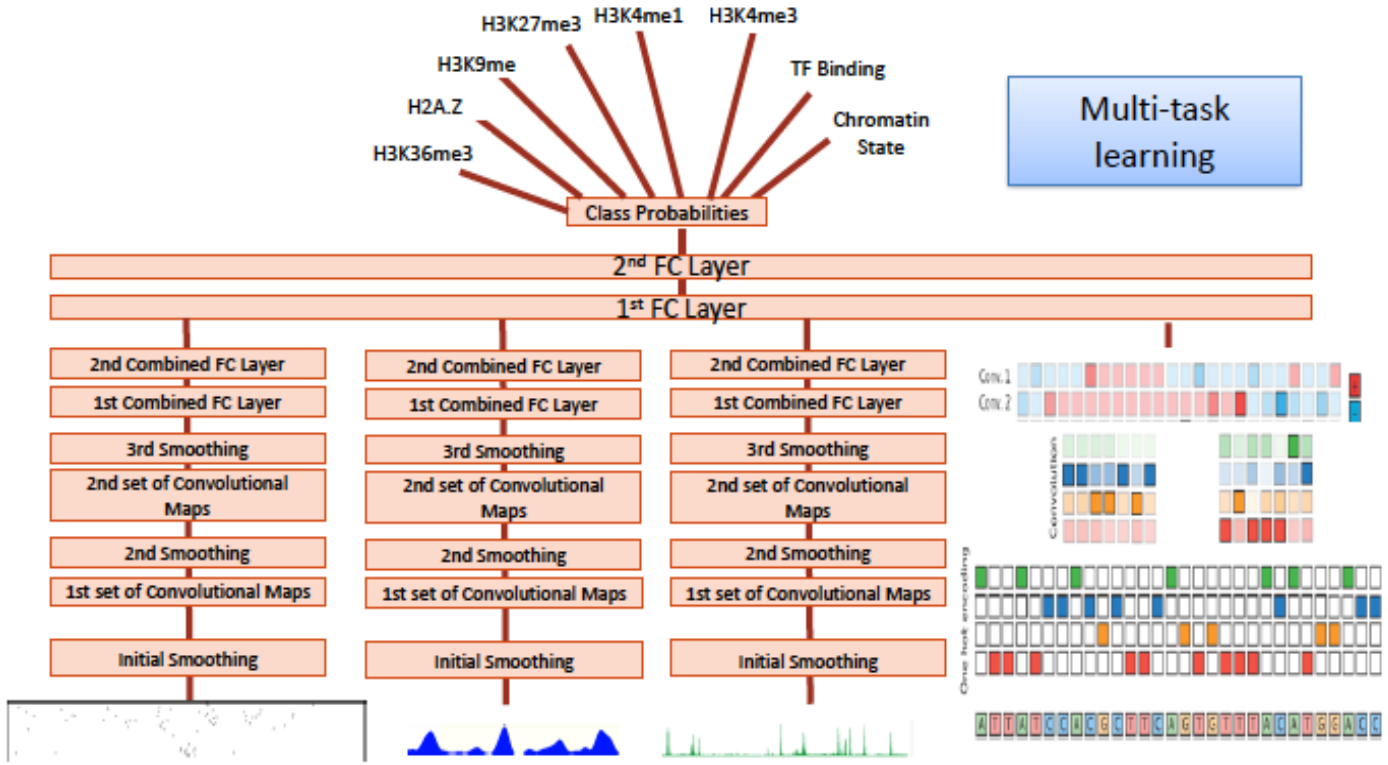
#REF: Eser, Umut, and L. Stirling Churchman. "FIDDLE: An integrative deep learning framework for functional genomic data inference." *bioRxiv* (2016): 081380.





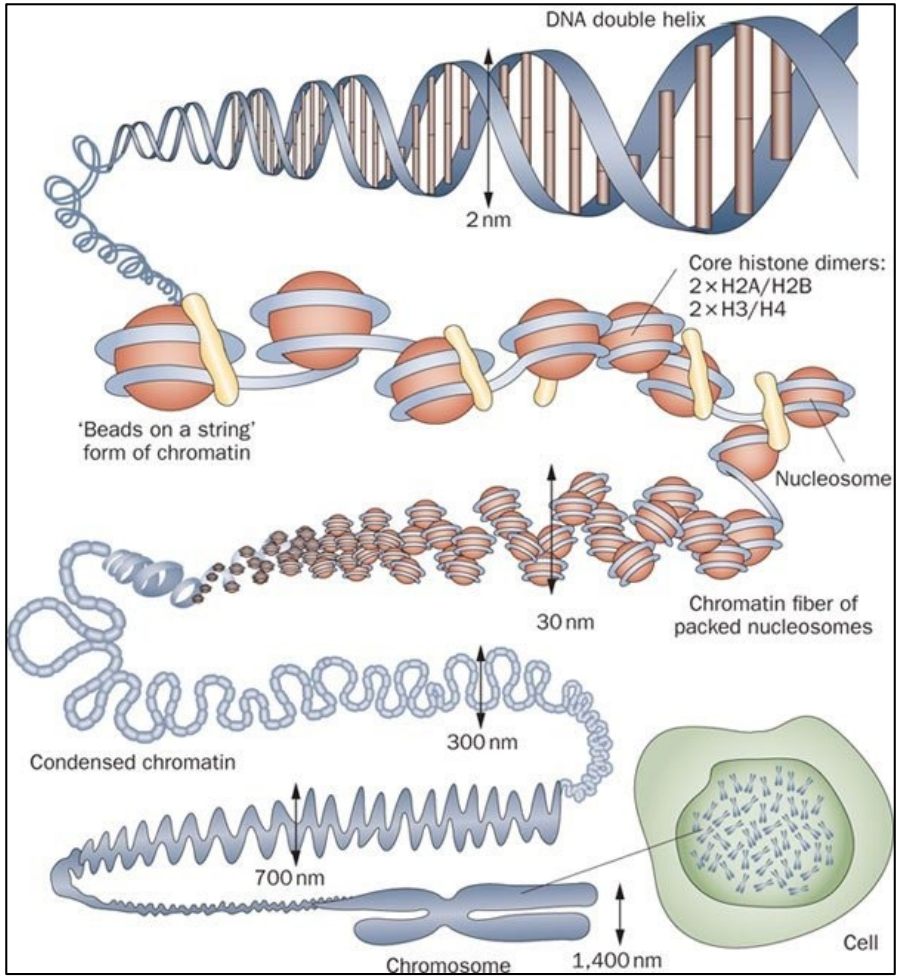
THE CHROMPUTER

Integrating multiple inputs (1D, 2D signals, sequence) to simulatenously **predict multiple outputs**



Source: https://simons.berkeley.edu/sites/default/files/docs/4575/2016-kundaje-simonsinstitute-deeplearning.pdf

Chromatins



https://qph.ec.quoracdn.net

More models/frameworks

DragoNN

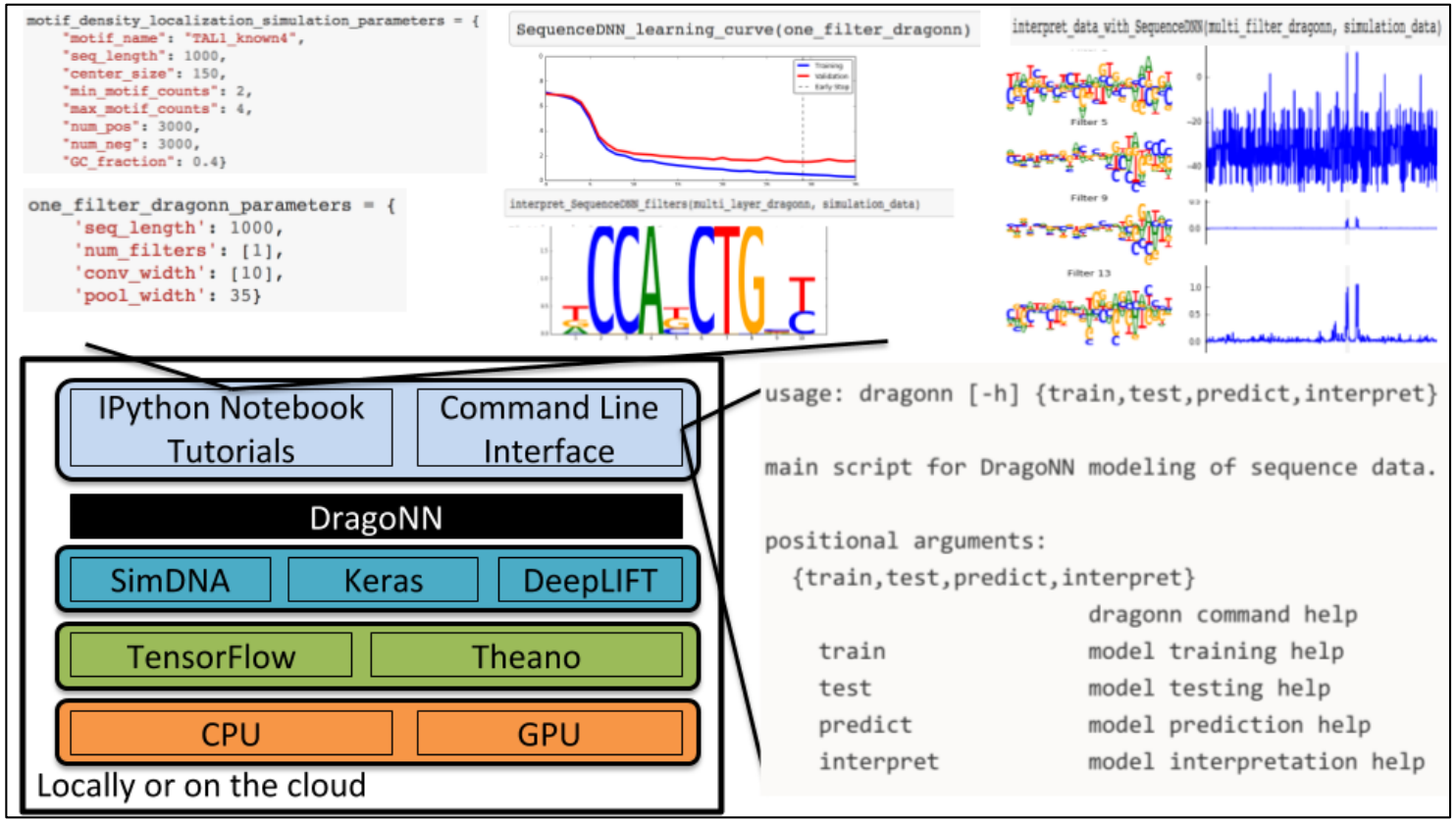
DeepChrome

DeepSEA

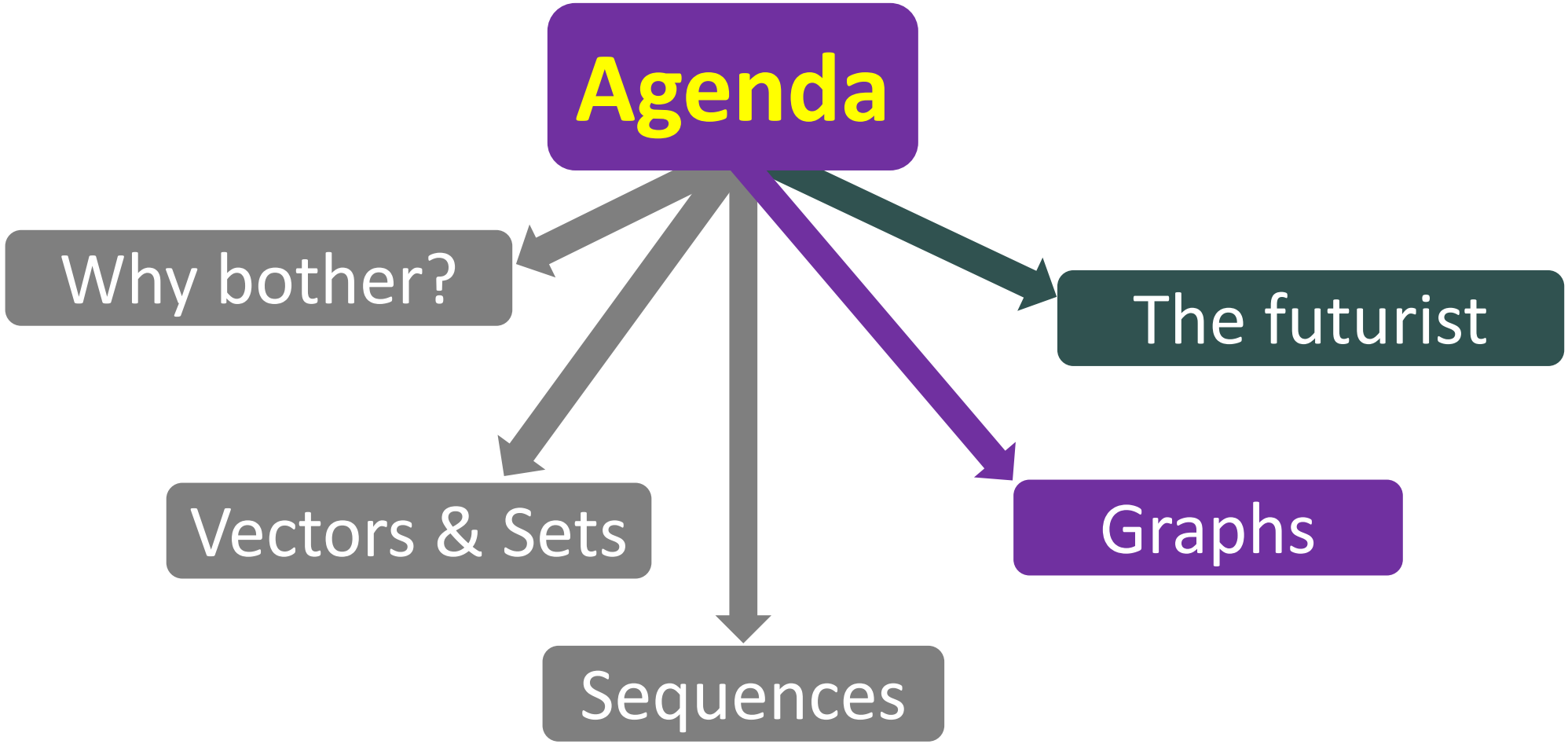
Basset

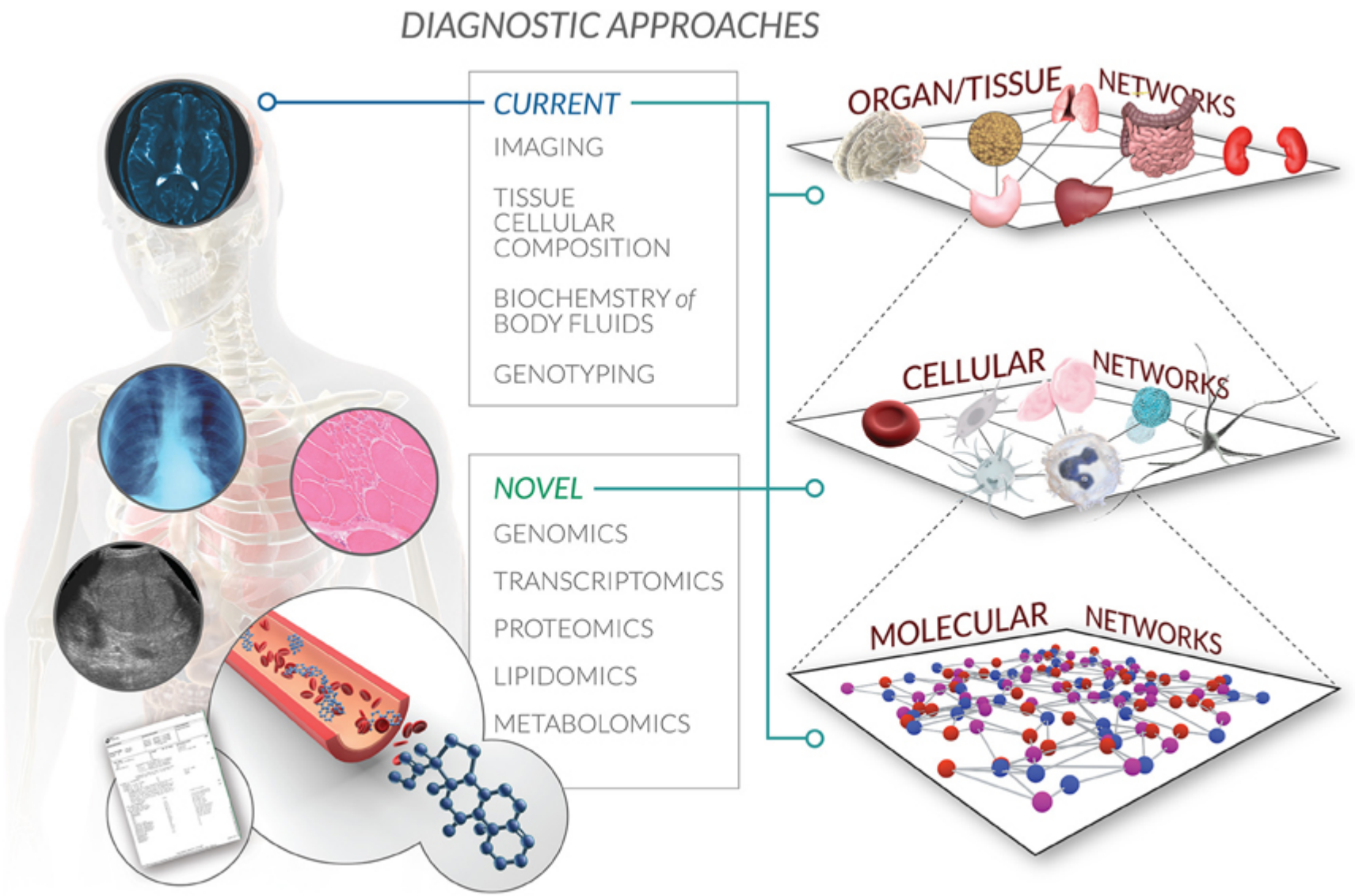
DeepBound

…



http://kundajelab.github.io/dragonn


Agenda
Why bother?
Vectors & Sets
Sequences
Graphs
The futurist



System medicine

https://www.frontiersin.org/articles/10.3389/fphys.2015.00225/full

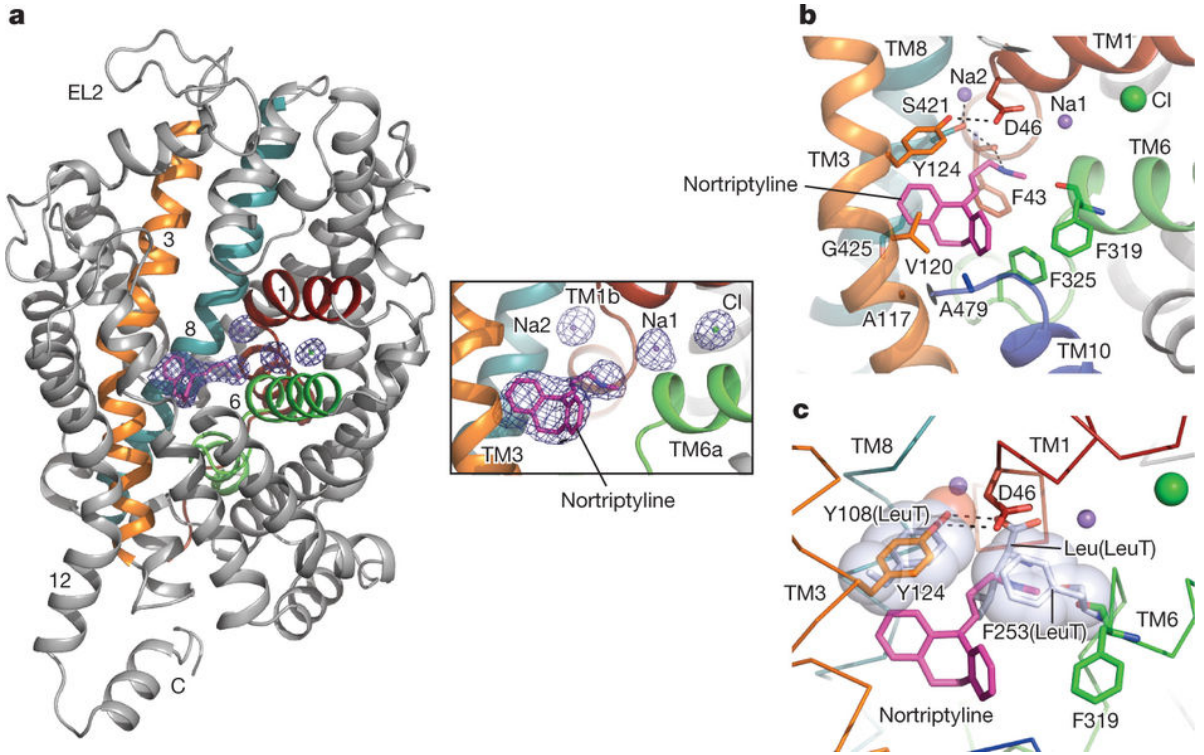
Biology & pharmacy

Traditional techniques:

- Graph kernels (ML)
- Molecular fingerprints (Chemistry)

Modern techniques

- Molecule as graph: atoms as nodes, chemical bonds as edges

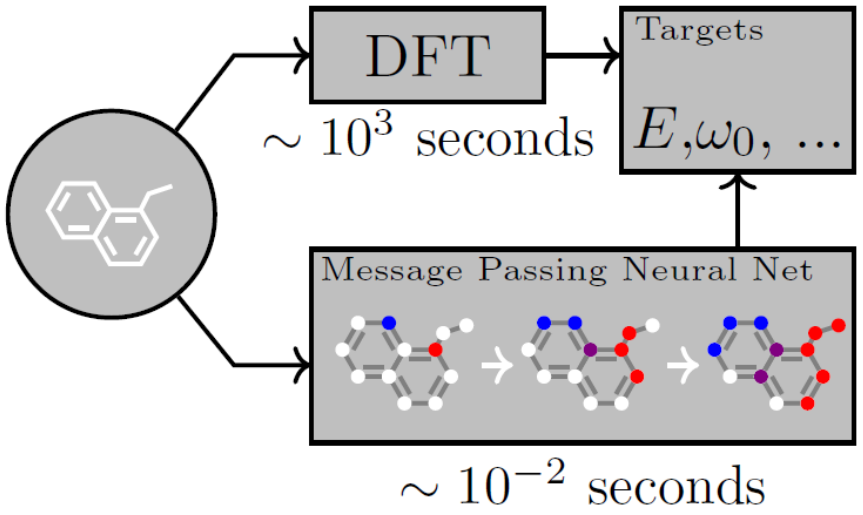


#REF: Penmatsa, Aravind, Kevin H. Wang, and Eric Gouaux. "X-ray structure of dopamine transporter elucidates antidepressant mechanism." *Nature* 503.7474 (2013): 85-90.

Chemistry

DFT = Density Functional Theory

Gilmer, Justin, et al. "Neural message passing for qu
chemistry." *arXiv preprint arXiv:1704.01212* (2017)



- Molecular properties
- Chemical-chemical interaction
- Chemical reaction
- Synthesis planning

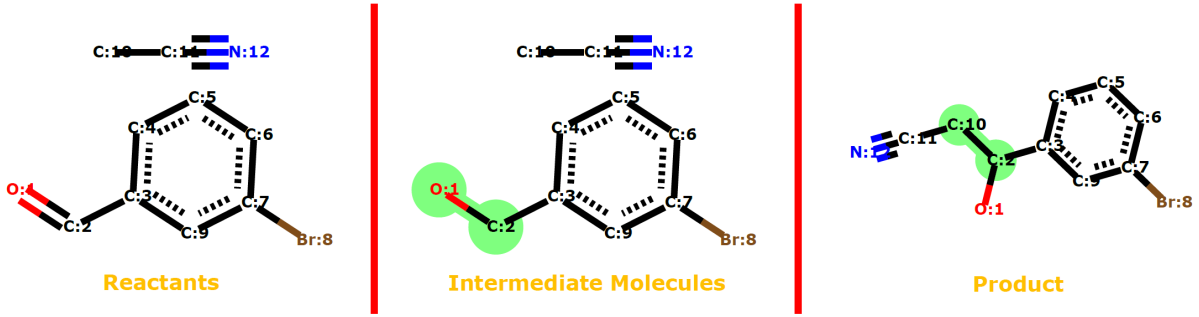
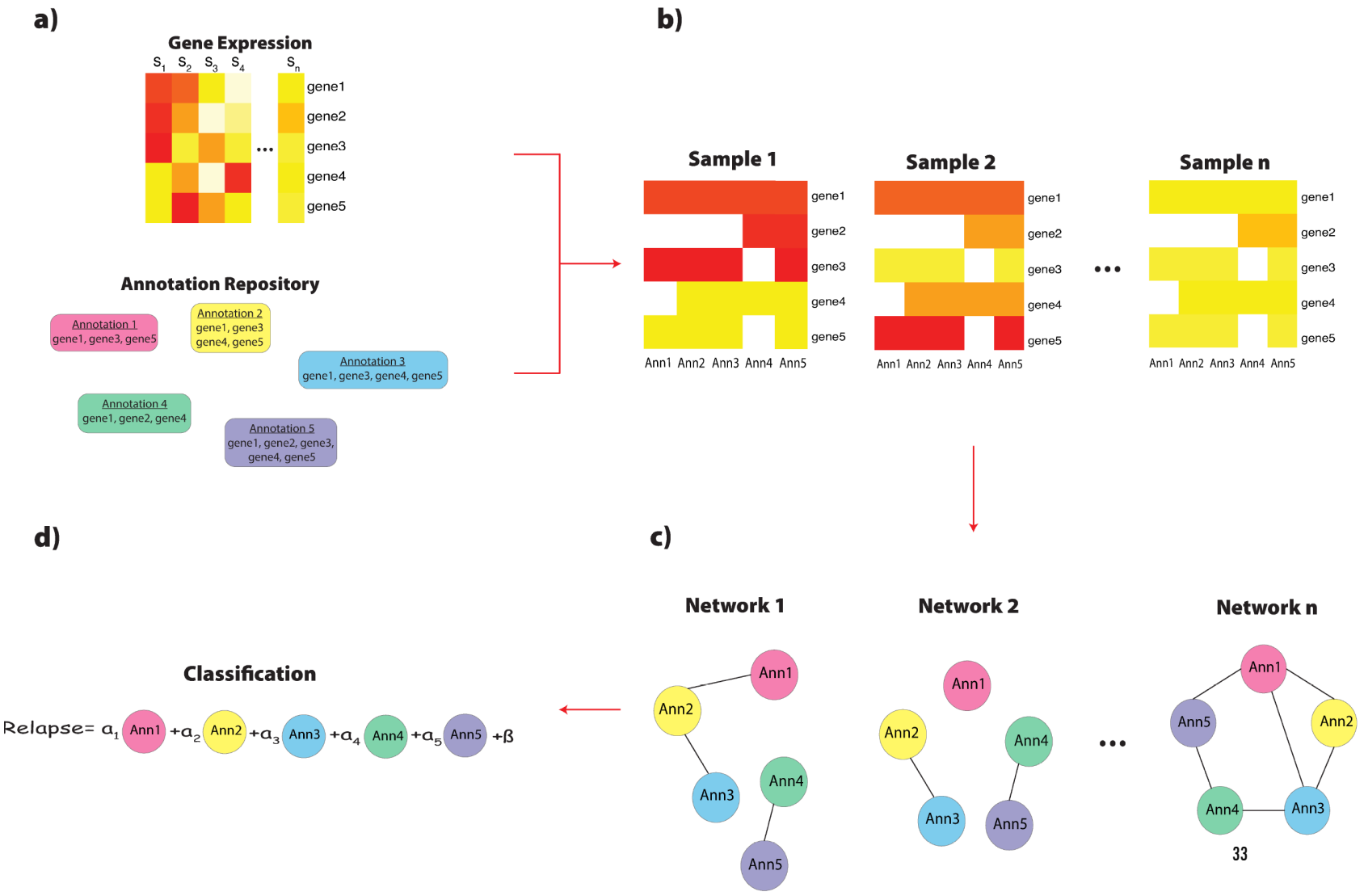


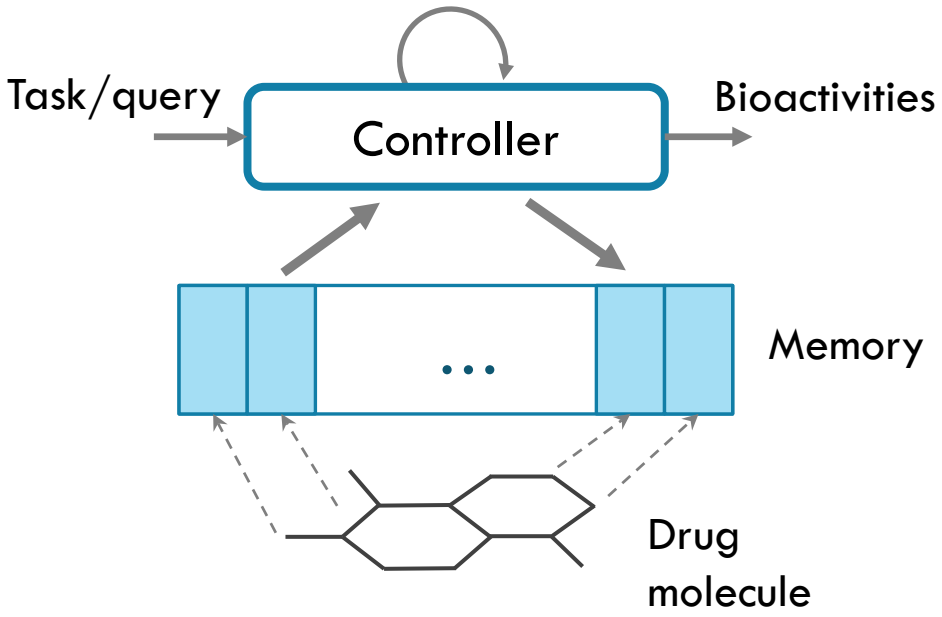
Figure 1: A sample reaction represented as a set of graph transformations from reactants (leftmost) to products (rightmost). Atoms are labeled with their type (Carbon, Oxygen,...) and their index (1, 2,...) in the molecular graph. The atom pairs that change connectivity and their new bonds (if existed) are highlighted in green. There are two bond changes in this case: 1) The double bond between O:1 and C:2 becomes single. 2) A new single bond between C:2 and C:10 is added.

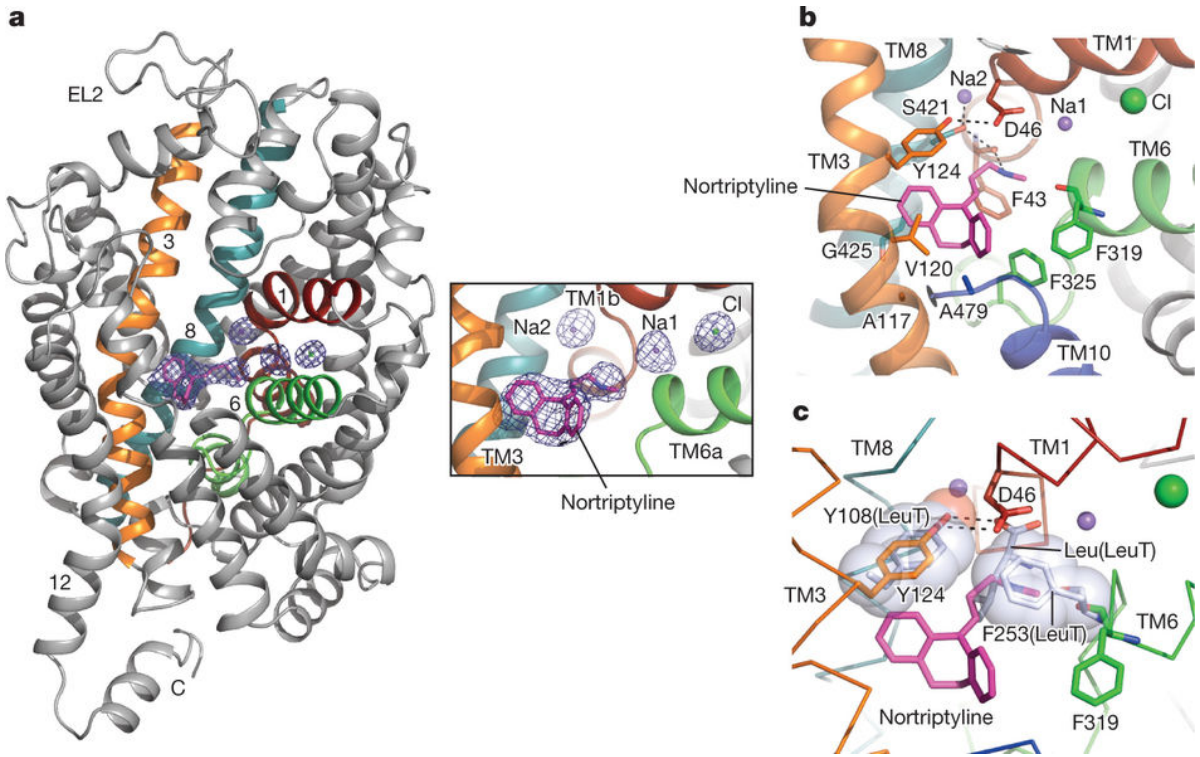
From vector to graph with PAN: Personalized Annotation Networks



Nguyen, Thin, Samuel C. Lee, Thomas P. Quinn, Buu Truong, Xiaomei Li, Truyen Tran, Svetha Venkatesh, and Thuc Duy Le. "Personalized Annotation-based Networks (PAN) for the Prediction of Breast Cancer Relapse." *bioRxiv* (2019): 534628.

Predicting molecular bioactivities as querying a graph

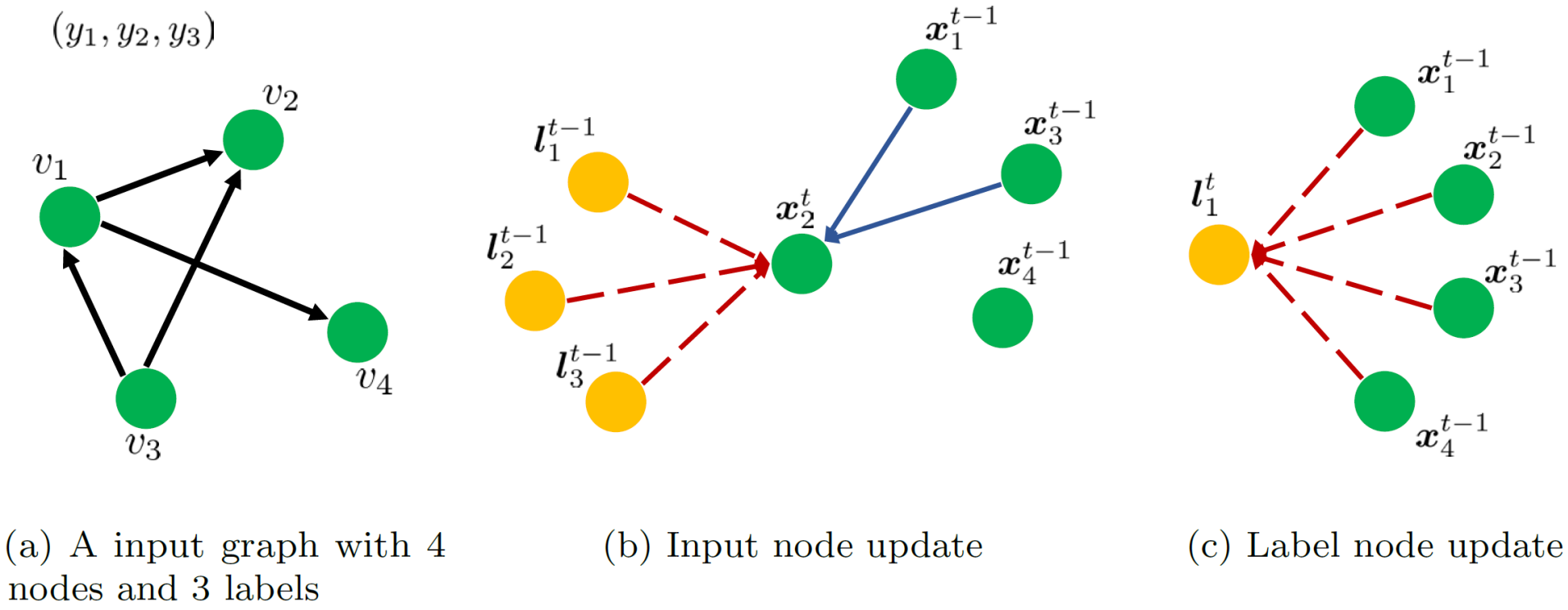




#REF: Penmatsa, Aravind, Kevin H. Wang, and Eric Gouaux. "X-ray structure of dopamine transporter elucidates antidepressant mechanism." *Nature* 503.7474 (2013): 85-90.

#Ref: Pham, Trang, Truyen Tran, and Svetha Venkatesh. "Graph Memory Networks for Molecular Activity Prediction." *ICPR'18*.

Multi-target binding for drug repurposing as graph multi-labeling



(a) A input graph with 4 nodes and 3 labels

(b) Input node update

(c) Label node update

#REF: Do, Kien, et al. "Attentional Multilabel Learning over Graphs-A message passing approach." *Machine Learning, 2019*.

Dataset	Metrics	Fingerprint		SMILES	Molecular Graph		
		SVM	HWN	GRU	WL+SVM	CLN	GAML
9cancers	m-AUC	81.94	85.95	83.29	86.06	88.35	**88.78**
	M-AUC	81.37	85.85	82.74	85.74	88.23	**88.50**
	m-F1	50.63	57.44	55.97	54.55	59.48	**62.03***
	M-F1	50.71	57.29	55.99	54.54	59.50	**62.14***
50proteins	m-AUC	79.85	77.46	79.11	81.62	82.08	**82.82**
	M-AUC	74.77	73.78	75.25	77.60	78.36	**79.35***
	m-F1	17.21	16.37	16.08	17.04	18.37	**20.47***
	M-F1	18.40	15.87	14.96	18.66	17.72	**19.83***

Table 4: The performance in the multi-label classification with graph-structured input (m-X: micro average of X; M-X: macro average). SVM and HWN work on fingerprint representation; GRU works on string representation of molecule known as SMILES; WL+BR and CLN work directly on graph representation. Bold indicates better values. (*) $p < 0.05$.

#REF: Do, Kien, et al. "Attentional Multilabel Learning over Graphs-A message passing approach." *arXiv preprint arXiv:1804.00293*(2018).

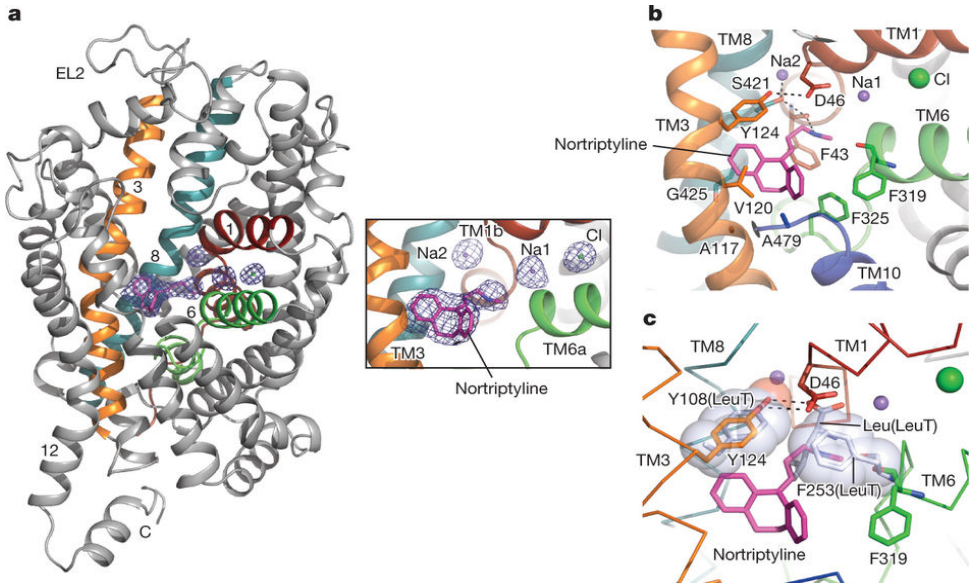
Drug-target binding as graph reasoning

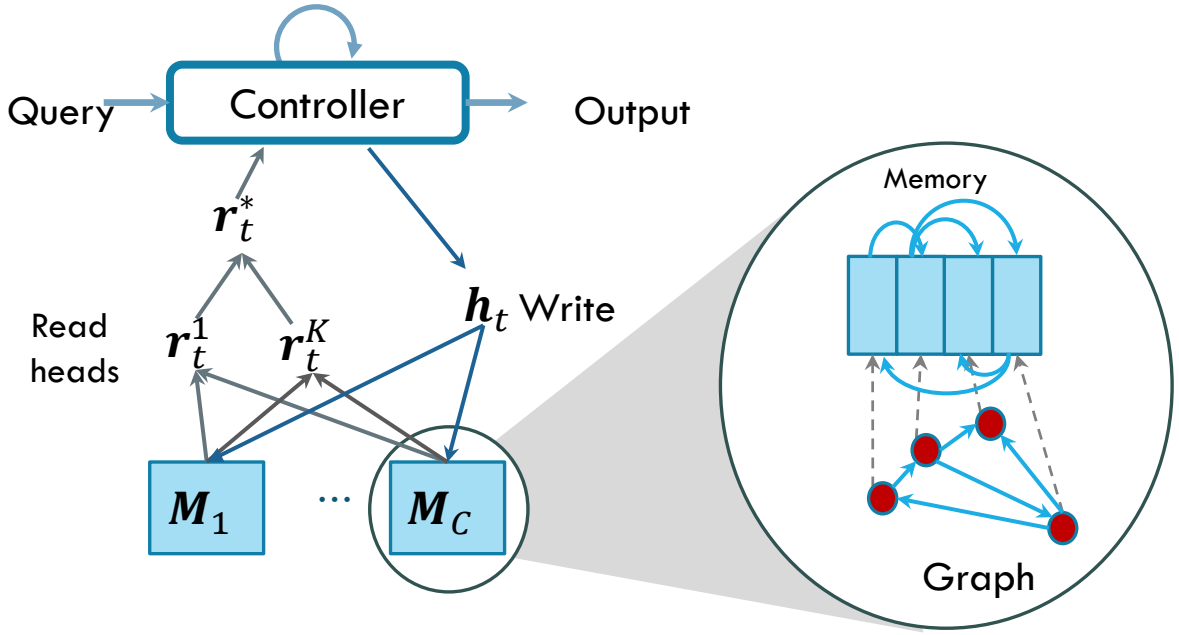
- Reasoning is to deduce knowledge from previously acquired knowledge in response to a query (or a cues)

Can be formulated as Question-Answering or Graph-Graph interaction:

- Knowledge base: Binding targets (e.g., RNA/protein sequence, or 3D structures), as a graph
- Query: Drug (e.g., SMILES string, or molecular graph)
- Answer: Affinity, binding sites, modulating effects

Drug-drug, drug-target & protein-protein as graph-graph interaction

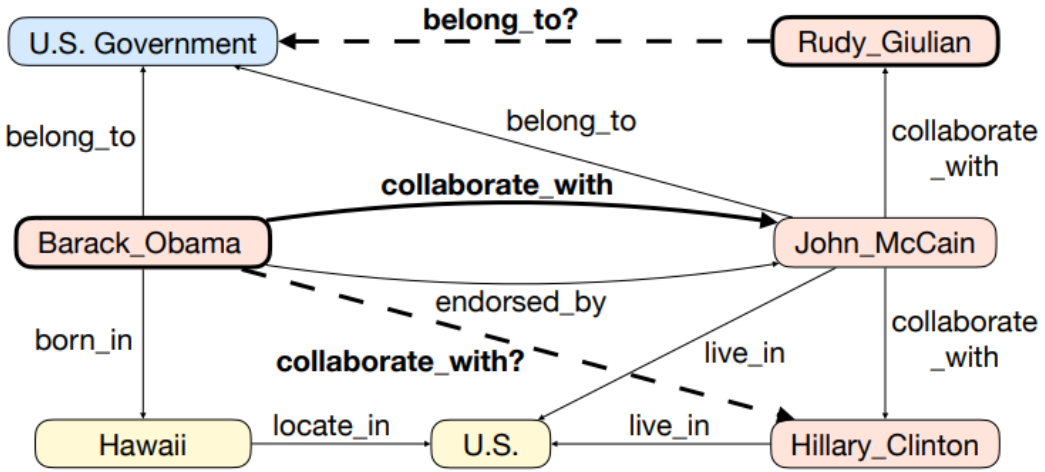




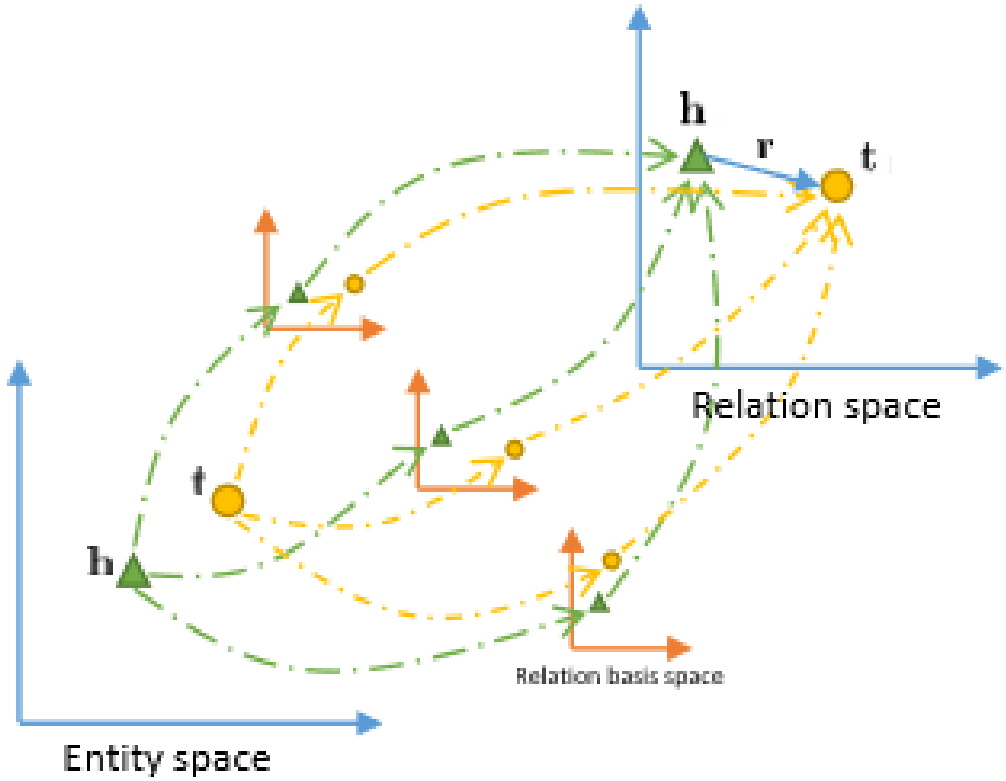
#REF: Penmatsa, Aravind, Kevin H. Wang, and Eric Gouaux. "X-ray structure of dopamine transporter elucidates antidepressant mechanism." *Nature* 503.7474 (2013): 85-90.

Pham, Trang, Truyen Tran, and Svetha Venkatesh. "Relational dynamic memory networks." *arXiv preprint arXiv:1808.04247*(2018).

Inferring (bio) relations as knowledge graph completion



https://www.zdnet.com/article/salesforce-research-knowledge-graphs-and-machine-learning-to-power-einstein/



Do, Kien, Truyen Tran, and Svetha Venkatesh. "Knowledge graph embedding with multiple relation projections." *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018.

Drug design as structured machine translation, aka conditional generation

Can be formulated as structured machine translation:

- Inverse mapping of (knowledge base + binding properties) to (query) → One to many relationship.

Representing graph as string (e.g., SMILES), and use sequence VAEs or GANs.

Graph VAE & GAN

- Model nodes & interactions
- Model cliques

Sequences

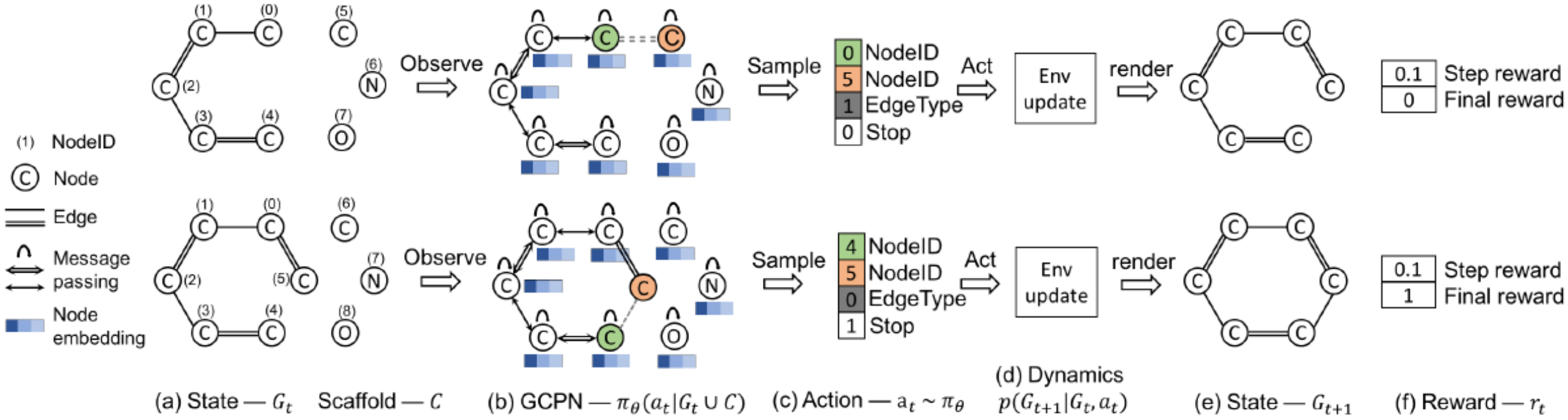
- Iterative methods

Reinforcement learning

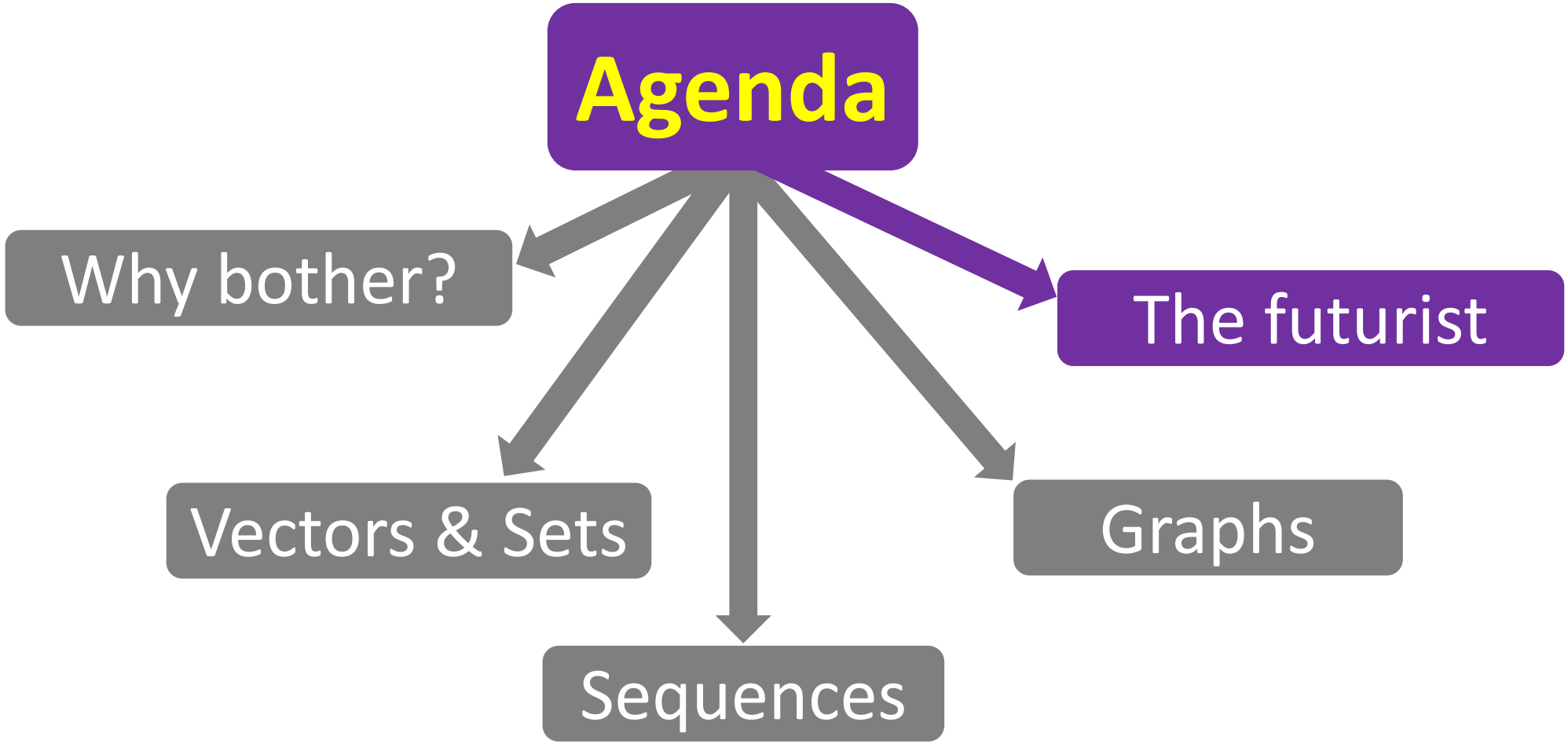
- Discrete objectives

Any combination of these + memory.

Drug design as reinforcement learning



You, Jiaxuan, et al. "Graph Convolutional Policy Network for Goal-Directed Molecular Graph Generation." *NeurIPS* (2018).


Agenda
Why bother?
The futurist
Vectors & Sets
Graphs
Sequences

What can DL do to genomics?

Deep learning offerings

Function approximation

Program approximation

Program synthesis

Deep density estimation

Disentangling factors of variation

Capturing data structures

Generating realistic data (sequences)

Question-answering

Information extraction

Knowledge graph construction and completion

Genomic problems

GWAS, gene-disease mapping

Binding site identification

Function prediction

Drug-target binding

Drug design

Structure prediction

Sequence generation

Functional genomics

Optimizing sequences

Organizing the (knowledge about) omics universe

Deep learning versus genomics

Bertolero, M. A., Blevins, A. S., Baum, G. L., Gur, R. C., Gur, R. E., Roalf, D. R., ... & Bassett, D. S. (2019). The network architecture of the human brain is modularly encoded in the genome. *arXiv preprint arXiv:1905.07606*.

Neuron ↔ Nucleotide, amino acid (building bricks)

Neural networks ↔ Chemical/biological networks (the house)

Message passing ↔ Signalling (the communication)

Neural programs ↔ Proteins/RNAs (the operating machines)

Neural Turing machine ↔ DNA (data + instruction + control)

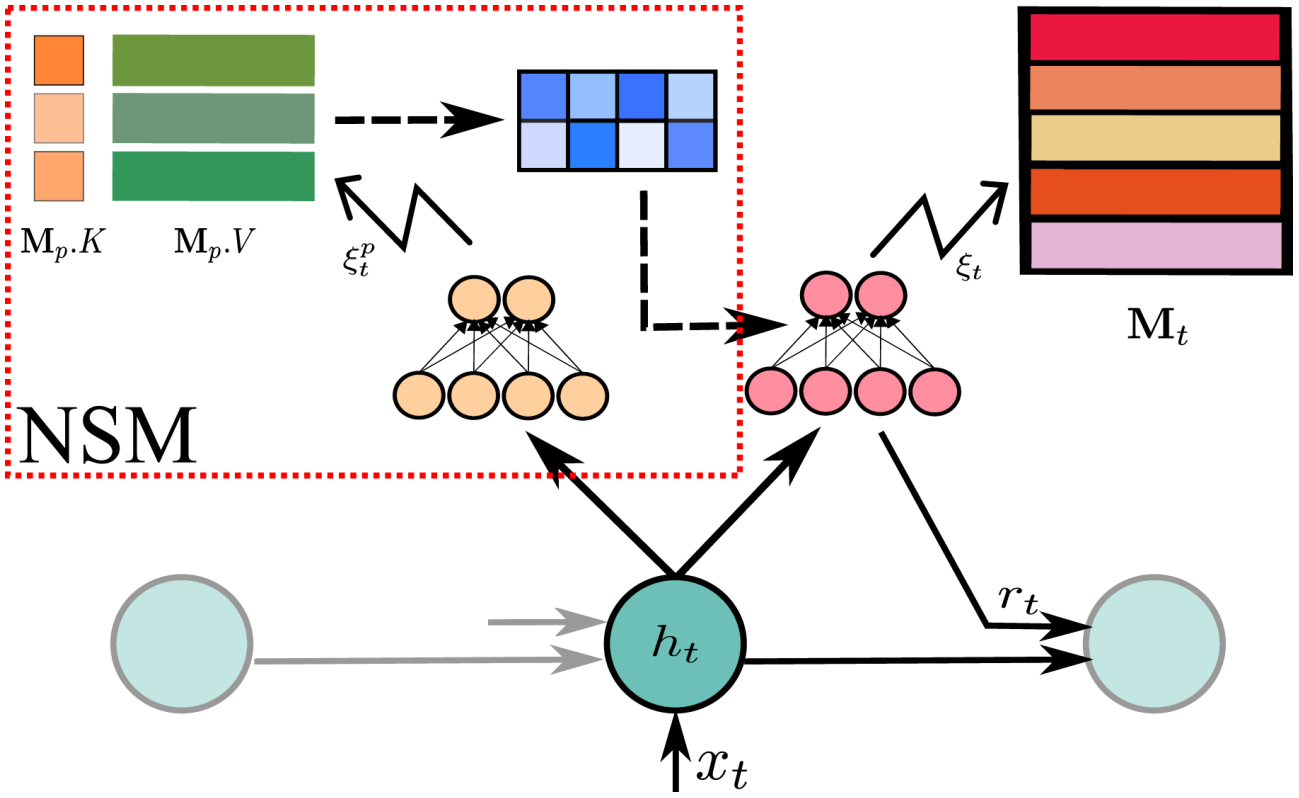
Neural universe ↔ Omics universe (the computational universe)

Learning over time ↔ Co-evolution (adaptation)

Super Neural Turing machine ↔ DNA + Evolution (data + program + adaption)

Living bodies as multiple programs interacting

- We need new (neural) capabilities:
  - Truly Turing machine: programs can be stored and called when needed.
  - Can solve BIG problem with many sub-modules.
    - → Composionality
  - Can reason given existing structures and knowledge bases



Neural Stored-program Memory

Le, Hung, Truyen Tran, and Svetha Venkatesh. "Neural Stored-program Memory." *arXiv preprint arXiv:1906.08862*(2019).

Living in the future: AI for health care

We tend to overestimate the short-term and underestimate the long-term.

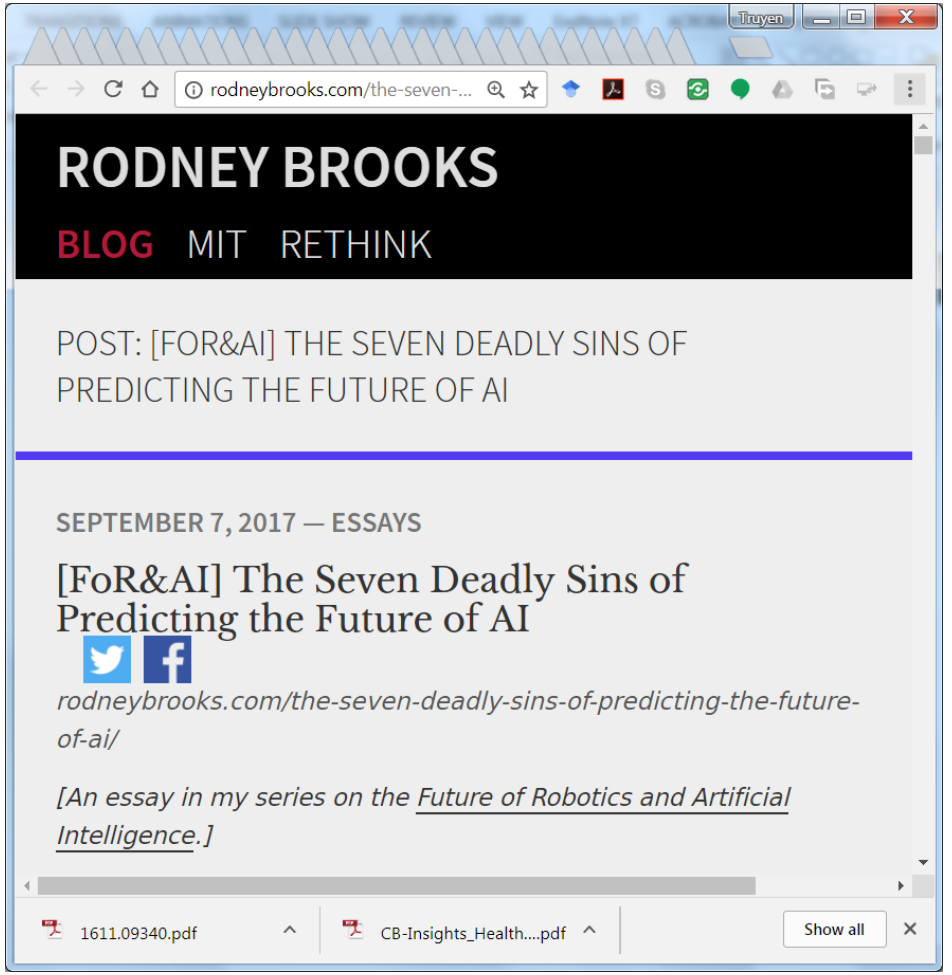
Bear in mind that anything beyond 5 years are nearly impossible to predict!

Let's map Kai-Fu Lee's vision:

- Wave 1: Internet data (→ PubMed, social media)
- **Wave 2: Business data (→EMR)**
- **Wave 3: Digitalize the physical world (→Drugs)**
- Wave 4: Full automation (→ Robot surgeons, GPs)

Some speculations (by me):

- https://letdataspeak.blogspot.com.au/2017/02/living-in-future-deep-learning-for.html



The team

DEAKIN UNIVERSITY AUSTRALIA
Worldly

THE UNIVERSITY OF SYDNEY

GARVAN INSTITUTE

Barwon Health

Epworth HealthCare

We're hiring

PhD & Postdocs

truyen.tran@deakin.edu.au

https://truyentran.github.io/scholarship.html

http://ahsanqawl.com/2015/10/qa/