

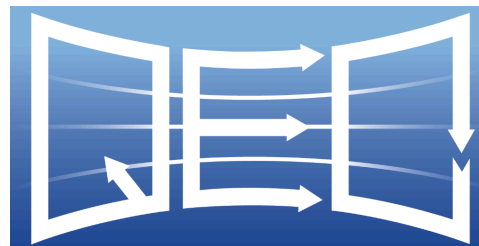
# Case studies in machine learning via quantum annealing

Presented by: Richard Li, University of Southern California

September 26, 2018



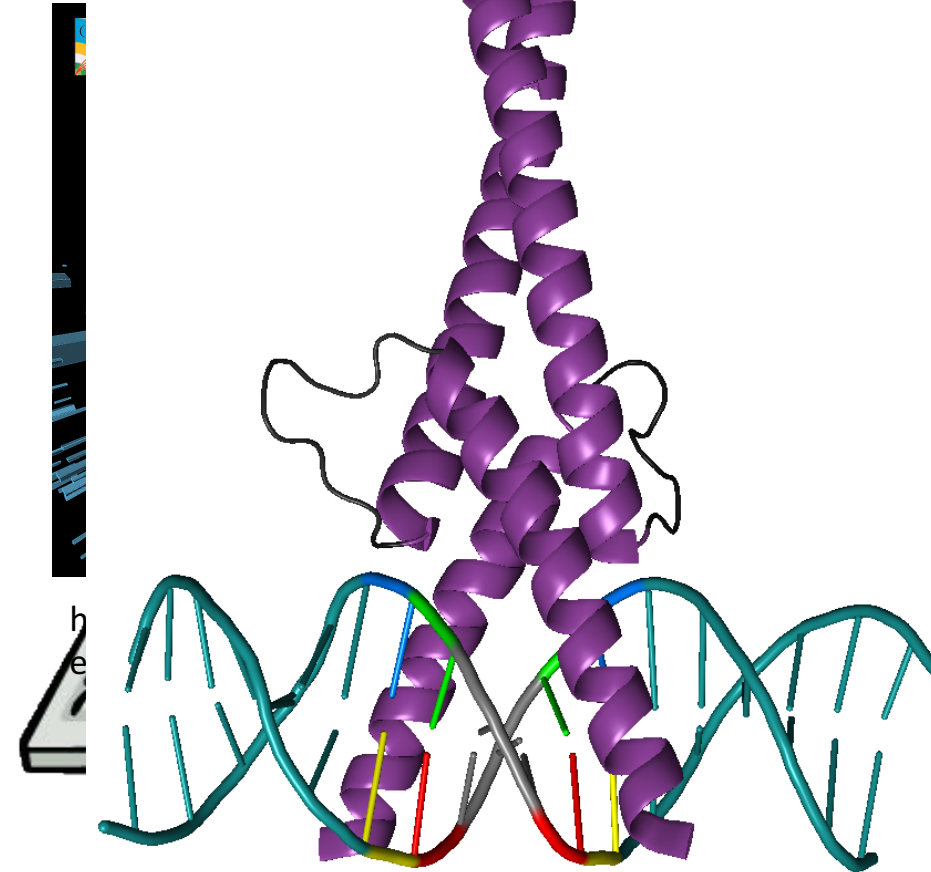
**Caltech**



**WuXiNextCODE**

# Overview

- Machine learning
- Higgs boson
- TF-DNA binding
- Cancer classification



PDB ID: 1NLW

Sun, WJ, et al., *Int. J. Mol. Sci.* 2017

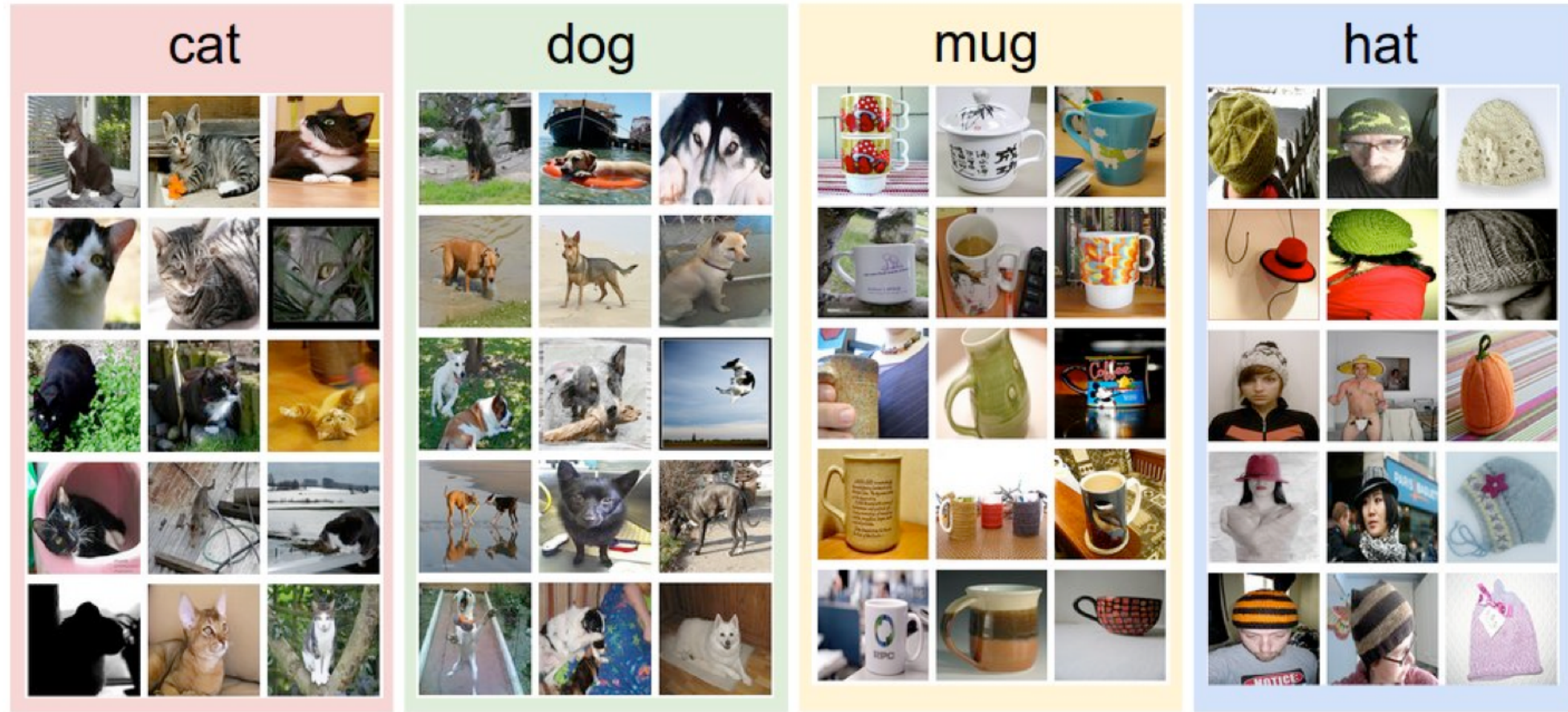
# Machine learning

- Spam detection
- Natural language processing
- Recommender systems



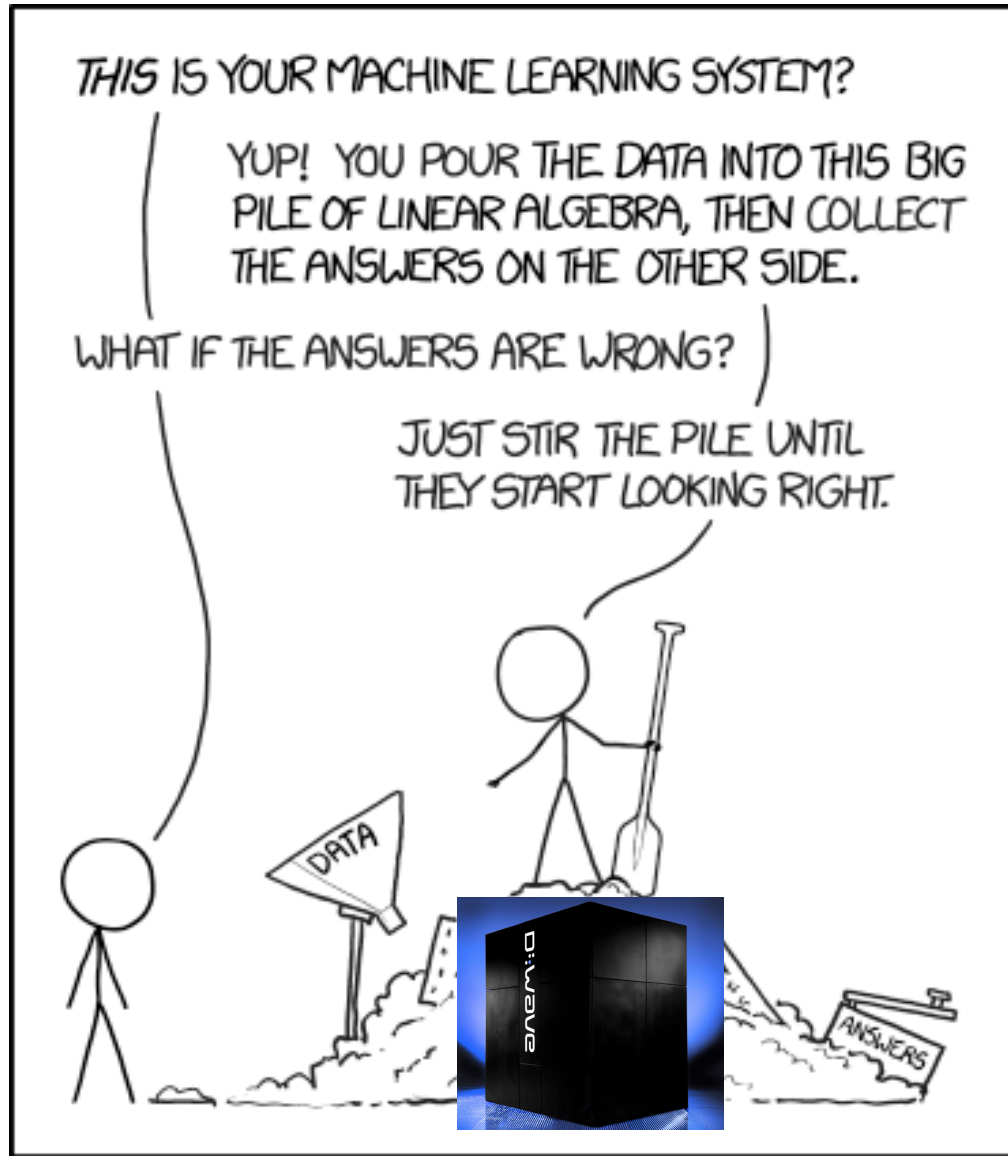
# Machine learning

Give computers ability to learn without being explicitly programmed



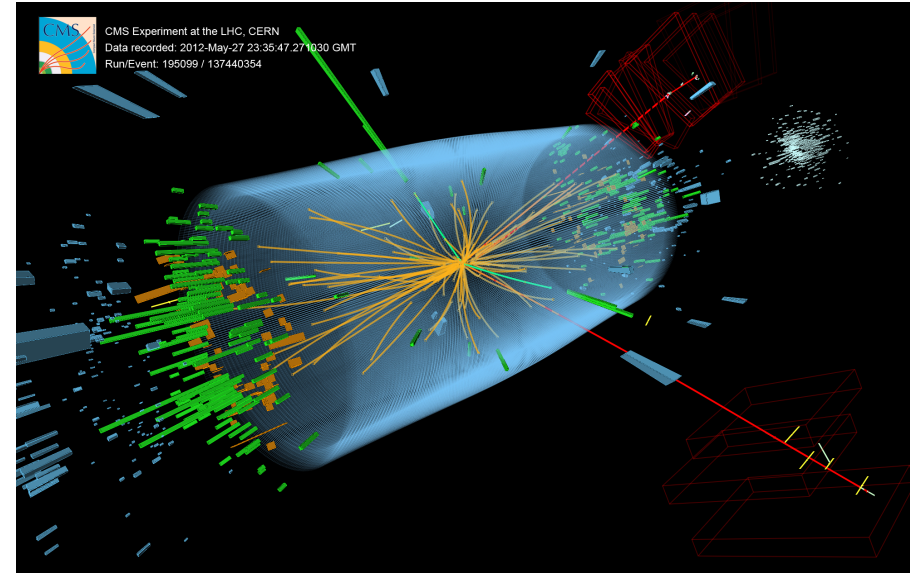


# (Supervised) Machine learning



# Overview

- Machine learning
- Higgs boson
- TF-DNA binding
- Cancer classification



“Solving a Higgs optimization problem with quantum annealing for machine learning”, *Nature* (2017)

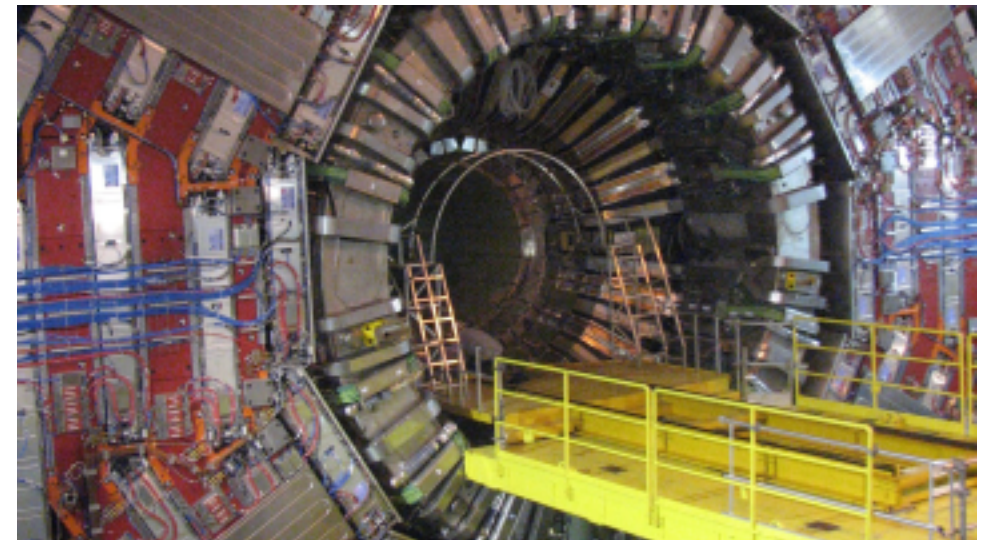
Authors: Alex Mott<sup>1</sup>, Joshua Job<sup>2</sup>, Jean-Roch Vlimant<sup>1</sup>, Daniel Lidar<sup>3</sup>, Maria Spiropulu<sup>1</sup>

Associations: 1. Department of Physics, California Institute of Technology 2. Department of Physics, Center for Quantum Information Science & Technology, University of Southern California 3. Departments of Electrical Engineering, Chemistry, and Physics, Center for Quantum Information Science & Technology, University of Southern California

# Higgs boson – Standard Model

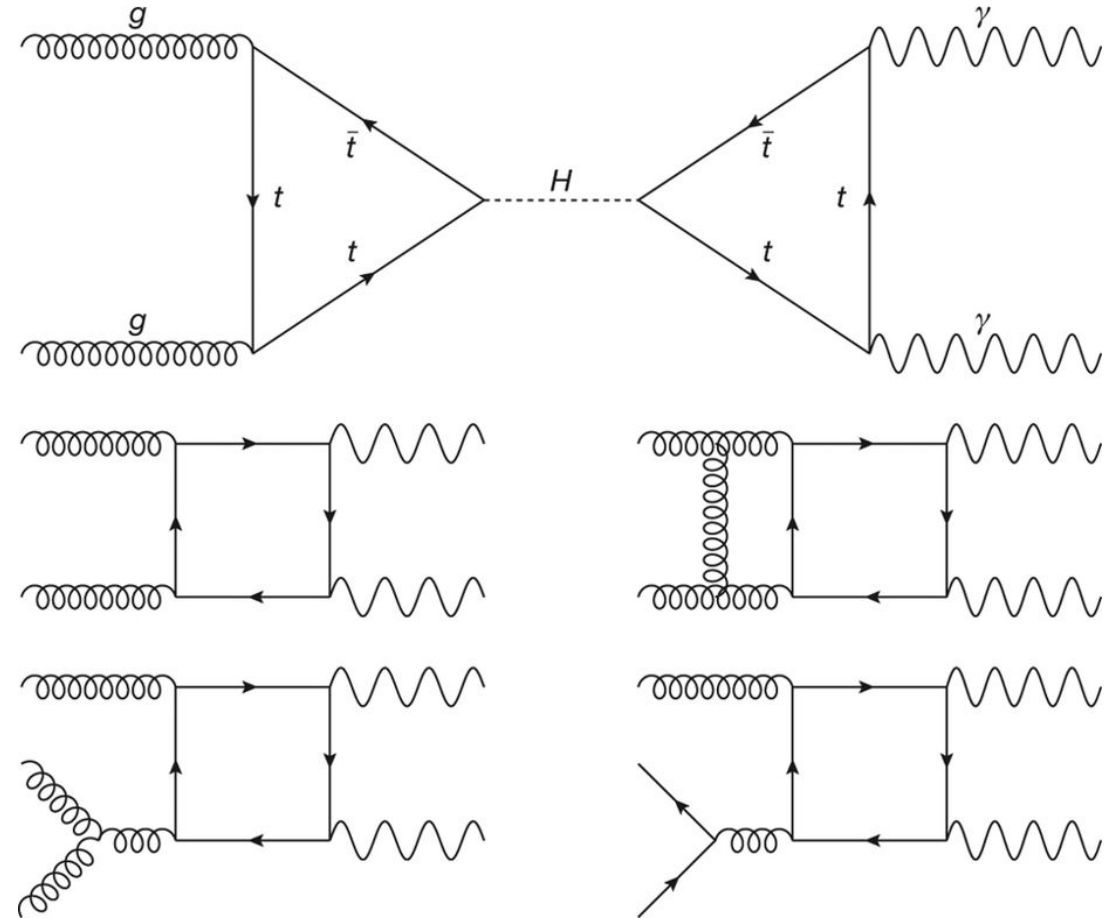
## Standard Model of Elementary Particles

three generations of matter (fermions)							
	I	II	III				
mass	$\approx 2.2 \text{ MeV}/c^2$	$\approx 1.28 \text{ GeV}/c^2$	$\approx 173.1 \text{ GeV}/c^2$	0	$\approx 125.09 \text{ GeV}/c^2$		
charge	$2/3$	$2/3$	$2/3$	0	0		
spin	$1/2$	$1/2$	$1/2$	1	0		
	<b>u</b> up	<b>c</b> charm	<b>t</b> top	<b>g</b> gluon	<b>H</b> Higgs		
<b>QUARKS</b>	<b>d</b> down	<b>s</b> strange	<b>b</b> bottom	<b><math>\gamma</math></b> photon		<b>SCALAR BOSONS</b>	
	$\approx 4.7 \text{ MeV}/c^2$	$\approx 96 \text{ MeV}/c^2$	$\approx 4.18 \text{ GeV}/c^2$	0			
	$-1/3$	$-1/3$	$-1/3$	0			
	$1/2$	$1/2$	$1/2$	1			
	<b>e</b> electron	<b><math>\mu</math></b> muon	<b><math>\tau</math></b> tau	<b>Z</b> Z boson			
<b>LEPTONS</b>	$\approx 0.511 \text{ MeV}/c^2$	$\approx 105.66 \text{ MeV}/c^2$	$\approx 1.7768 \text{ GeV}/c^2$	$\approx 91.19 \text{ GeV}/c^2$			
	0	0	0	0			
	$1/2$	$1/2$	$1/2$	1			
	<b><math>\nu_e</math></b> electron neutrino	<b><math>\nu_\mu</math></b> muon neutrino	<b><math>\nu_\tau</math></b> tau neutrino	<b>W</b> W boson			
	$< 2.2 \text{ eV}/c^2$	$< 1.7 \text{ MeV}/c^2$	$< 15.5 \text{ MeV}/c^2$	$\approx 80.39 \text{ GeV}/c^2$			
	0	0	0	$\pm 1$			
	$1/2$	$1/2$	$1/2$	1			
						<b>GAUGE BOSONS</b>	



# Signal and background

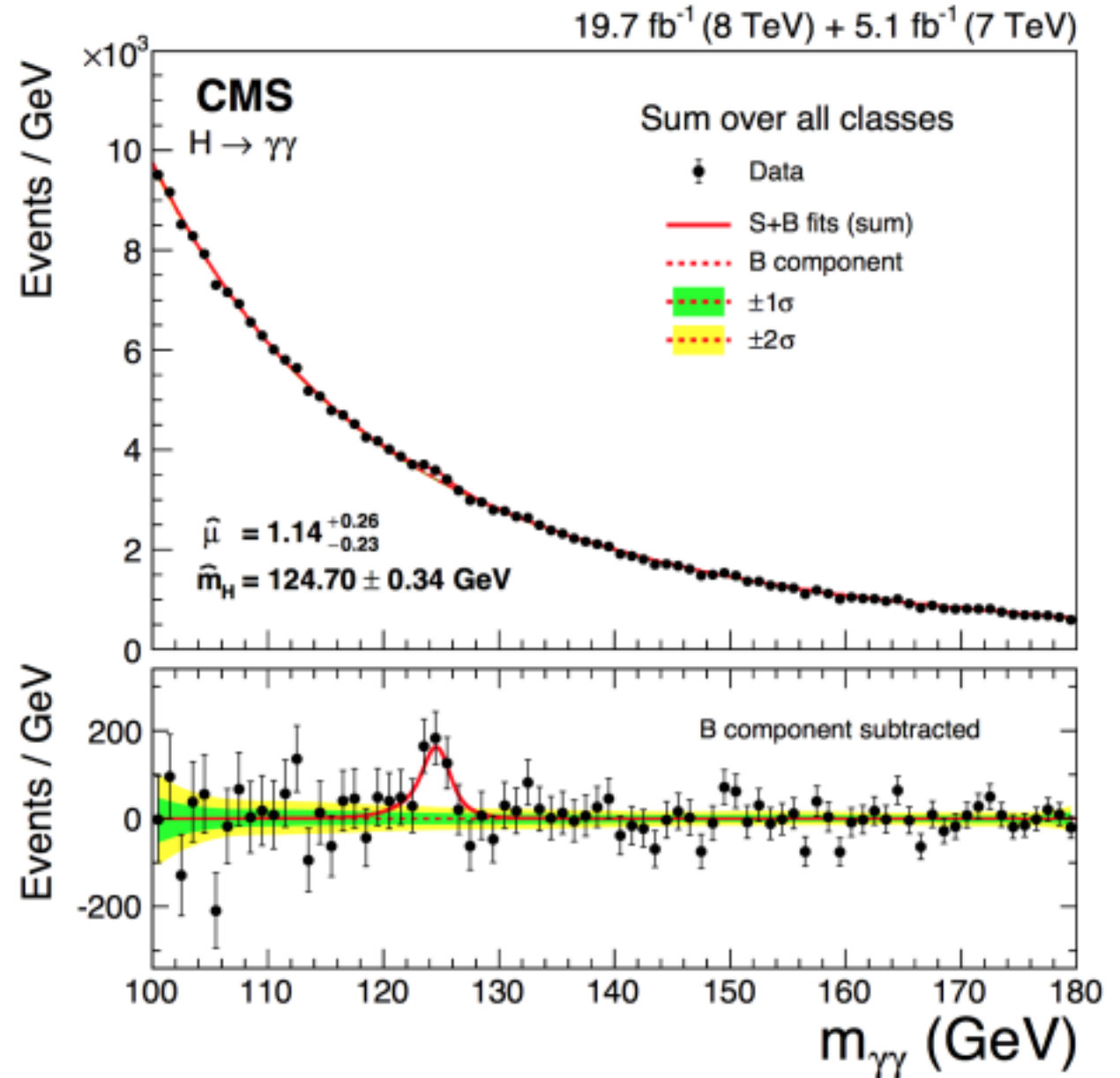
- Signal: a Higgs decaying into two photons; i.e. the  $H \rightarrow \gamma\gamma$  process
- Background processes two gluons producing two photons  $gg \rightarrow \gamma\gamma$





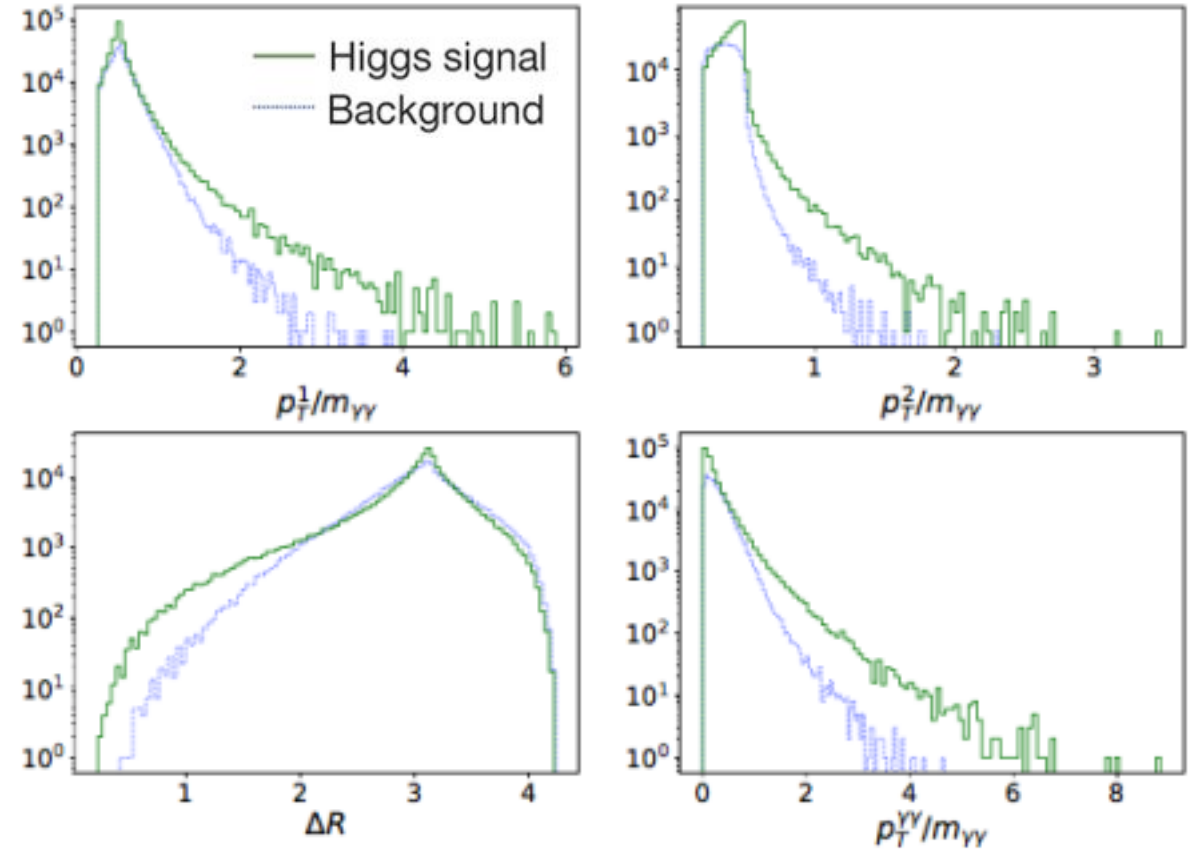
# The data

- <5% of events are signal of Higgs, >95% are background events of other Standard Model processes
- (Use data simulated from event generators)



# The data

- Start with 8 kinematic variables, mostly involving momenta of output photon pairs
- Take various products, end with 36 final variables
- Predict whether event is Higgs or background
- Construct data so that equal number of signal and background events



# The approach

- Construct weak classifiers  $h(\mathbf{x}_i)$  based on the 36 variables
- Use *boosting* to construct a strong classifier (learn the  $w$ 's)
- Minimize error  $\delta(\vec{w})$  between prediction and output

$$p(y_i = 1 | h(\mathbf{x}_i) > 0) > p(y_i = 1) > 0$$
$$-1 \leq h(\mathbf{x}_i) \leq 1, \quad \forall \mathbf{x}_i$$

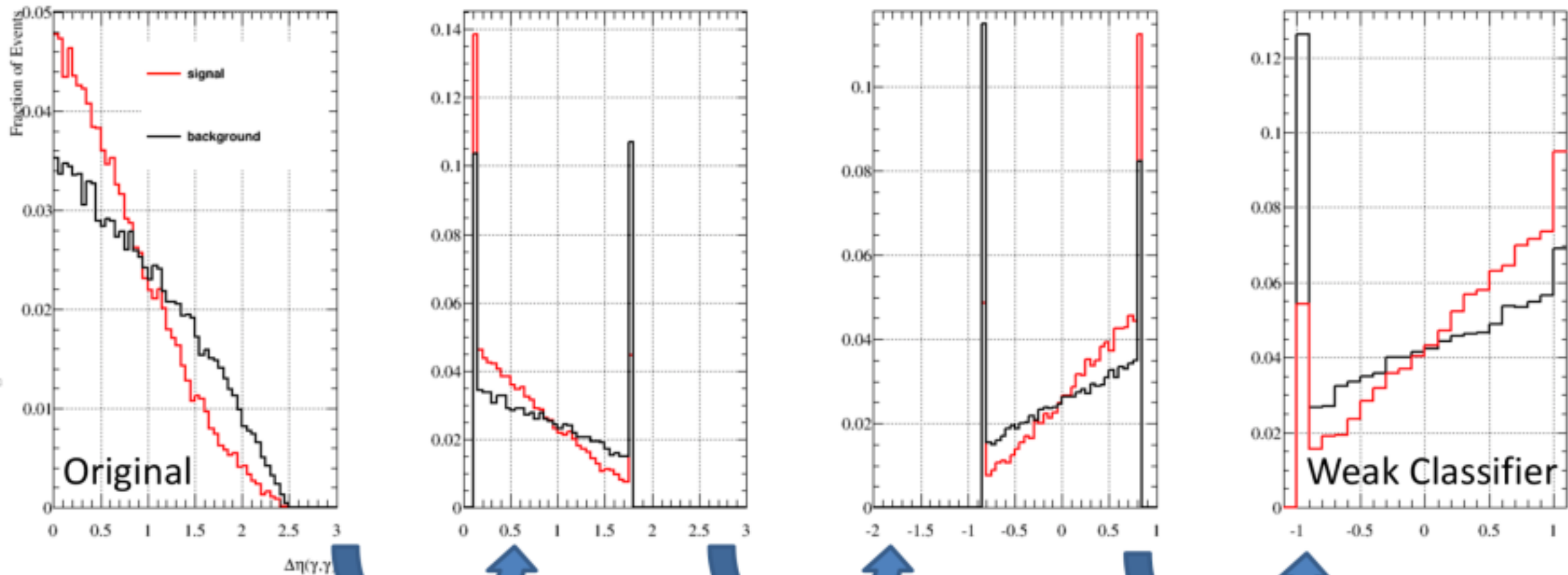
$$O(\mathbf{x}) = \sum_j w_j h_j(\mathbf{x})$$

$$O(\mathbf{x}) < O_c \rightarrow B$$

$$O(\mathbf{x}) > O_c \rightarrow S$$

$$\delta(\vec{w}) = \sum_{x \in T} (y(x) - \sum_i w_i h_i(x))^2$$

# Construction of the Weak Classifiers



- Truncate Distributions
- Assign overflowing events to high/low bins

- Invert distribution
- Center at 0

- Rescale distribution so it fills [-1,1]

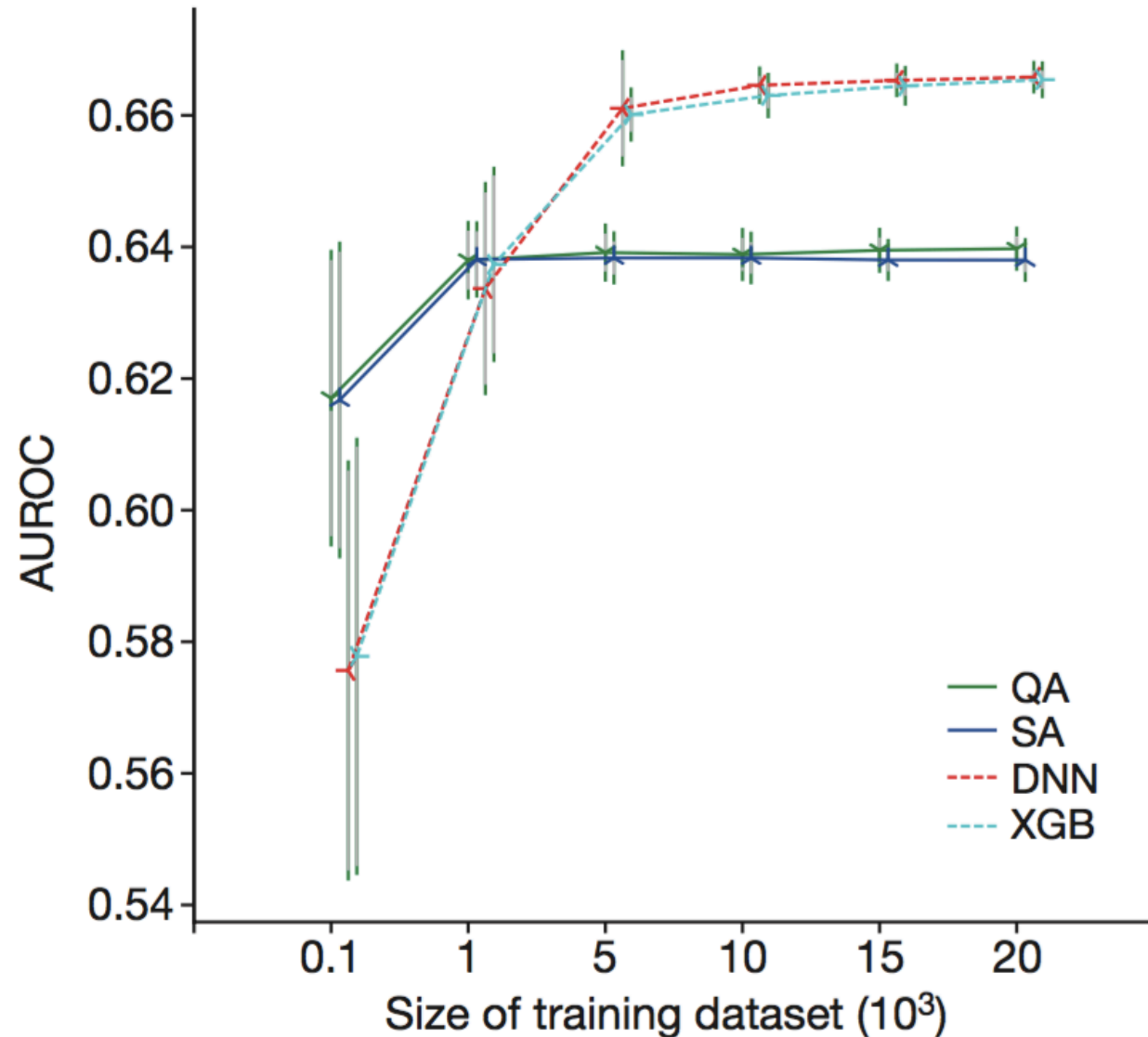


# Results

- Interpretability: found variables that appear the most important for classification

$$\left[ p_T^2, (\Delta R p_T^{\gamma\gamma})^{-1}, \text{ and } \frac{p_T^2}{p_T^{\gamma\gamma}} \right]$$

- Classification performance (ROC)

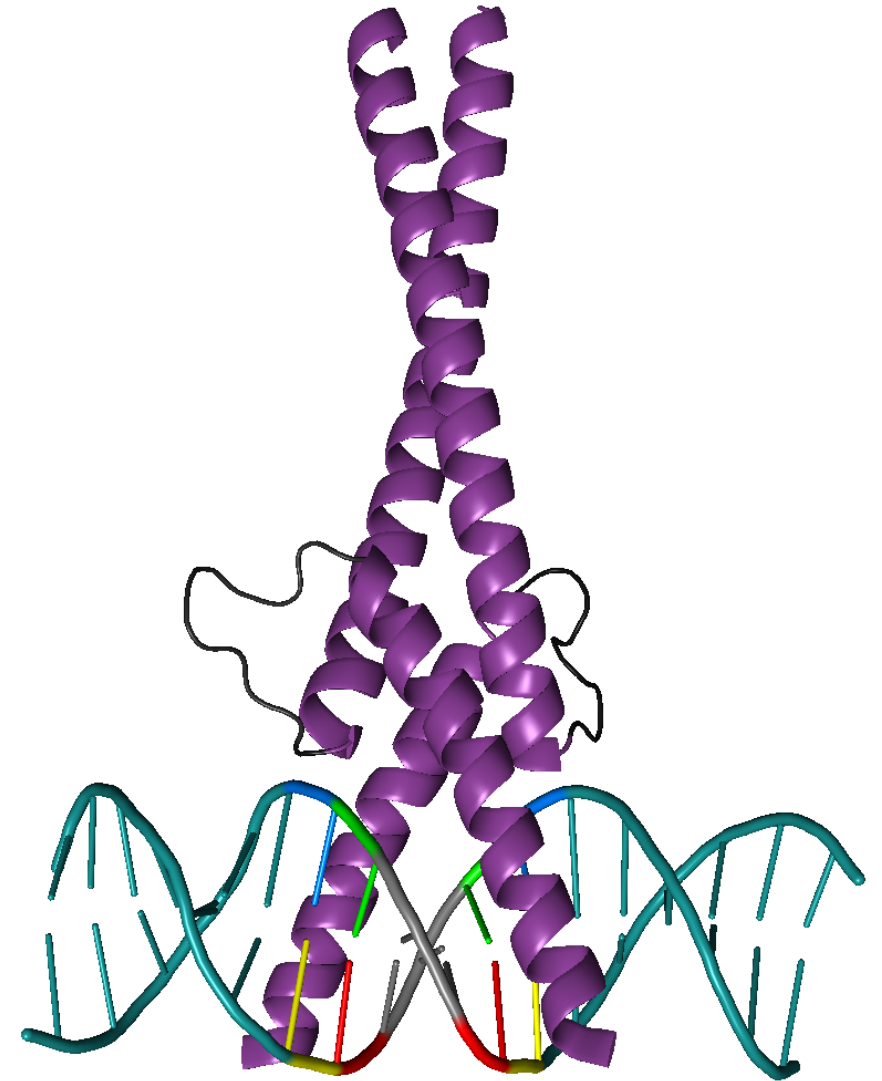


# Overview

- Machine learning
- Higgs boson
- **TF-DNA binding**
- Cancer classification

“Quantum annealing versus classical machine learning applied to a simplified computational biology problem”, *npjQI* (2018)

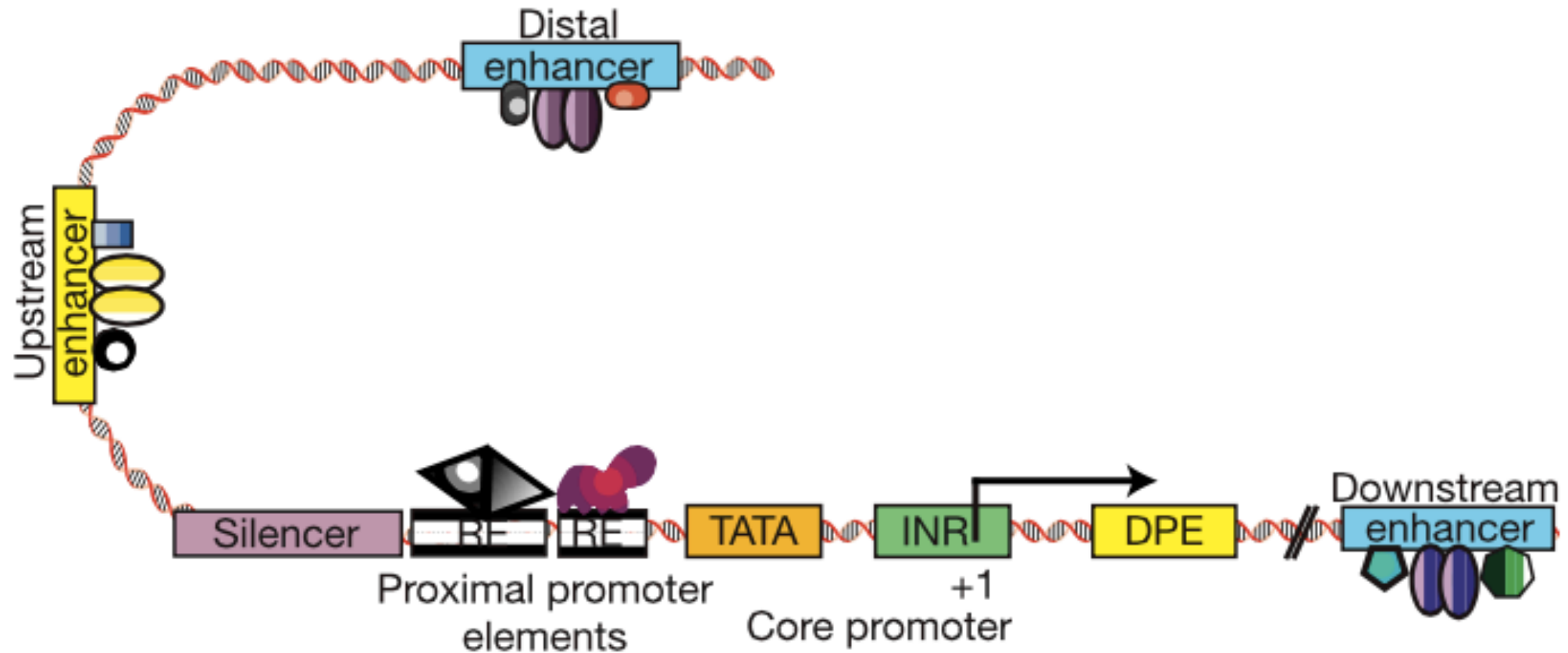
Authors: Richard Li, Rosa Di Felice, Remo Rohs, Daniel Lidar



PDB ID: 1NLW

# Background

## Central dogma of molecular biology



# The data

- Three PBM datasets: Mad, Max, Myc

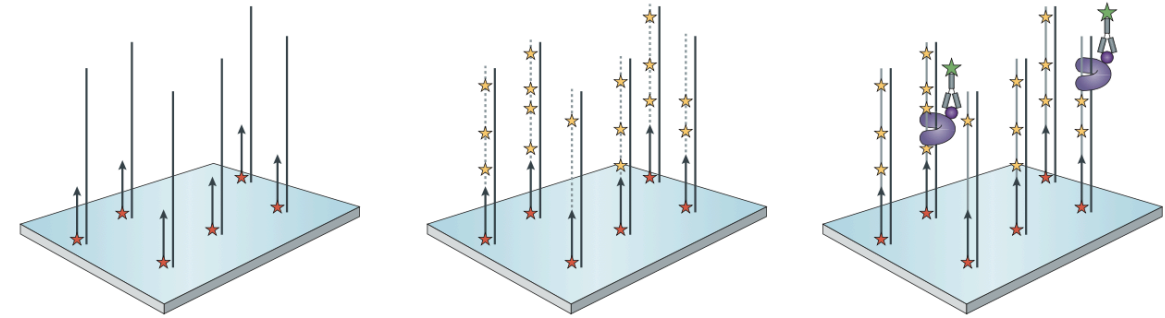
- Get sequences and log signal intensities  $\{(\vec{x}_n, y_n)\}$

$$\vec{x}_n = (x_{n,1}, x_{n,2}, \dots, x_{n,L})$$

$$x_{n,j} \in \{A, C, G, T\}, j = 1, \dots, L$$

- Encode sequences with one-hot encoding

$$\vec{\phi}_n \equiv \phi(\vec{x}_n) = (\phi_1(\vec{x}_n), \phi_2(\vec{x}_n), \dots, \phi_M(\vec{x}_n))$$



Stormo, G. D. et al. (2010) *Nat Rev Genet.*

$y_1 = 4.95$	$x_1 = \text{ACA ACTAA}$
$y_2 = 5.32$	$x_2 = \text{ACA ACTAC}$
...	...
$y_n = 4.81$	$x_n = \text{TCGCGTGT}$
...	...



Zhou, T et al. (2015) *PNAS*



(A slide of math)

$$f_{\mathbf{w}}(\vec{\phi}_n) = \mathbf{w}^\top \vec{\phi}_n = \sum_{m=1}^M w_m \phi_m(\vec{x}_n)$$

$$\mathbf{w}_{\text{opt}} = \arg \min_{\mathbf{w}} \sum_{n=1}^N \left( y_n - f_{\mathbf{w}}(\vec{\phi}_n) \right)^2 + \lambda \|\mathbf{w}\|_1$$

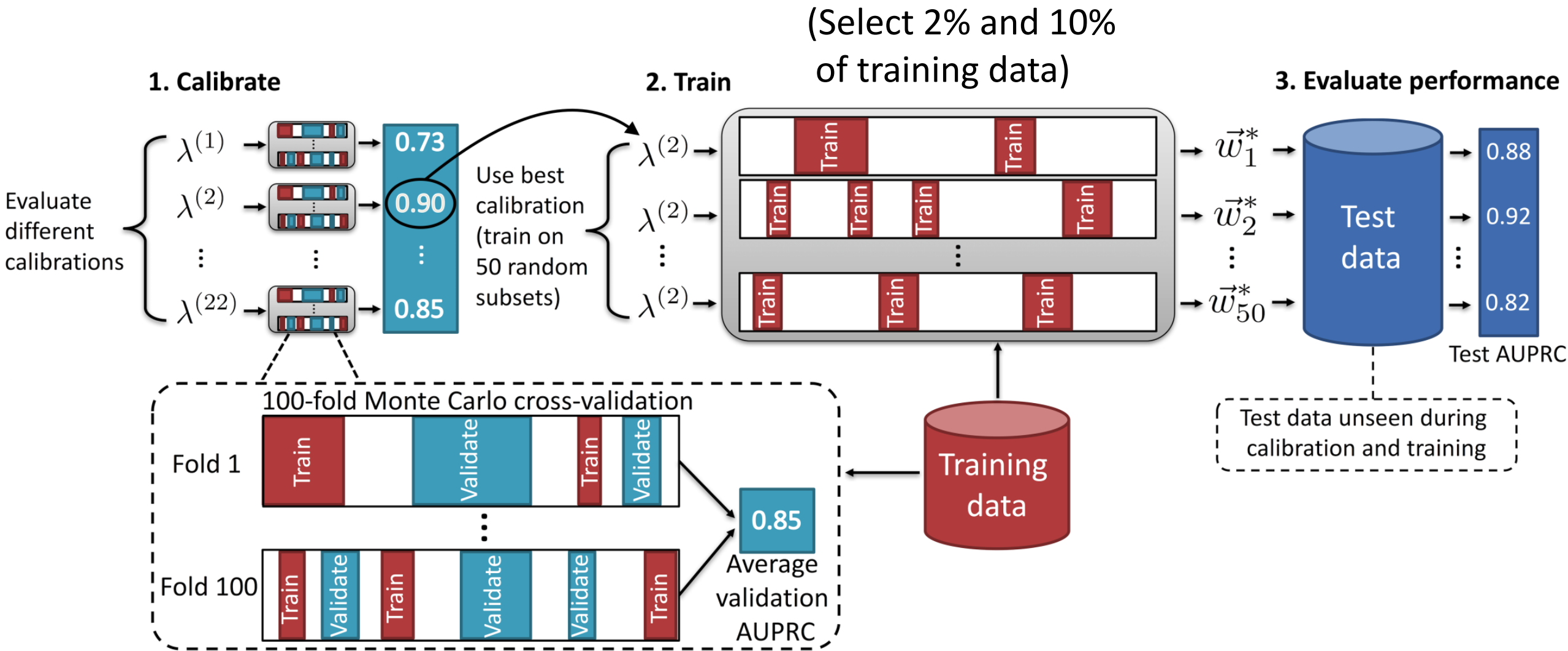
$$= \arg \min_{\mathbf{w}} \sum_n y_n^2 - 2y_n f_{\mathbf{w}}(\vec{\phi}_n) + f_{\mathbf{w}}(\vec{\phi}_n)^2 + \lambda \|\mathbf{w}\|_1$$

$$= \arg \min_{\mathbf{w}} \sum_m \sum_n -2y_n \phi_m(\vec{x}_n) w_m + \sum_{m,m'} \left( \sum_n \phi_m(\vec{x}_n) \phi_{m'}(\vec{x}_n) \right) w_m w_{m'} + \sum_m \lambda w_m$$

$$= \arg \min_{\mathbf{w}} \sum_m h_m w_m + \sum_{m,m'} J_{m,m'} w_m w_{m'}, \text{ where}$$

$$h_m = \lambda - 2 \sum_n y_n \phi_m(\vec{x}_n) \text{ and } J_{m,m'} = \sum_n \phi_m(\vec{x}_n) \phi_{m'}(\vec{x}_n)$$

# Workflow



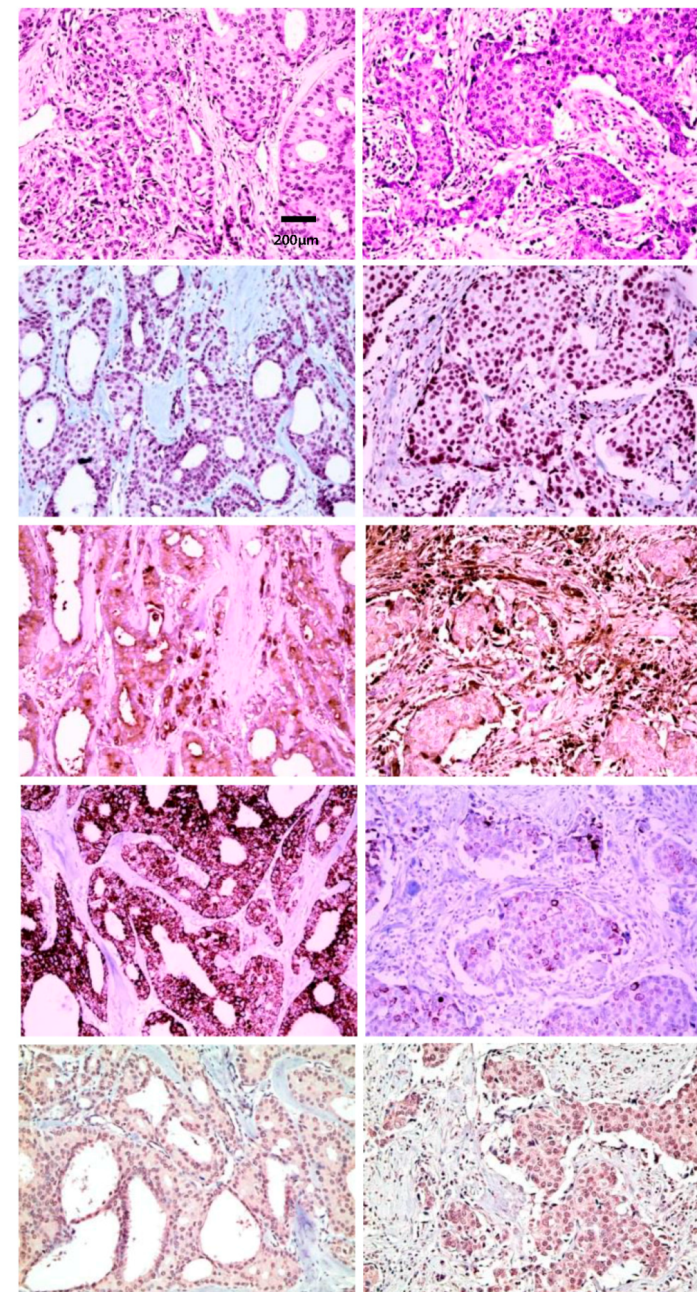


# Overview

- Machine learning
- Higgs boson
- TF-DNA binding
- **Cancer classification**

“Classifying cancer multi-omics data on a quantum computer” (in preparation)

Omar E. Gamel, Richard Y. Li, Nicholas A. Cilfone, Daniel A. Lidar, Thomas W. Chittenden



Sun, WJ, et al., *Int. J. Mol. Sci.* **2017**



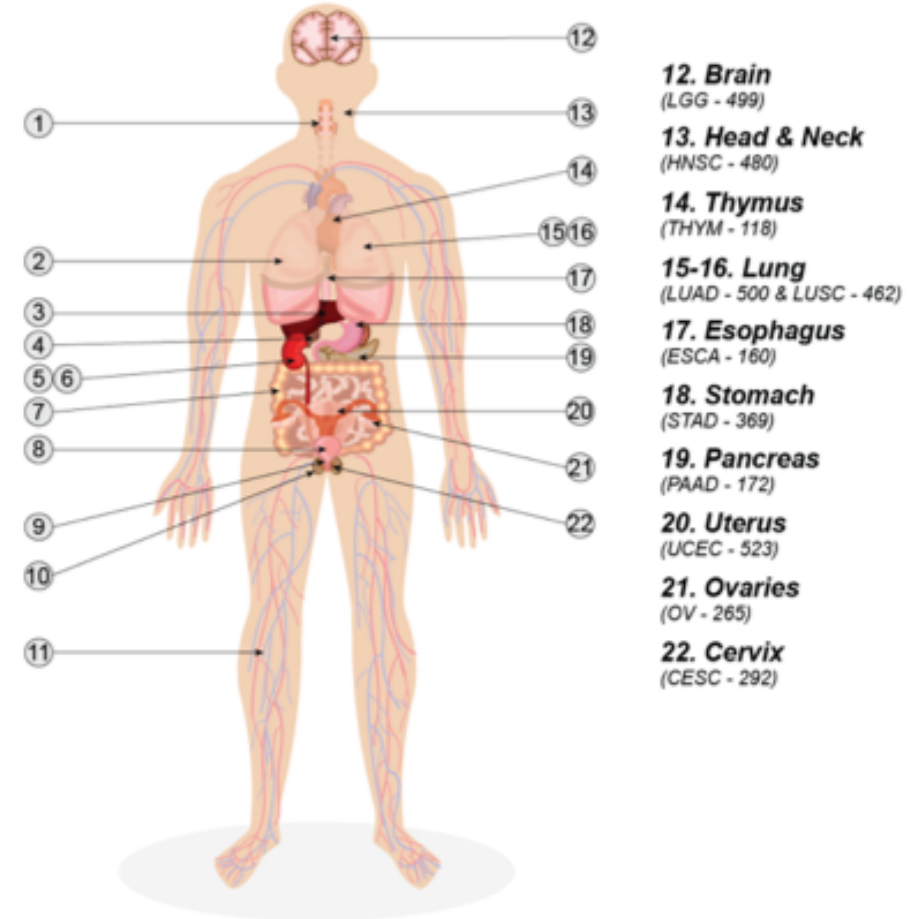
# The data

- An ensemble multi-omics cancer data pipeline
  - A variety of different signals
  - Construct metagenes
- Used cancer types from the cancer genome atlas

(i) Data

1. **Thyroid**  
(THCA - 481)
2. **Breast**  
(BRCA - 1006\*\*)
3. **Liver**  
(LIHC - 358)
4. **Adrenal Gland & Misc.<sup>†</sup>** (PCPG - 159)
- 5-6. **Kidney**  
(KIRP - 284 & KIRC - 327)
7. **Colon & Rectum**  
(COAD\* & READ\* - 551)
8. **Bladder**  
(BLCA - 401)
9. **Prostate**  
(PRAD - 483)
10. **Testis<sup>†</sup>**  
(TGCT - 133)
11. **Soft Tissue**  
(SARC - 249)

The Cancer Genome Atlas (TCGA)



# The approach

- Assign probability to each class by using softmax
- Find set of weights that minimize negative log-likelihood
- Take 2<sup>nd</sup> order Taylor approximation

$$Pr(y_i = c) = \frac{\exp \mathbf{w}_c^\top \mathbf{x}_i}{1 + \sum_{k=1}^{K-1} \exp \mathbf{w}_k^\top \mathbf{x}_i}, \quad c = 1, \dots, K-1$$

$$Pr(y_i = K) = \frac{1}{1 + \sum_{k=1}^{K-1} \exp \mathbf{w}_k^\top \mathbf{x}_i}$$

$$\{\mathbf{w}_k^*\} = \arg \min_{\{\mathbf{w}_k\}} \mathcal{L}$$

$$\mathcal{L} = -\log \prod_i Pr(y_i) = -\sum_i \log Pr(y_i)$$

$$= -\sum_{i: y_i \in 1 \dots K-1} \mathbf{w}_{y_i}^\top \mathbf{x}_i + \sum_i \log \left( 1 + \sum_{k=1}^{K-1} \exp \mathbf{w}_k^\top \mathbf{x}_i \right)$$

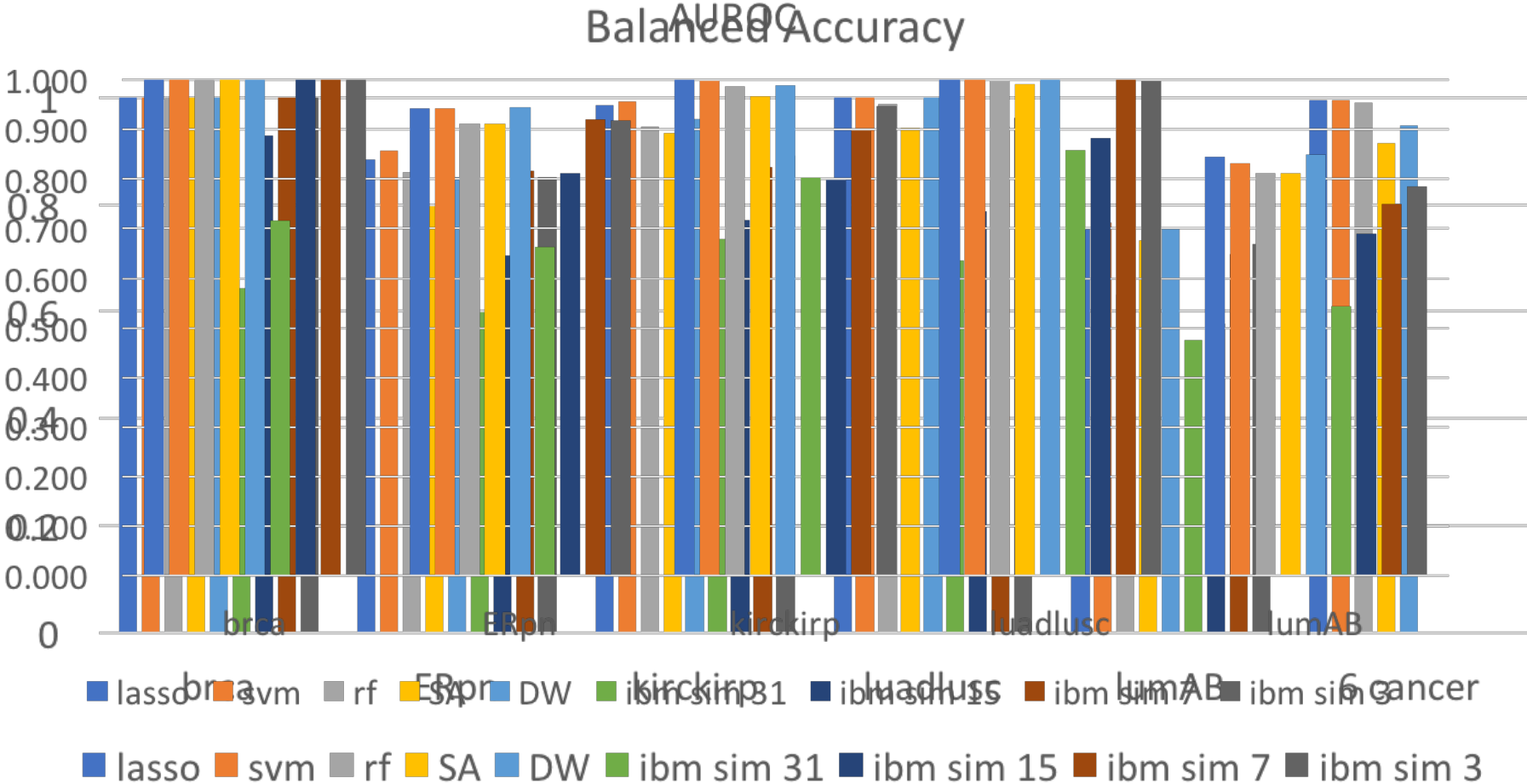
$$\mathcal{L} \approx -\sum_{i: y_i \in 1 \dots K-1} z_i^{(y_i)} + \sum_i \left[ \log K + \frac{1}{K} \sum_k z_i^{(k)} + \frac{1}{2} \frac{K-1}{K^2} \sum_k \left( z_i^{(k)} \right)^2 - \frac{1}{2K^2} \sum_{j \neq k} z_i^{(j)} z_i^{(k)} \right]$$

$$= \sum_{k=1}^{K-1} \mathbf{w}_k^\top (\mathbf{b}_k + \mathbf{h}') + \sum_{k=1}^{K-1} \mathbf{w}_k^\top \mathbf{J}' \mathbf{w}_k - \sum_{k=1}^{K-1} \sum_{j \neq k} \mathbf{w}_j^\top \mathbf{J}'' \mathbf{w}_k$$

$$\mathbf{b}_k = \sum_{i: y_i=k} -\mathbf{x}_i, \quad \mathbf{h}' = \sum_i \mathbf{x}_i, \quad \mathbf{J}' = \frac{K-1}{2K^2} \sum \mathbf{x}_i \mathbf{x}_i^\top, \quad \text{and } \mathbf{J}'' = \frac{1}{2K^2} \sum \mathbf{x}_i \mathbf{x}_i^\top$$

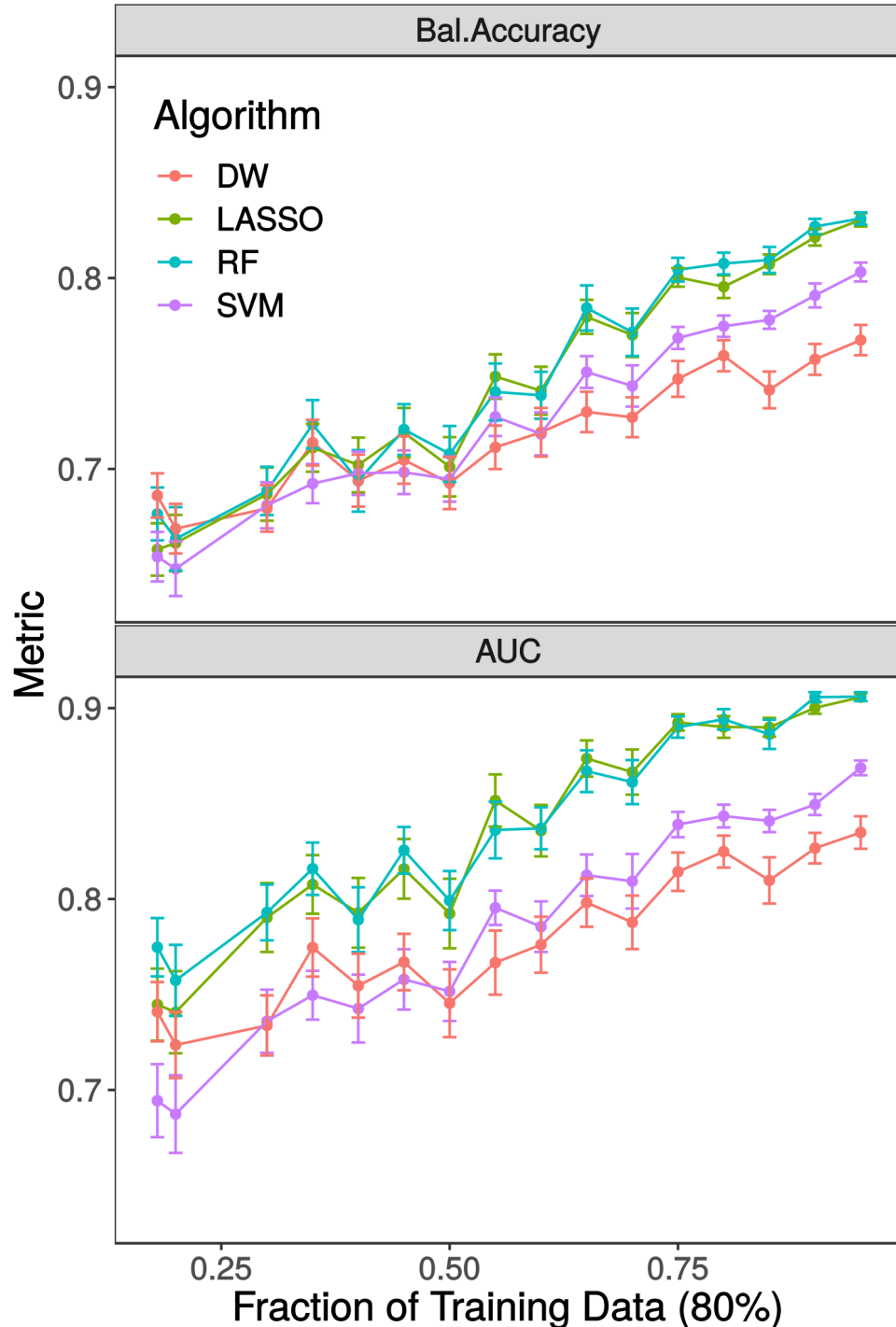
# Results

- Breast cancer tumor vs normal
- Estrogen receptor + vs -
- Kidney clear vs papillary cell carcinoma
- Lung squamous cell vs adenocarcinoma
- Luminal A vs Luminal B
- 6 cancer types



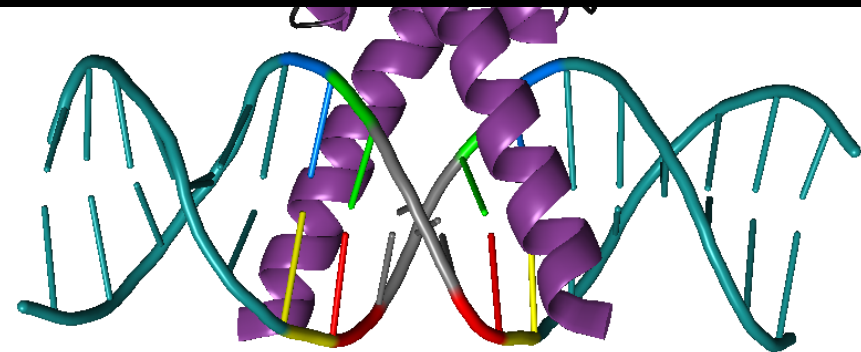
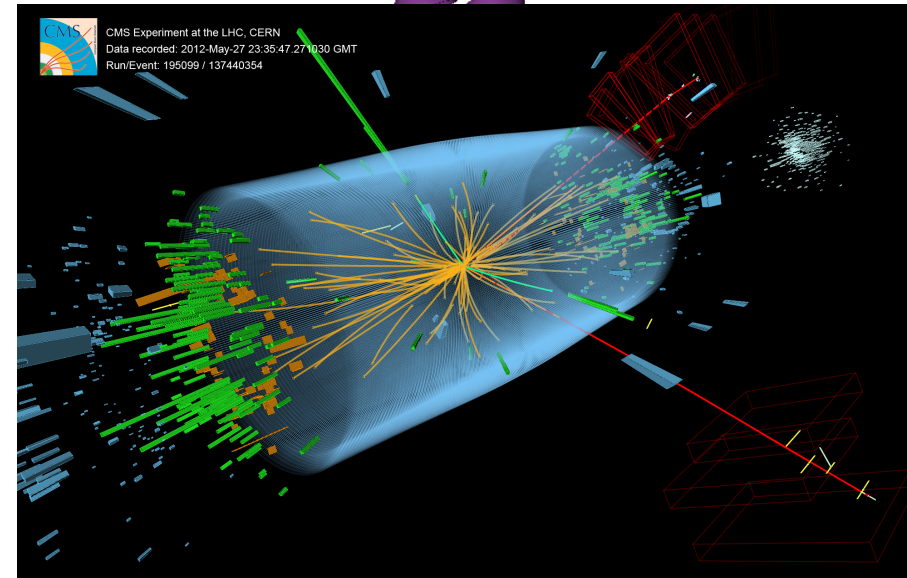
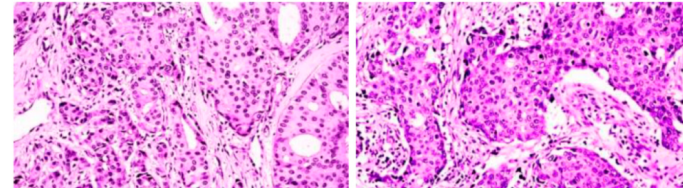
# Results

- Comparison on one dataset using “raw” meta-genes
- DW has a slight advantage for some performance metrics, not for some others



# Summary

- Higgs Boson
  - Construct weak classifiers on kinematic variables
  - Use boosting to predict class
- TF-DNA binding
  - sequence-based features
  - Predict binding affinity
- Cancer classification
  - Multi-omics data
  - Predict probabilities of classes



# Summary

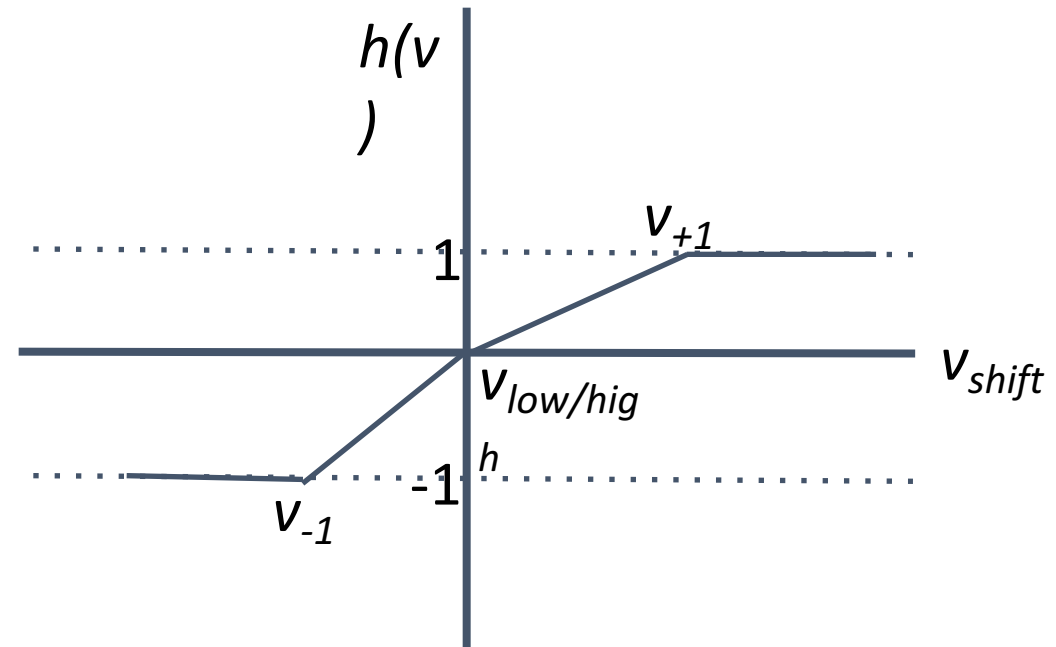
- Seems to do slightly better on small training sizes
- Not making a claim or a quantum advantage in general (usually SA does as well)
- Enforcing binary weights means a simpler model
- Fairly interpretable



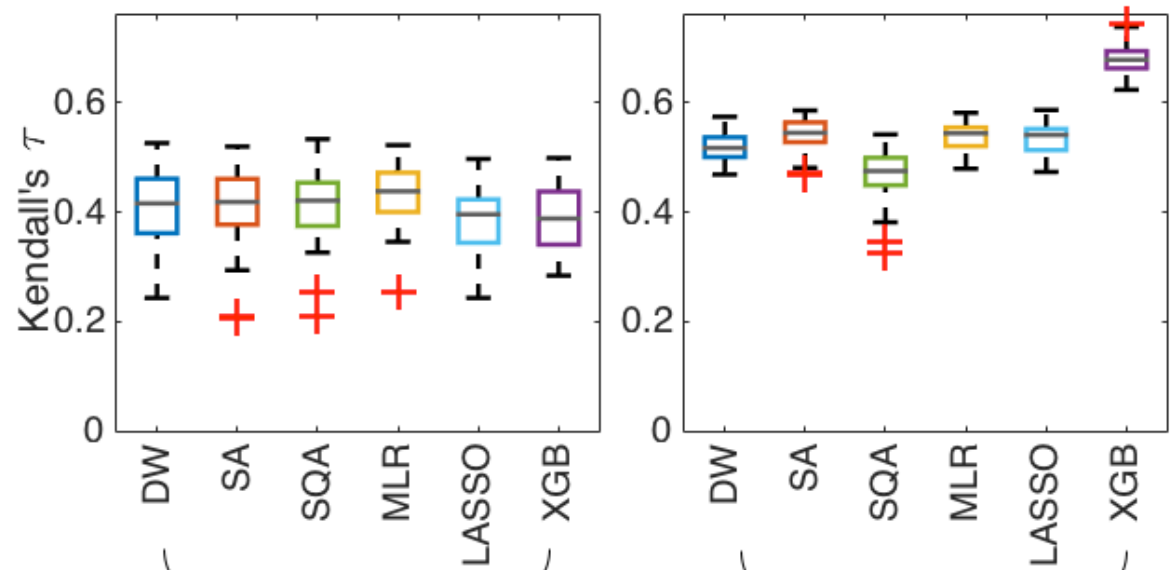
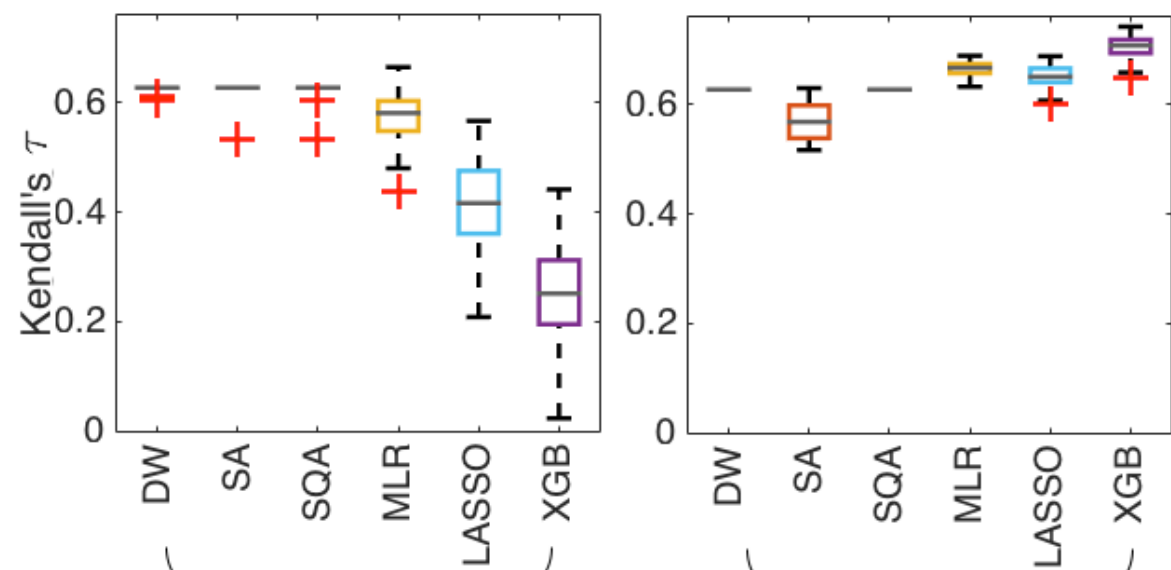
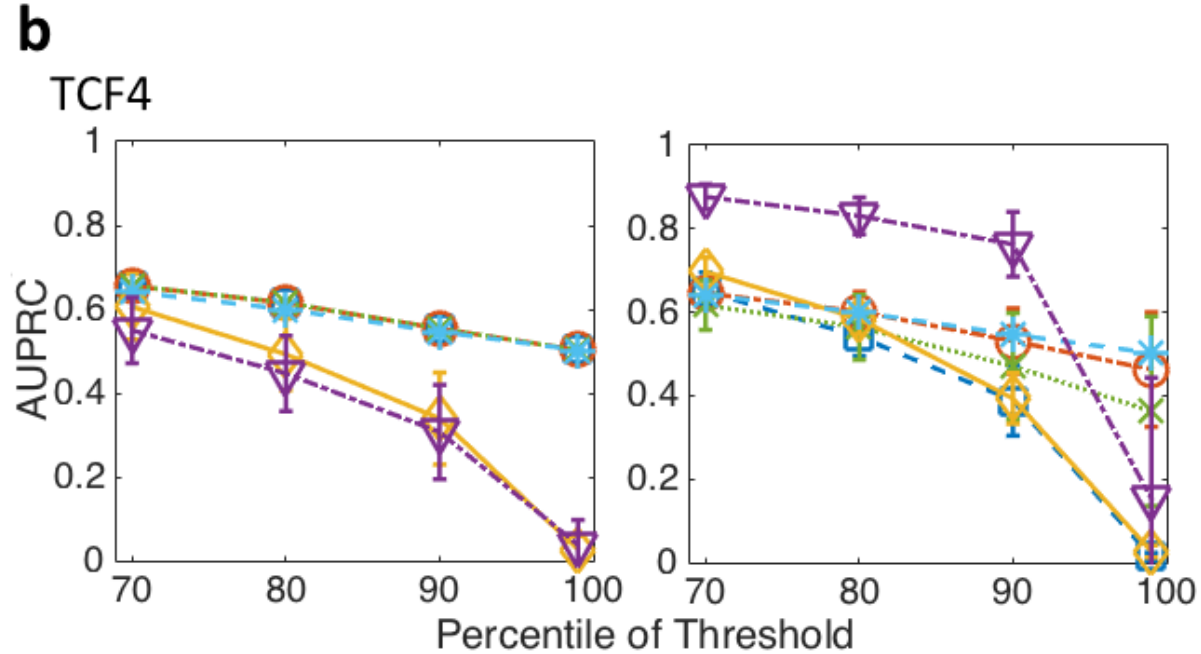
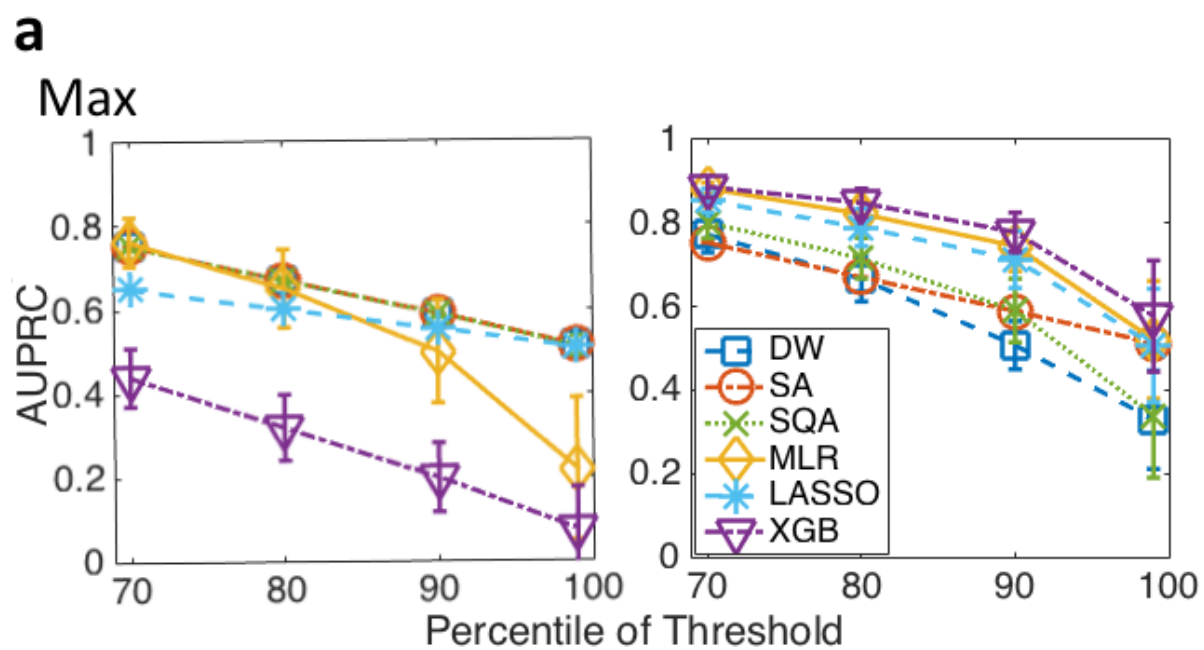
THANKS FOR LISTENING!

## Math sketch:

- $S$  is the signal distribution,  $B$  background,  $v$  is the variable
- $v_{low}$  and  $v_{high}$  are the 30th and 70th percentiles of  $S$ ,  $b_{low}$  and  $b_{high}$  the percentiles on  $B$  at those values
- If  $b_{high} < 0.7$  then define  $v_{shift} = v_{low} - v$ , else if  $b_{low} > 0.7$  then  $v_{shift} = v - v_{high}$ , else reject  $v$
- Define  $v_{+1}$  and  $v_{-1}$  as the 10th and 90th percentile of the transformed  $S$  distribution
- With this formulation, the weak classifier is given by
- Do this for all the variables and products (or, if flipped flipped, the ratio)



$$h(v) = \begin{cases} +1 & \text{if } v_{+1} < v^{shift}(v) \\ \frac{v^{shift}(v)}{v_{+1}} & \text{if } 0 < v^{shift}(v) \leq v_{+1} \\ \frac{v^{shift}(v)}{|v_{-1}|} & \text{if } v_{-1} < v^{shift}(v) \leq 0 \\ -1 & \text{if } v^{shift}(v) < v_{-1} \end{cases}$$



Training with 1%  
of training data

Training with 5%  
of training data

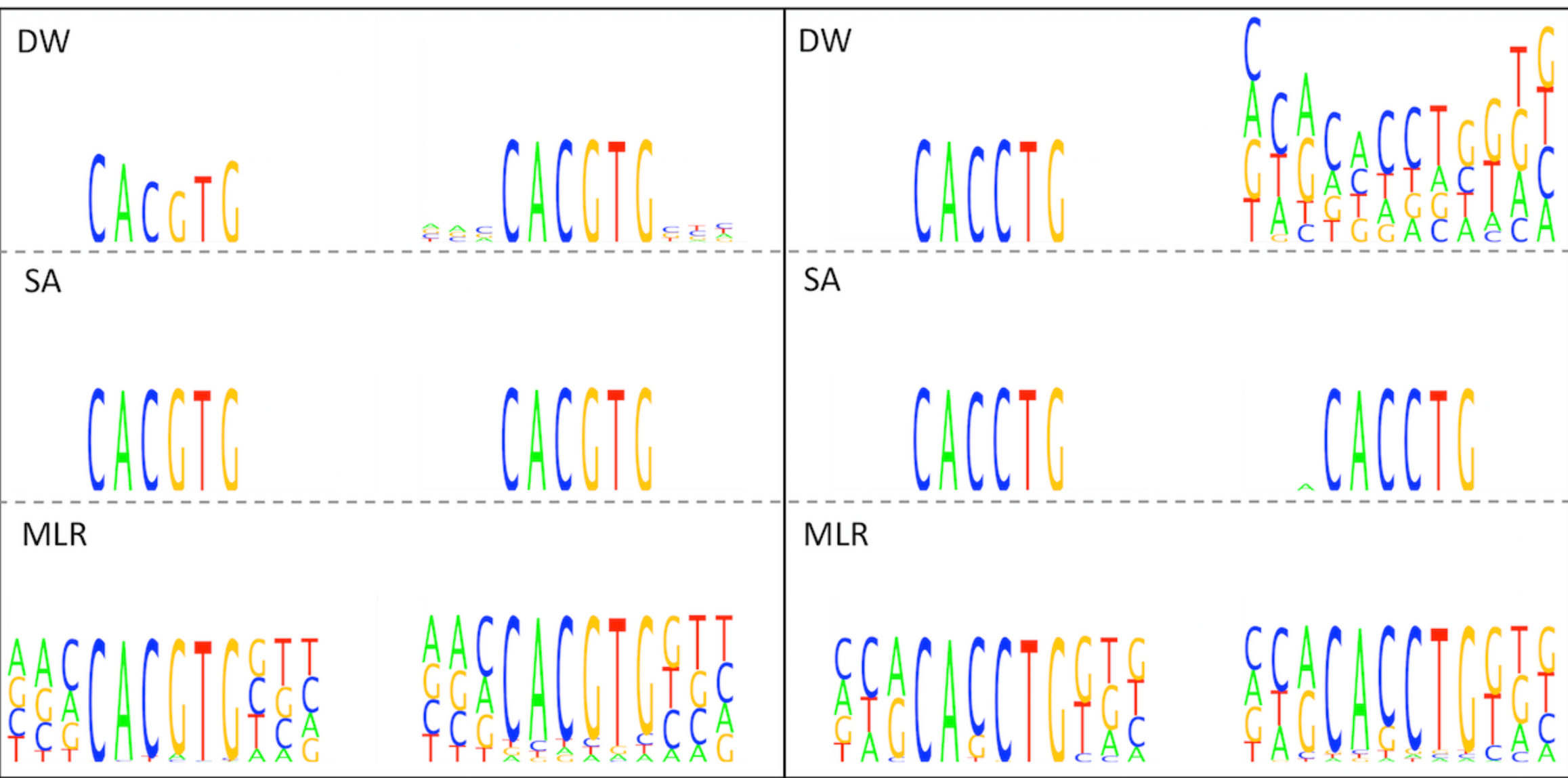
Training with 2%  
of training data

Training with 10%  
of training data

**c**

Max

TCF4



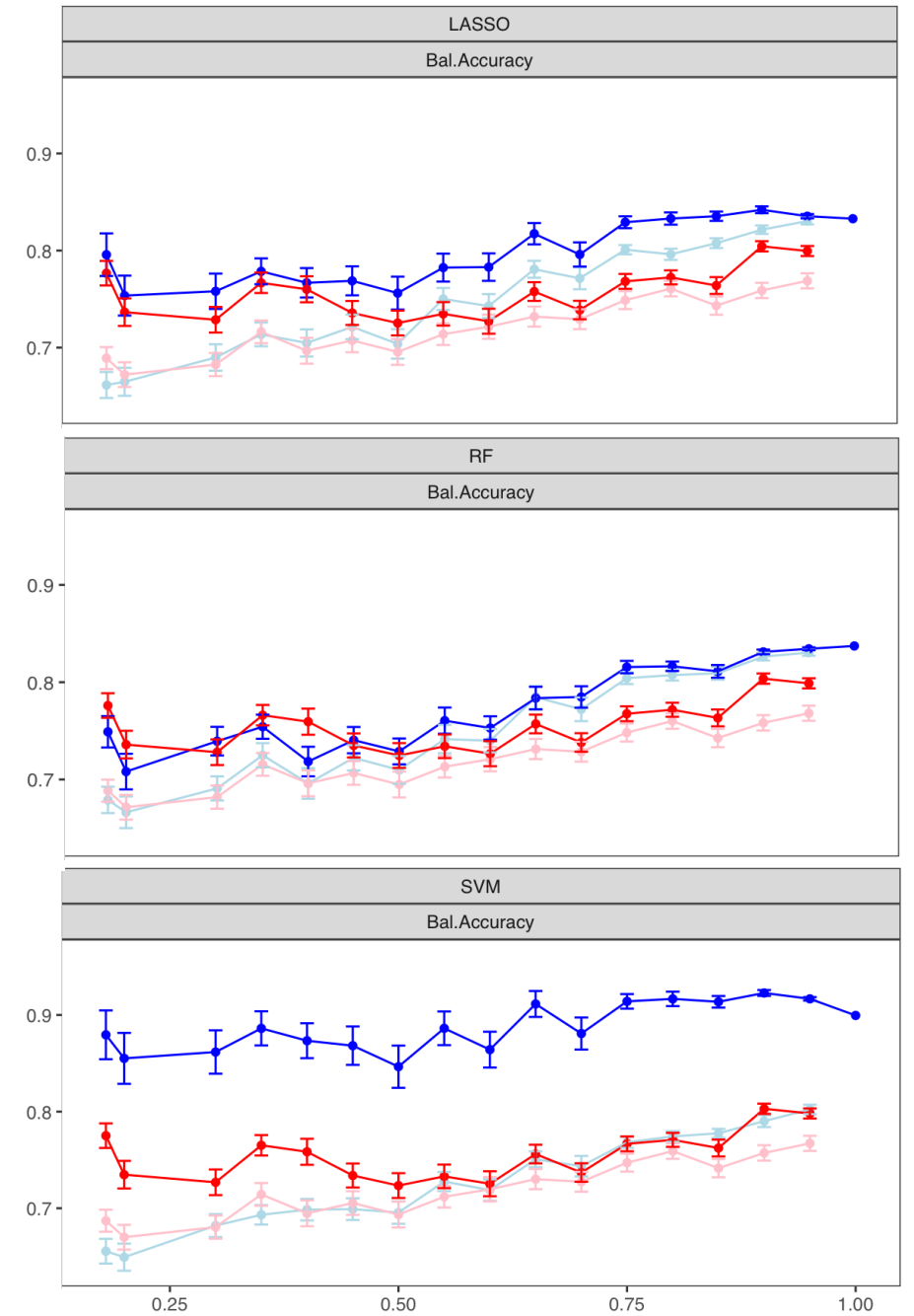
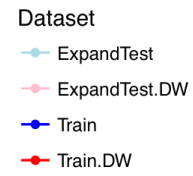
Training with 1%  
of training data

Training with 5%  
of training data

Training with 2%  
of training data

Training with 10%  
of training data

# Results





YOU'RE TRYING TO PREDICT THE BEHAVIOR OF <COMPLICATED SYSTEM>? JUST MODEL IT AS A <SIMPLE OBJECT>, AND THEN ADD SOME SECONDARY TERMS TO ACCOUNT FOR <COMPLICATIONS I JUST THOUGHT OF>.

EASY, RIGHT?

SO, WHY DOES <YOUR FIELD> NEED A WHOLE JOURNAL, ANYWAY?



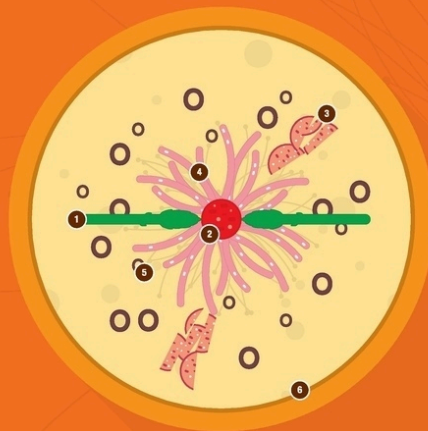
LIBERAL-ARTS MAJORS MAY BE ANNOYING SOMETIMES BUT THERE'S NOTHING MORE OBNOXIOUS THAN A PHYSICIST FIRST ENCOUNTERING A NEW SUBJECT

<https://xkcd.com/793/>

## Higgs Boson Pizza Day at CERN

What's happening on my Ham & Cheese pizza?

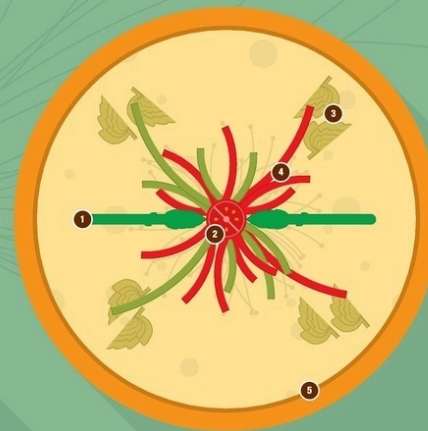
A two asparagus (proton-proton) collision produces a spicy Higgs boson (chorizo) decaying into two high-energy salami (photon) clusters and a lot of charged (sliced ham) and neutral (olive) particles that are detected in the pizza (detector) entirely covered with mozzarella sensors.



- Asparagus Proton
- Chorizo Higgs boson
- Salami Photon
- Ham Charged particle
- Olive Neutral particle
- Cheese Detector

What's happening on my Vegetarian pizza?

A two asparagus (proton-proton) collision produces a juicy Higgs boson (cherry tomato) decaying into four high-energy peppers producing a tasty signal in the artichoke (muon) chambers and a lot of charged (red and green peppers) particles that are detected in the pizza (detector) entirely covered with mozzarella sensors.



- Asparagus Proton
- Cherry tomato Higgs boson
- Artichoke Muon
- Pepper Charged particles
- Cheese Detector

Original recipe discovered in Naples by Fernando Ferroni and Pierluigi Paolucci, INFN.

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.



<https://xkcd.com/1838/>