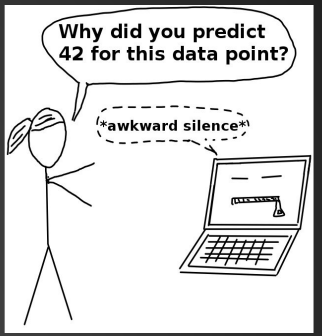
Learning Feature Importance for a Deep Learning Cancer Classifier

Jacob Chmura
Wednesday, July 31, 2019
Supervisor: Quaid Morris
University of Toronto Undergraduate Summer Research Program

Lack of Interpretability in Deep Learning is a Limitation



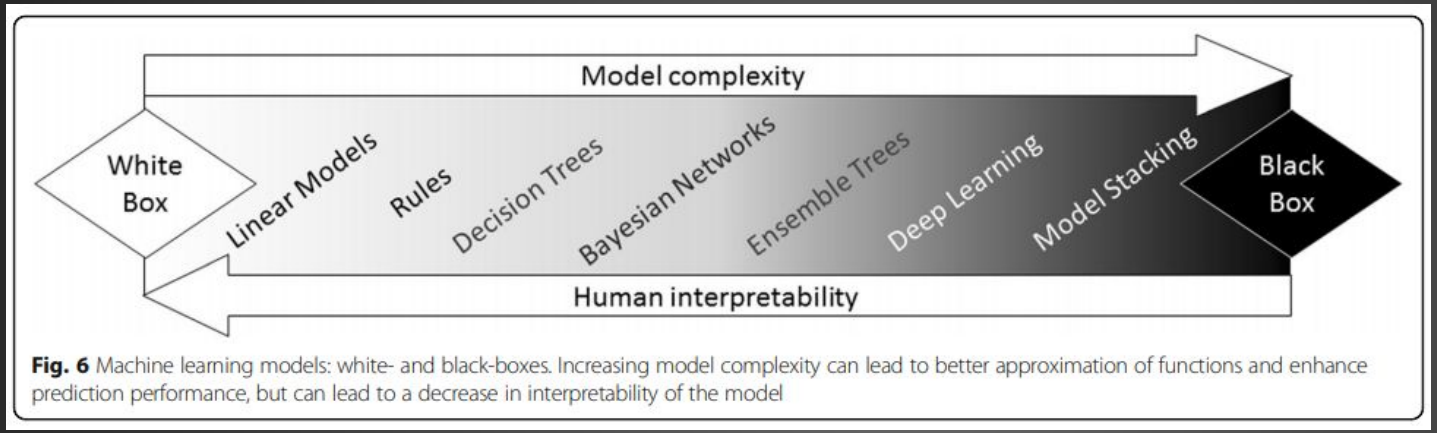


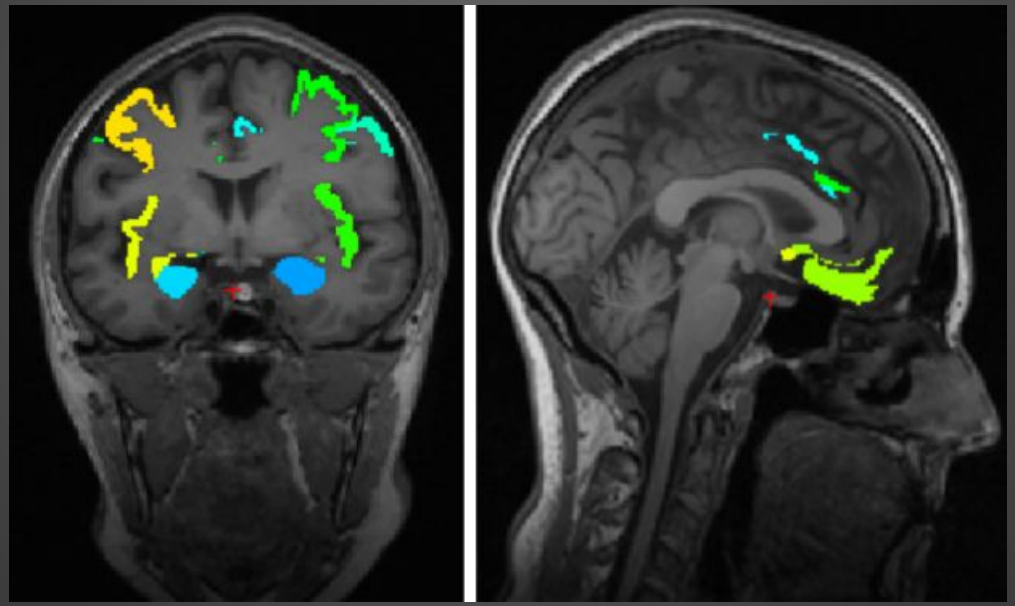
Fig. 6 Machine learning models: white- and black-boxes. Increasing model complexity can lead to better approximation of functions and enhance prediction performance, but can lead to a decrease in interpretability of the model

https://onehealth.ifas.ufl.edu/media/onehealthifasufledu/pdfs/publications/Prosperi-article.pdf

- *Object recognition network*: could tell us which pixels of the image responsible for a label being picked

Lack of Interpretability in Deep Learning is a Limitation

- *Medical imaging model*: could help inform the doctor of the part of the image that resulted in the recommendation. Knowing the strengths and weaknesses of a model is essential in clinical settings.



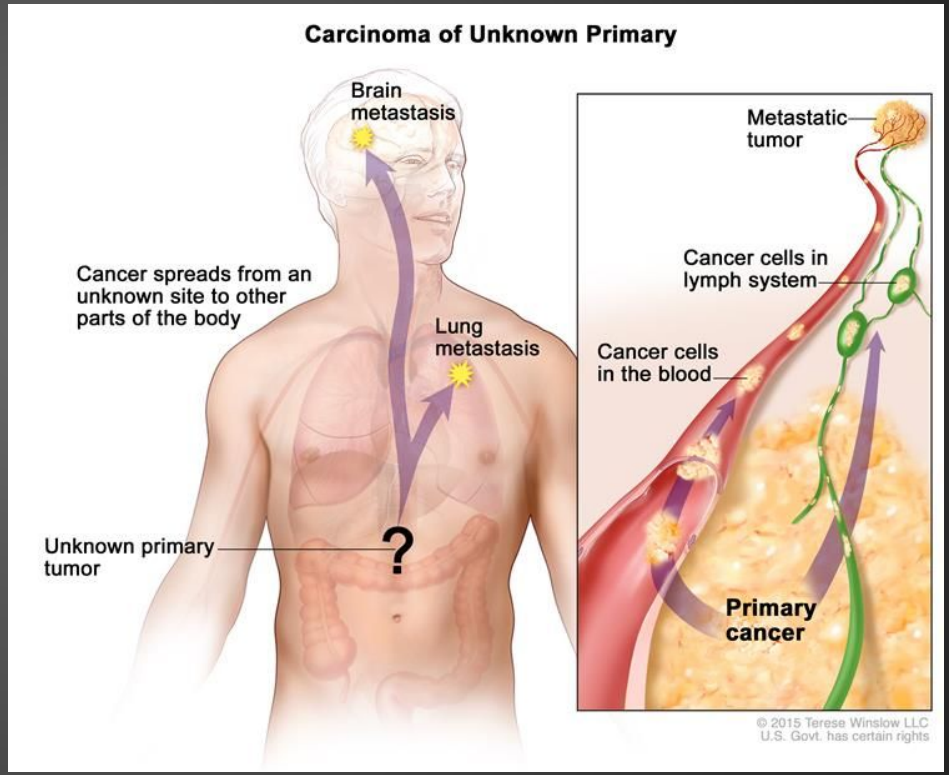
Classifying Primary and Metastatic Cancers based on Mutation Patterns using Deep Learning Techniques

Situation:

- pathologist can't identify primary cancer tumor well
- at autopsy, the primary cannot be identified roughly 70% of the time
- fourth most common cause of cancer death
- therapeutic options are driven by tissue of origin

Idea:

- Different tumour types have dramatically different patterns of mutation
- Use this to train a deep network to classify cancer classes



https://www.cancer.gov/types/unknown-primary/patient/unknown-primary-treatment-pdq

Features for the Model

~ 5000 length feature vector split into three main categories:

(1) *Mutation Distribution (~3000)*: encode information about cell type

- divided genome into bins and created features corresponding to the number of mutations per bin

Wild type gene	C G A C T G G C T G A C
Transition (AT pair replaced by GC pair)	C G **G** C T G G C **C** G A C
Transversion (AT pair replaced by TA pair)	C G **T** C T G G C **A** G A C
Insertion (GC pair inserted)	C G A **G** C T G G C T **C** G A C
Deletion (AT pair deleted)	C G C T G G C G A C

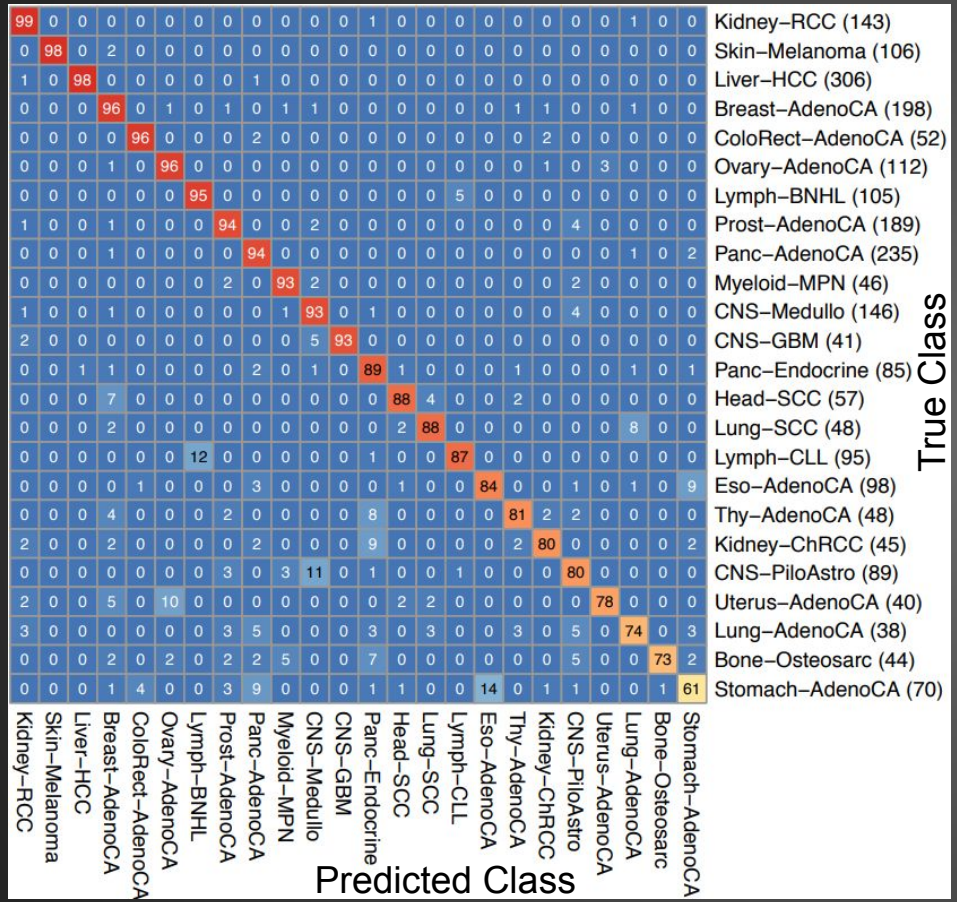
http://www.biochem.uthscsa.edu/med/06-Mechanisms-of-Mutation/PrereqMechanismsofMutation_print.html

(2) *Mutation Types (~150)*: reflect environmental exposures of the cell of origin

- Ex. skin cancers have mutation types strongly correlated with UV light-induced DNA damage.

(3) *Driver Genes (~2000):* tumour types are distinguished by high frequencies of alterations in particular driver genes and pathways

Results



- classifier achieves an accuracy of 91%
- most frequent classification errors for Stomach-AdenoCA samples were two other upper gastrointestinal tumours, (Eso-AdenoCA and Panc-AdenoCA)

(Wei Jiao et al., 2019)

Goals

- *Aim 1:* take existing model and assess feature importance

- *Aim 2*: produce certainty estimates and extend to rare cancer types

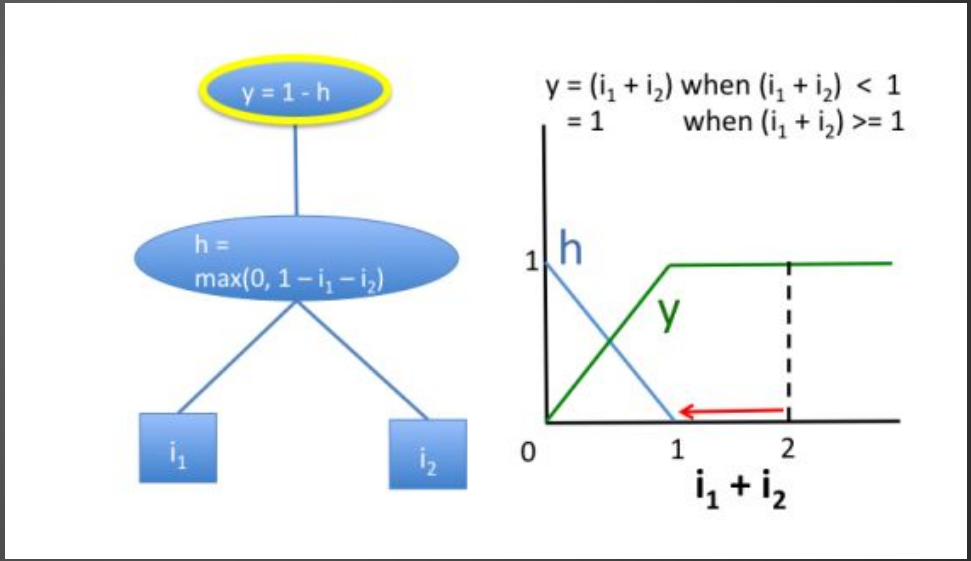
How do we learn which features are important?

Idea 1: *Perturbations*?

- Make small changes to individual inputs and observe the impact on later neurons in the network

Problems:

- computationally inefficient
- **Saturation problem**



(Avanti Shrikumar, 2017)

***Saturation problem**:* lack of local change need not imply zero importance

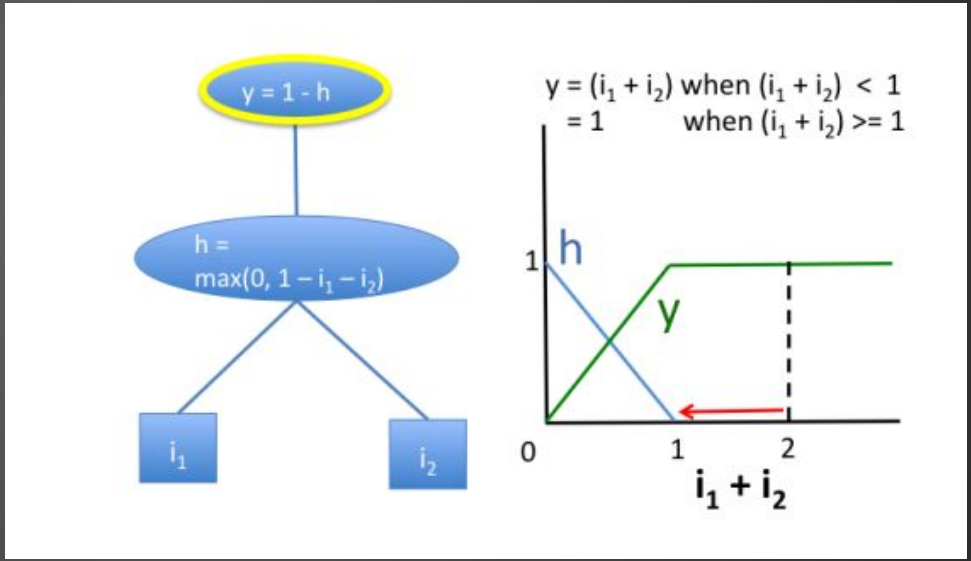
How do we learn which features are important?

Idea 2: *Gradients*?

- Compute the gradient of the outputs with respect to each input feature

Problems:

- nonlinearities
- **Saturation problem**



(Avanti Shrikumar, 2017)

***Saturation problem**:* lack of local change need not imply zero importance

DeepLIFT (Deep Learning Important FeaTures) *(Avanti Shrikumar, 2017)*

Philosophy: explain the difference in output from some ‘reference’ output in terms of the difference of the input from some ‘reference’ input.

reference input: represents default or *neutral* baseline that is chosen

- references for all neurons can be found by choosing a reference input and propagating activations through the net

- t represents target output neuron of interest
- $x_1, x_2, ..., x_n$ represent neurons in some intermediate layer
- t_0 represent the reference activation of t.
- Δt is difference-from-reference, that is $\Delta t = t - t_0$.

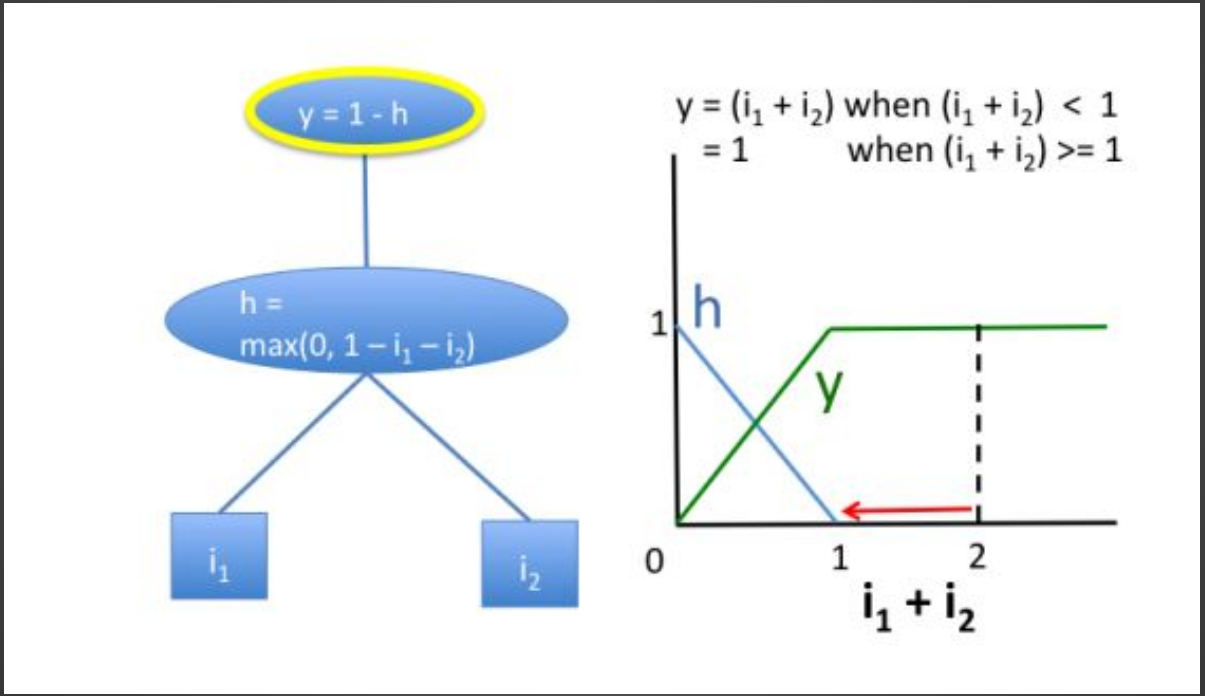
$$\sum_{i=1}^{n} C_{\Delta x_i \Delta t} = \Delta t$$

DeepLIFT assigns contribution scores:

- $C_{\Delta x_i \Delta t}$ is the amount in t that is *blamed* on difference-from-reference of x_i .

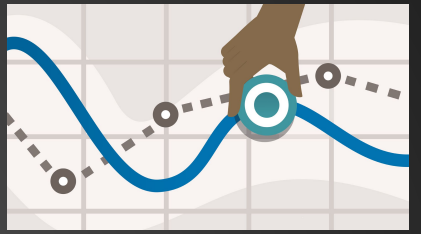
Saturation Problem Revisited

- C∆xi∆t can be non-zero even when ∂t / ∂xi is zero: a neuron can be signaling meaningful information even when its gradient is zero.



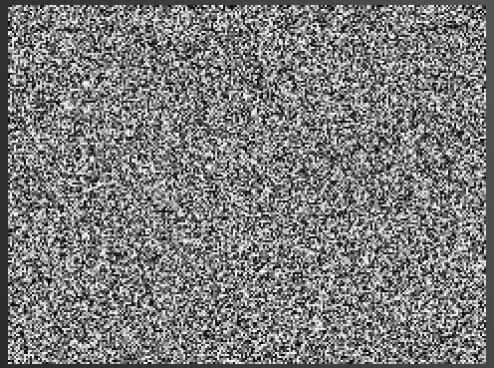
- DeepLIFT is sensitive to chosen reference

Important Question: *How do we choose a good reference?*



Intuition: ask ourselves *what am I interested in measuring differences against?*

- needs to convey a complete absence of signal (allows us to interpret the attributions as a function of the input)



black image or noisy image signifies the absence of objects

zero embedding vector is a good baseline

Choosing a reference for Cancer Classifier

~ 5500 length feature vector split into three main categories:

Mutation Distribution (~3000)	Mutation Type (~150)	Drivers (~2000)

★ Typical Feature Vector is extremely sparse

Null Reference

- A vector of all zeros

0 0 0 … 0 0 0	0 0 0 … 0 0 0	0 0 0 ... 0 0 0

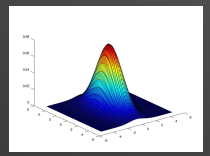
Shuffle Reference

- permute within the three main sections of the input vector and averaging the results over multiple such references

1.1 -1.6 … 0.7 0.2	0.0 0.5 … 1.1 -0.7	2.9 1.3 … 2.1 0.2

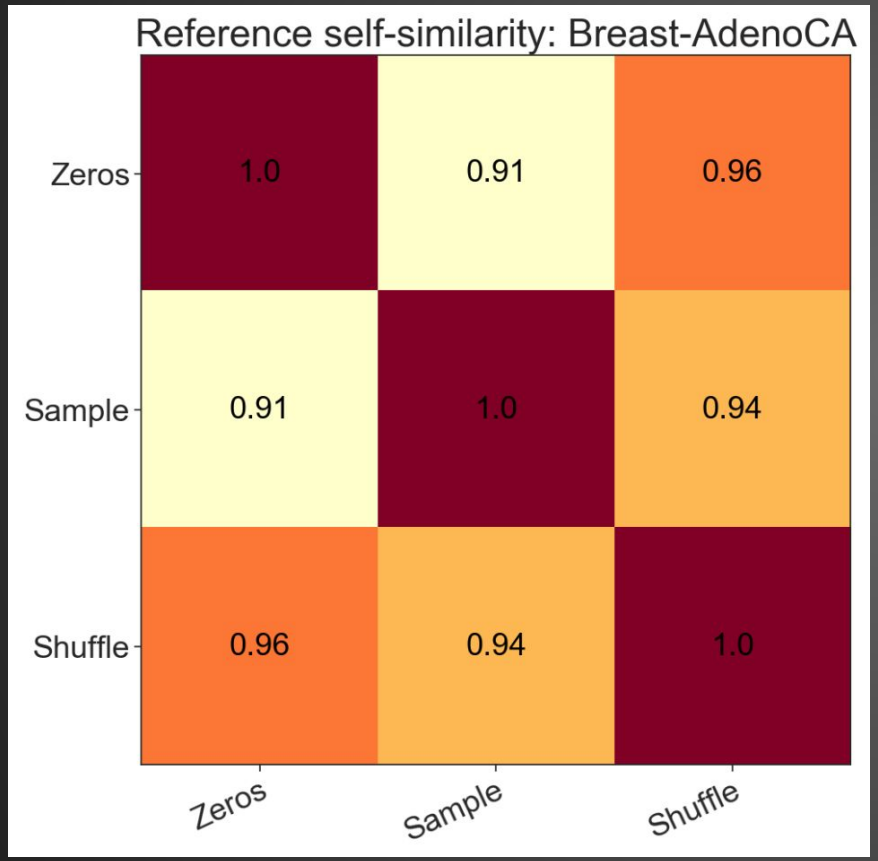
Sample Reference

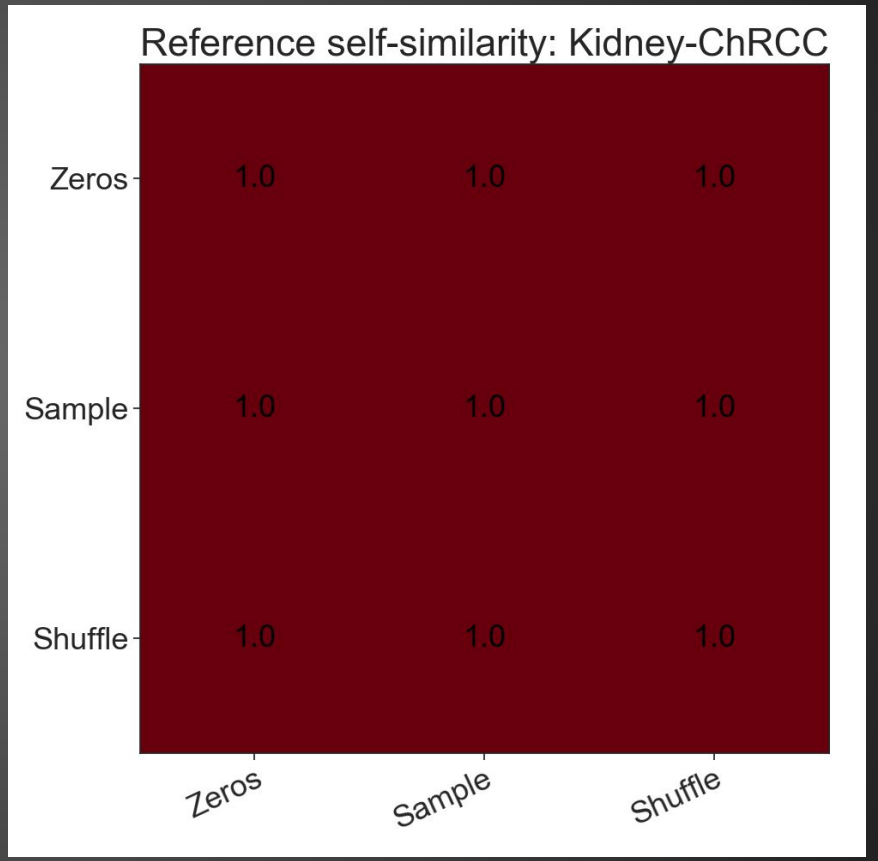
- For each z score feature, simply sample from a Gaussian
- For count features, create a distribution of frequencies and sample from it



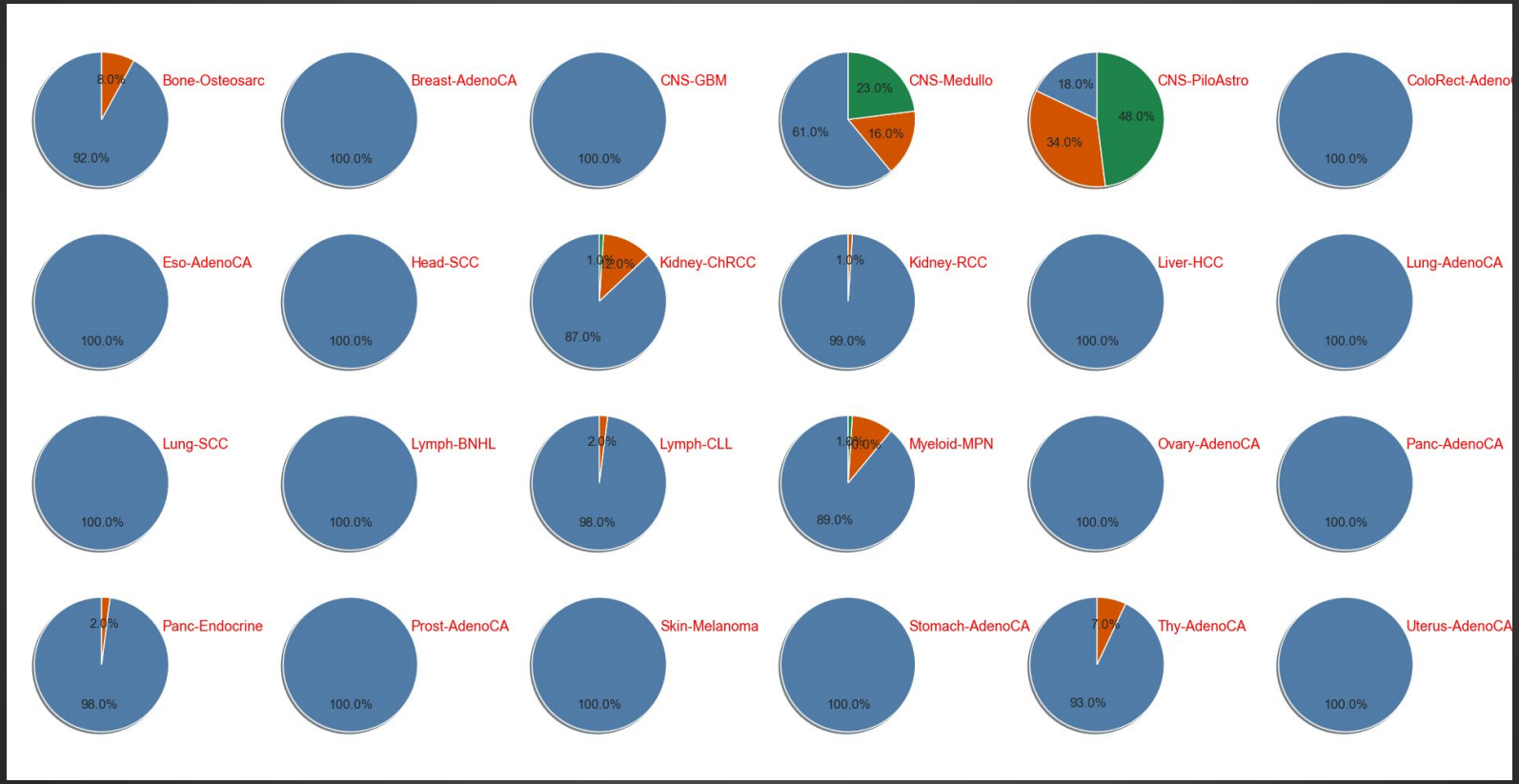
2.1 -0.6 … -1.7 0.1	0.0 3.5 … 0.1 0.7	1.9 1.3 … 2.1 0.2

Reference Self-Similarity Matrix

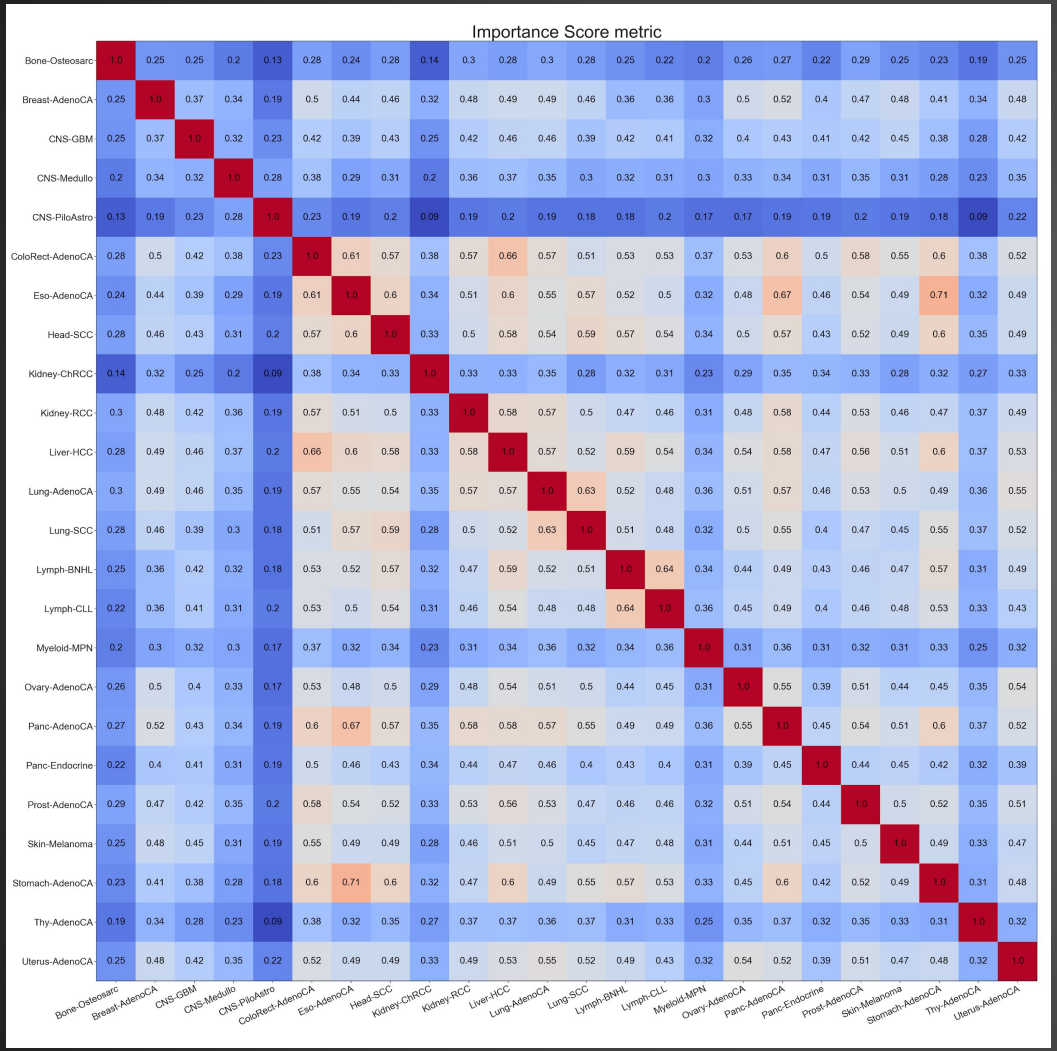




Feature Type Distribution over 100 Most Significant Inputs



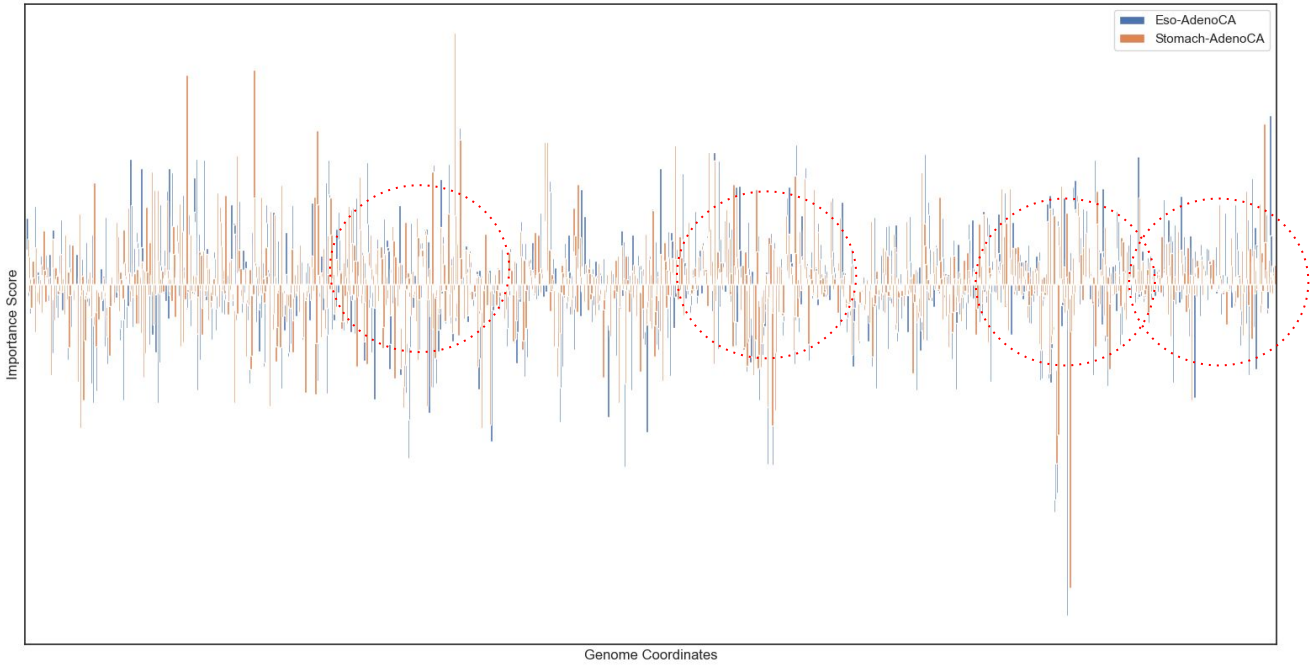
Retraining without driver features, the classifier does just as good



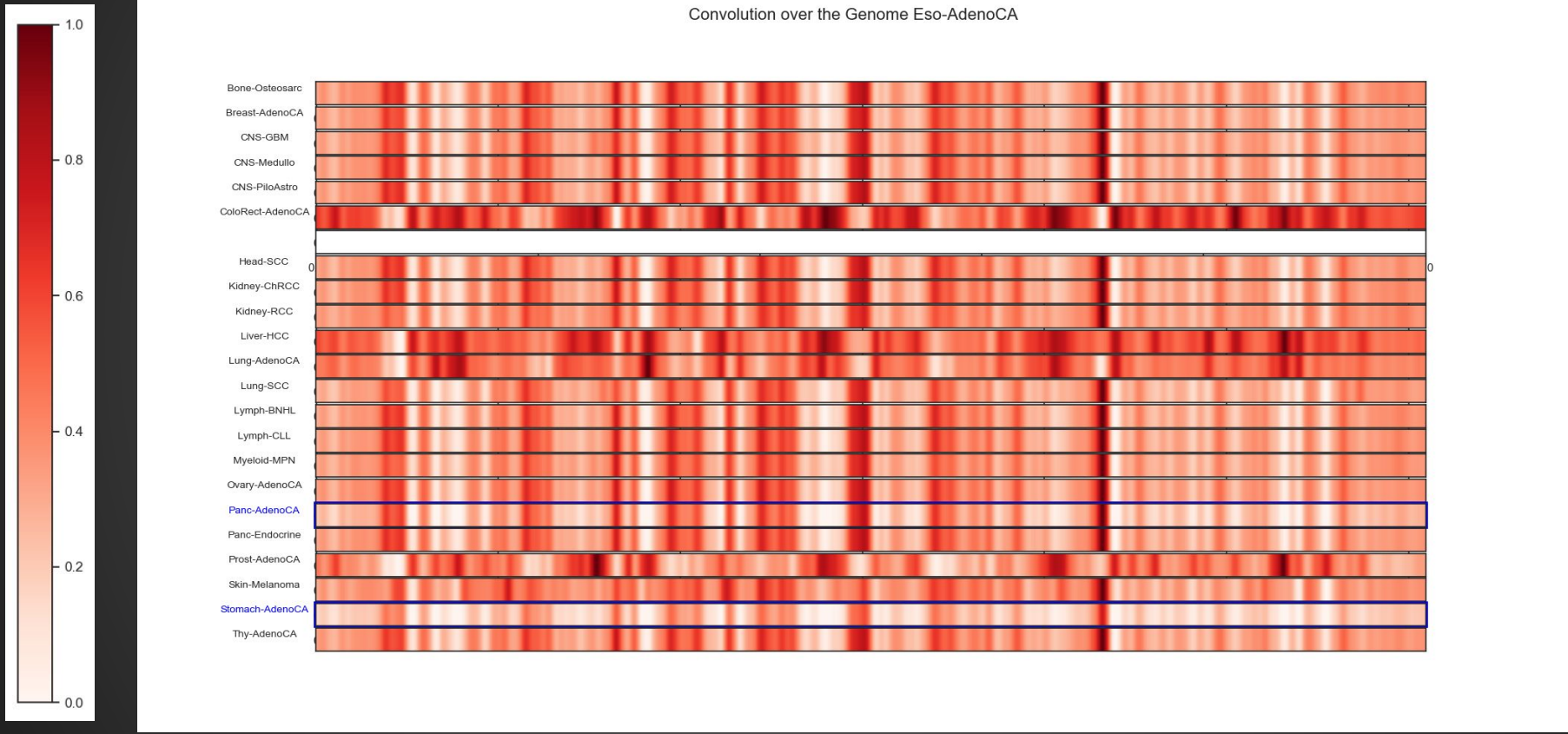
Cancer class Self-Similarity over 100 Most Significant Input Features

- Similarity score computed using adaptation of Jaccard Index over feature importance
- No drivers
- Average of all 3 references

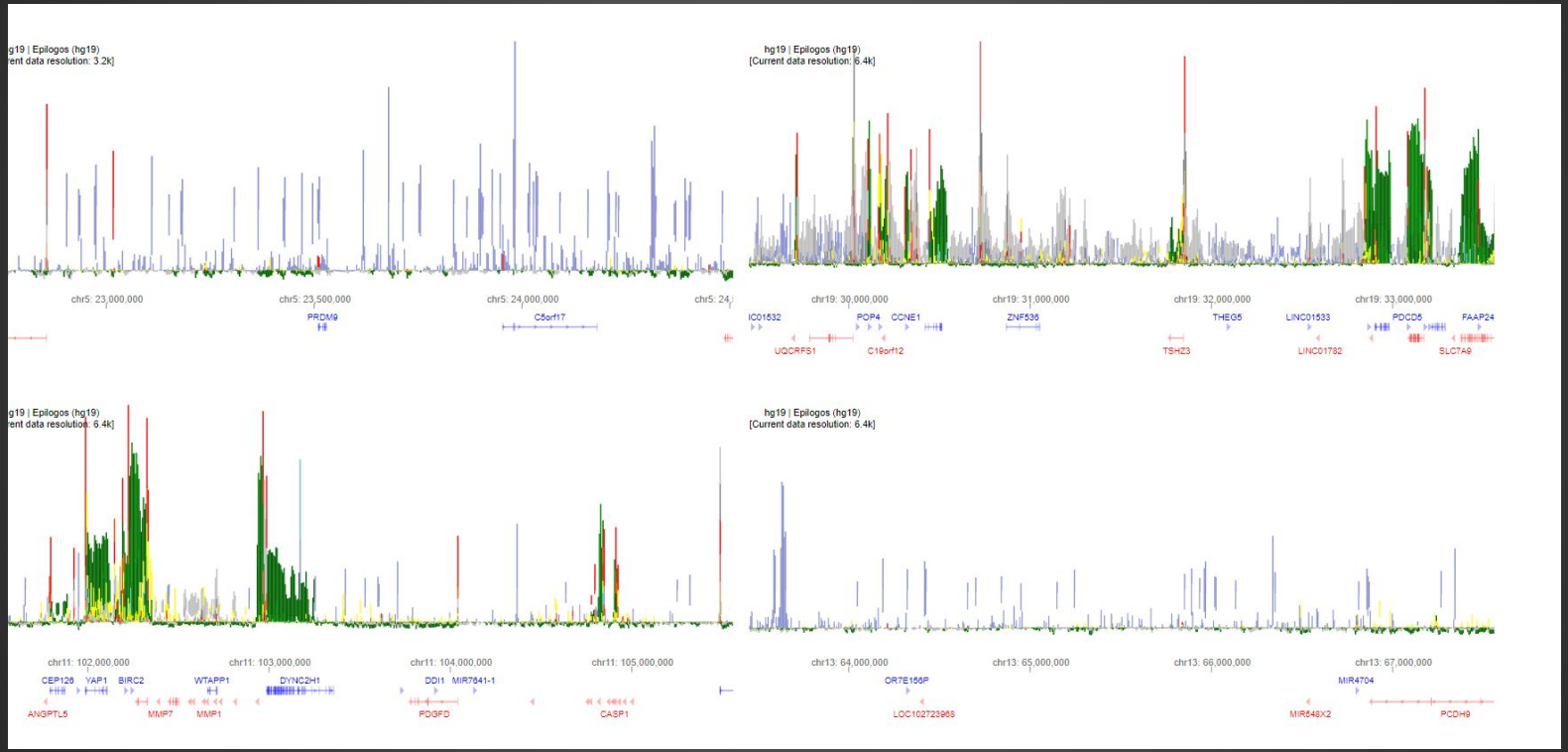
Genome Wide Mutation Importance: Chromosome 11-15



Convolution of Importance Scores Across Genome

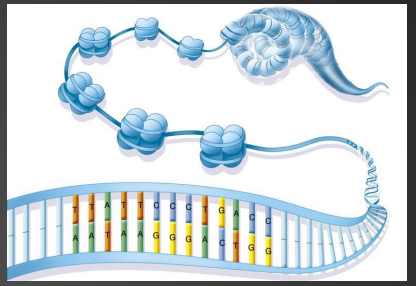


Genome Annotation Process: *Extract genes from most important DeepLIFT scoring genome regions*



Bone-Osteosarc Important Regions

Chromatin States Extraction Controlled for Cell Type



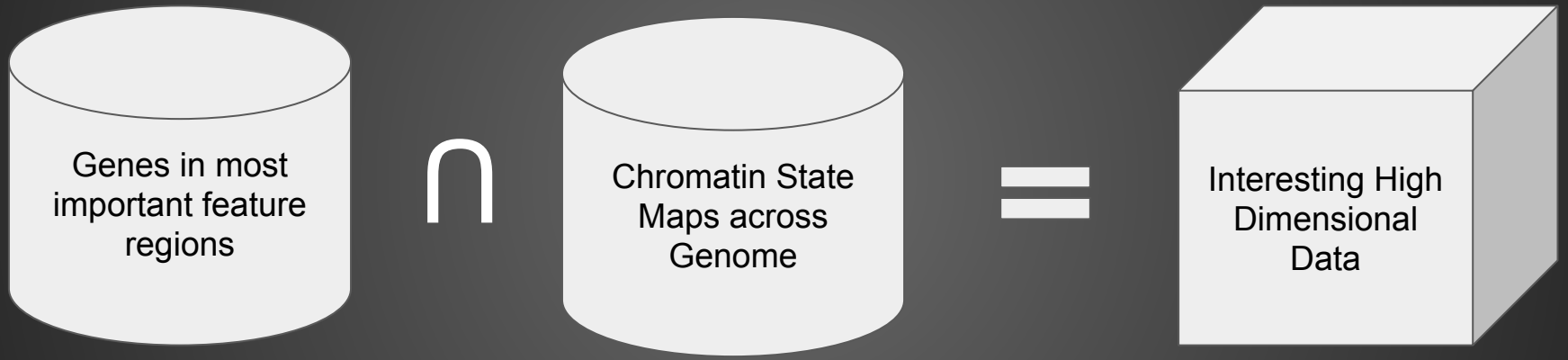
- Chromatin is understood to be a complex genome organizer
- Map chromatin function along the genome to discrete states

Does it transcribe?

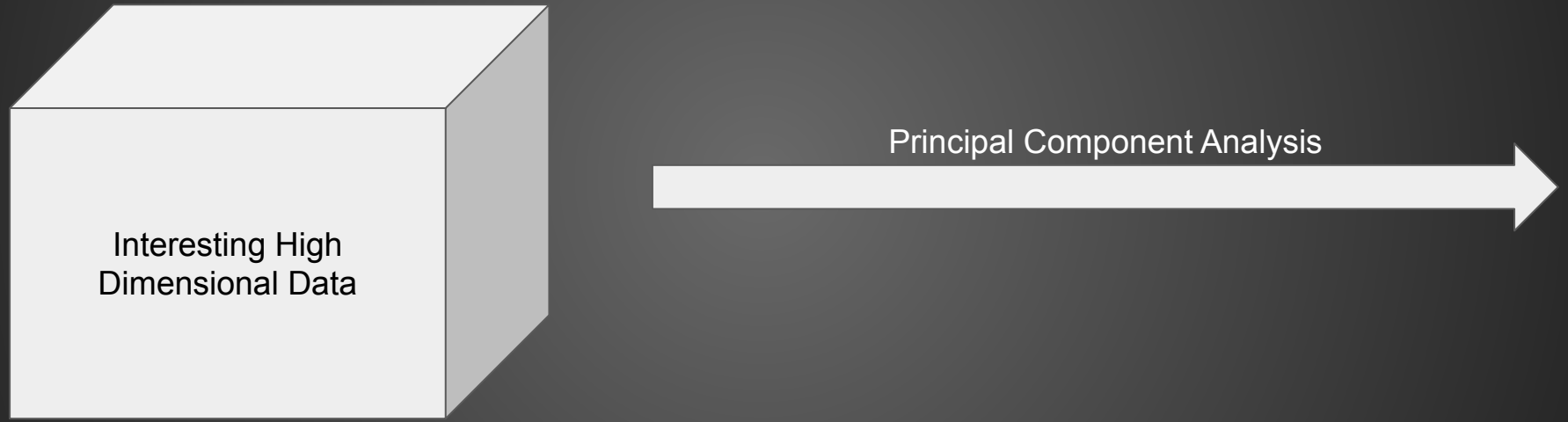
STATE NO.	MNEMONIC	DESCRIPTION	COLOR NAME	COLOR CODE
1	TssA	Active TSS	Red	255,0,0
2	TssAFlnk	Flanking Active TSS	Orange Red	255,69,0
3	TxFlnk	Transcr. at gene 5' and 3'	LimeGreen	50,205,50
4	Tx	Strong transcription	Green	0,128,0
5	TxWk	Weak transcription	DarkGreen	0,100,0
6	EnhG	Genic enhancers	GreenYellow	194,225,5
7	Enh	Enhancers	Yellow	255,255,0
8	ZNF/Rpts	ZNF genes & repeats	Medium Aquamarine	102,205,170
9	Het	Heterochromatin	PaleTurquoise	138,145,208
10	TssBiv	Bivalent/Poised TSS	IndianRed	205,92,92
11	BivFlnk	Flanking Bivalent TSS/Enh	DarkSalmon	233,150,122
12	EnhBiv	Bivalent Enhancer	DarkKhaki	189,183,107
13	ReprPC	Repressed PolyComb	Silver	128,128,128
14	ReprPCWk	Weak Repressed PolyComb	Gainsboro	192,192,192
15	Quies	Quiescent/Low	White	255,255,255

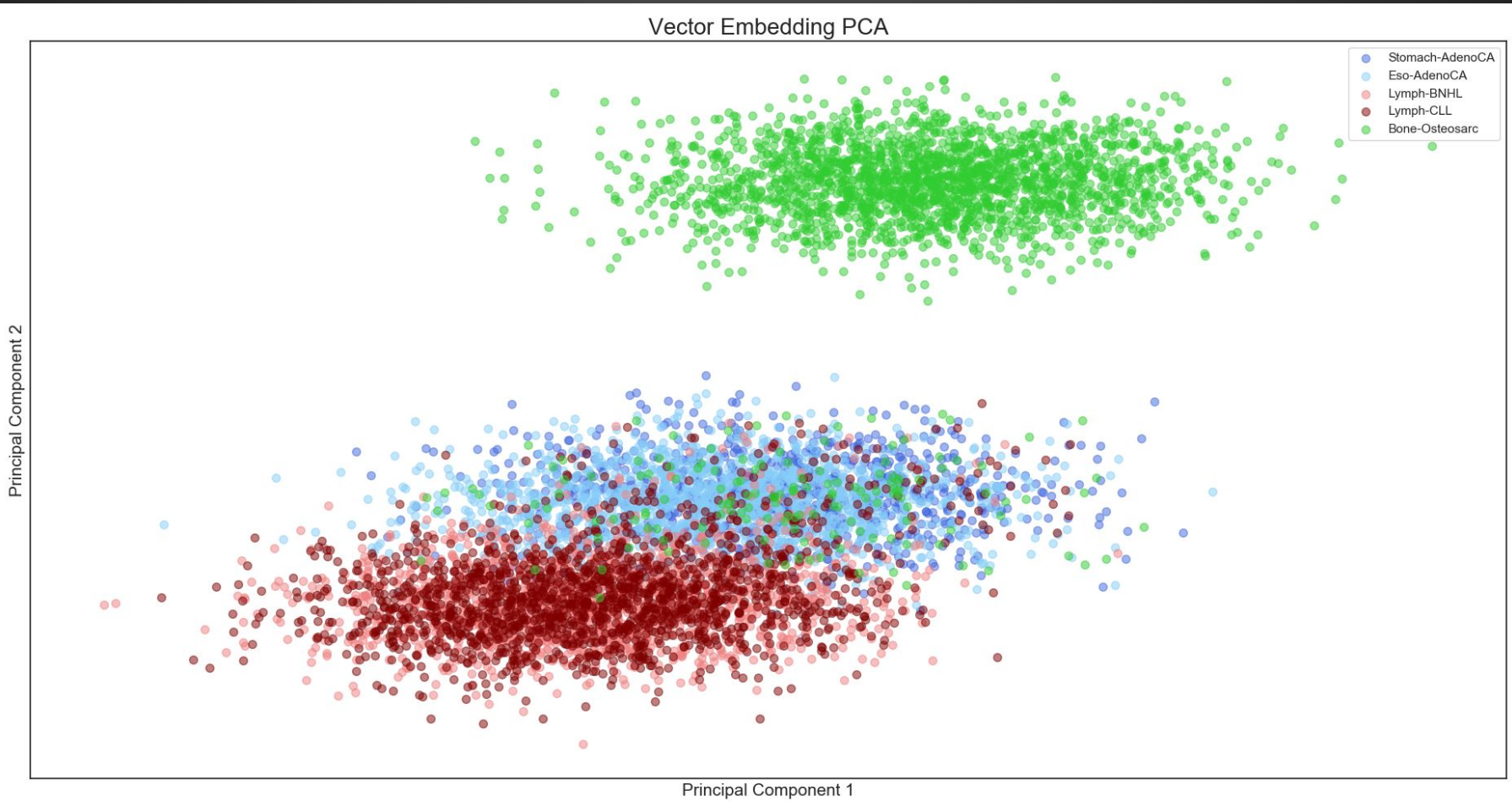
https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html

Vector Space Embedding: *We want to control for Suppressed Chromatin Markers*



Vector Space Embedding: *I want to see what this looks like ...*




Vector Embedding PCA
Stomach-AdenoCA
Eso-AdenoCA
Lymph-BNHL
Lymph-CLL
Bone-Osteosarc
Principal Component 2
Principal Component 1

Conclusions

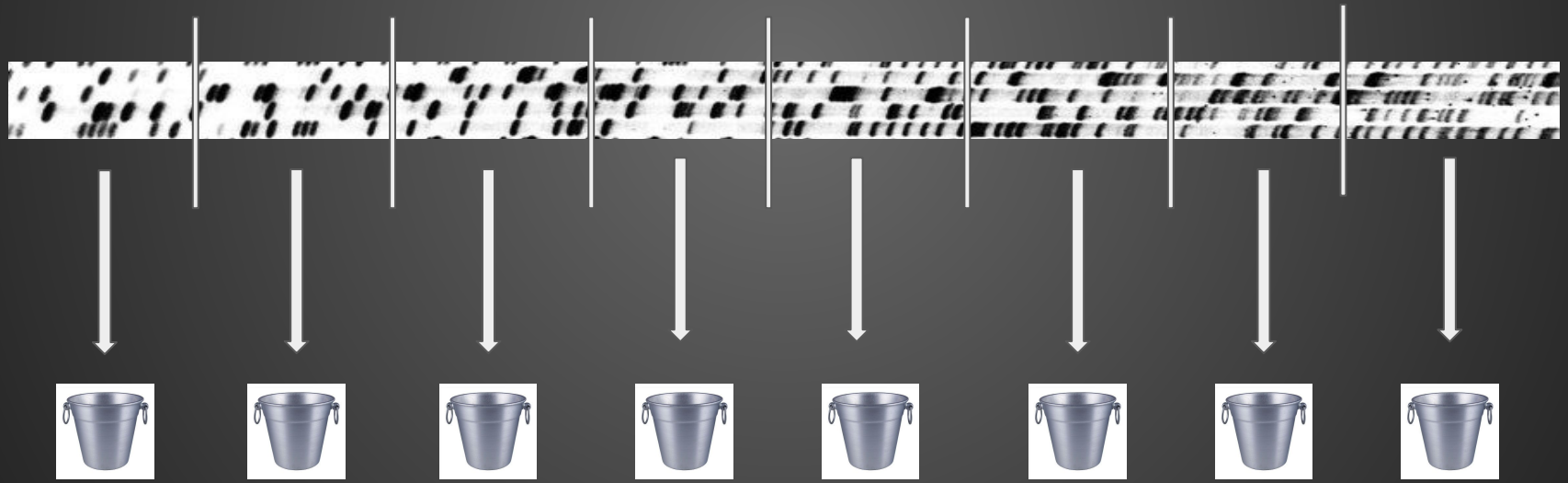
- Driver gene and pathway features don’t provide additional information for cancer classification. Training with mutation distribution and type is sufficient.
- Model misclassifications reflect shared biological characteristics in mutation topology.
- Discovered patterns in chromatin marks in misclassified cancers
- Research is fun
- I learned a lot of biology

Next Steps

Better genome segmentation

- Explore genome segmentations based on mutational process activity
- increase information content between mutational density and cell type

Naive Segmentation

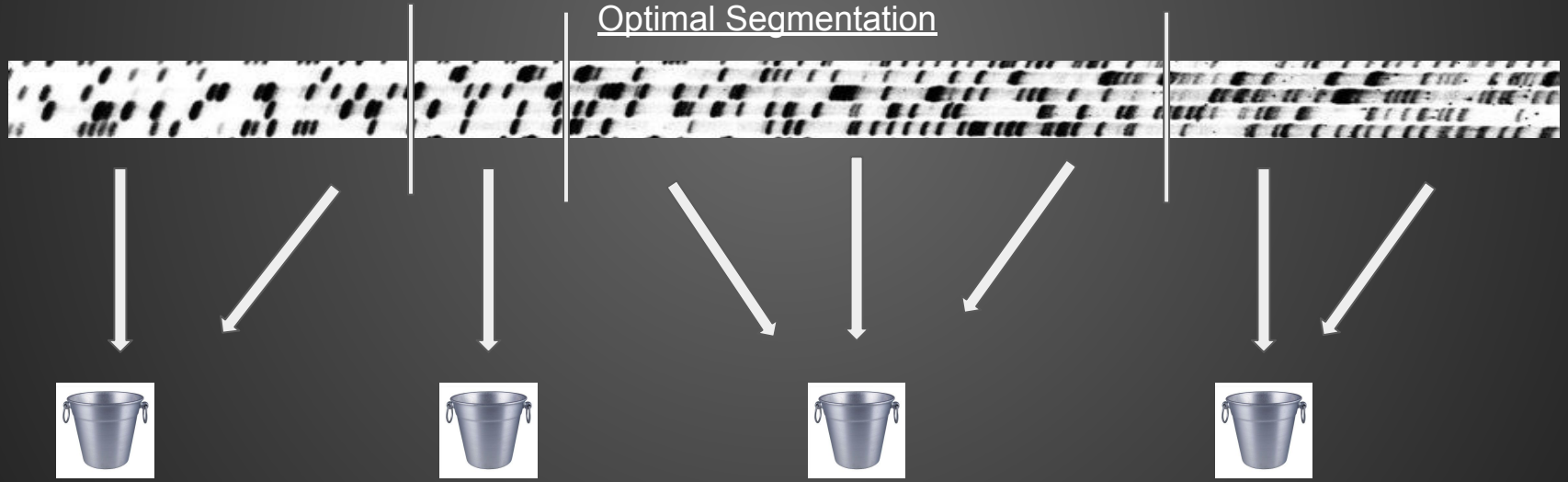


Next Steps

Better genome segmentation

- Explore genome segmentations based on mutational process activity
- increase information content between mutational density and cell type

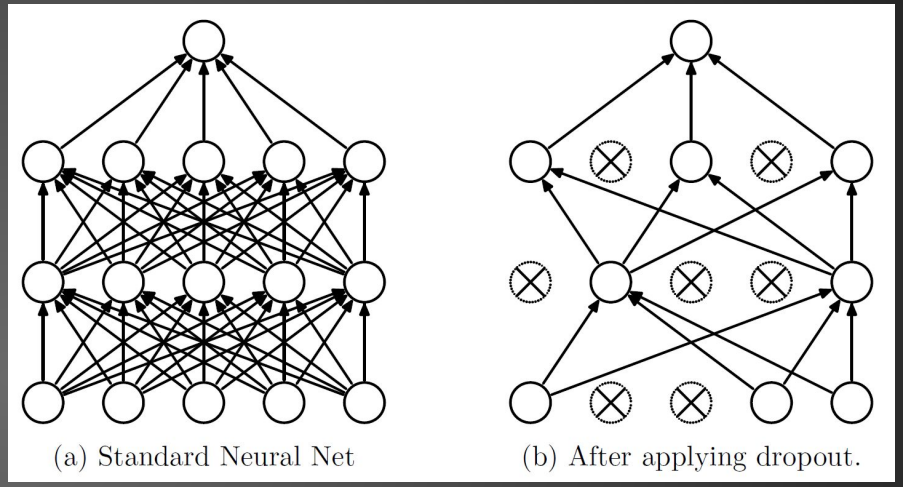
Optimal Segmentation



Next Steps

Extend classifier to rare cancer types

- Adopt a bayesian approach by implementing MC- dropout
- Certainty Estimates provide interpretable results

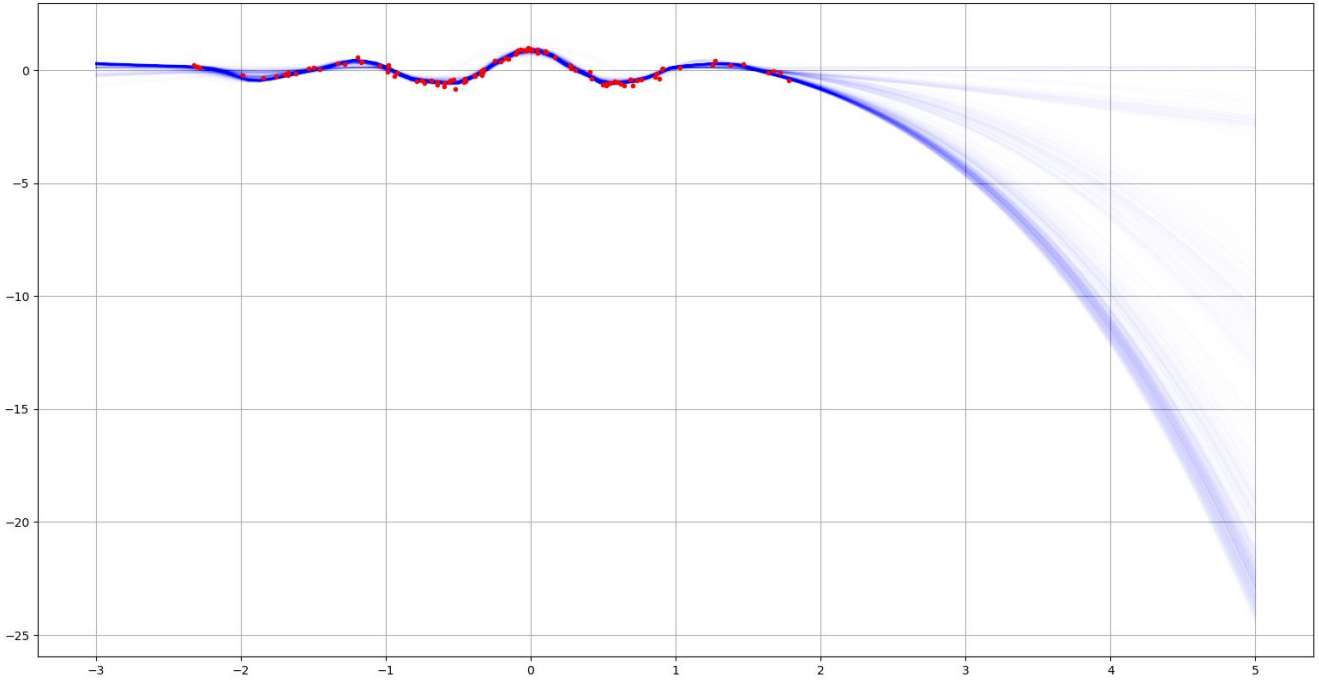


Source: Srivastava, Hinton, Krizhevsky, Sutskever and Salakhutdinov (2014)

MC-dropout (Monte Carlo Dropout)*: dropping neuron activations during test time ~ Non-deterministic*

- Generates predictions that one can interpret as samples from a probability distribution

Regression Example



Acknowledgements

Thanks to everyone in *The Morris Lab*:

Adamo Young
Alexander Sasse
Alinda Selega
Amir
Amit Deshwar
Audrina Zhou
Caitlyn
Chris
Chris Cole
Farzan Taj
Gurnit Atwal
Haoran Zhang
Ilyes Baali
Jarry Barber
Jeff Wintersinger
Jingping Qiao
Kaitlin laverty
Kate Nie
Kathryn Yu
Kim Skead
Linda
Nik Krosigk
Nil Sahin
Quaid Morris
Rozy Razavi
Yoonsik Park
Yulia Rubanova

Special thanks to Gurnit Atwal and Quaid Morris