Scripps Research
Translational Institute

AI for Genomics

Amalio Telenti

Scripps Research Translational Institute

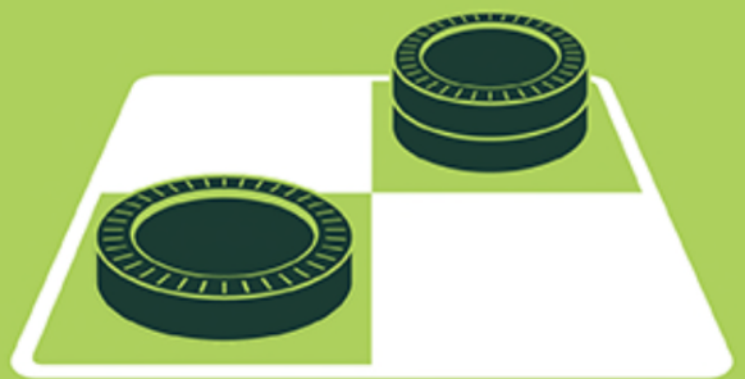
and Department of Integrative Structural and Computational Biology

The Scripps Research Institute

La Jolla, CA

ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



MACHINE LEARNING

Machine learning begins to flourish.

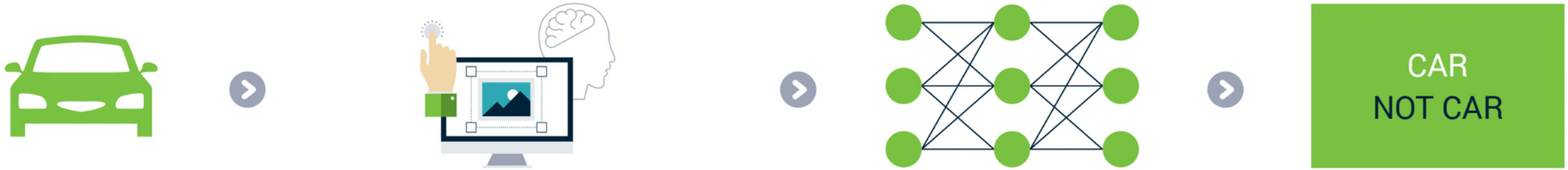
DEEP LEARNING

Deep learning breakthroughs drive AI boom.

1950's 1960's 1970's 1980's 1990's 2000's 2010's

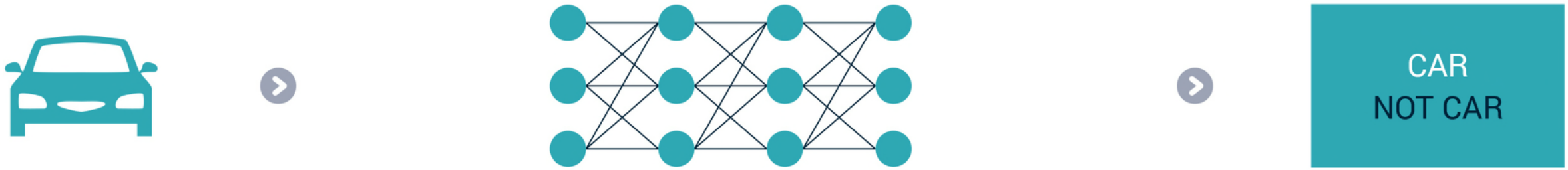
Nvidia

Machine Learning



Input Feature extraction Classification Output

Deep Learning



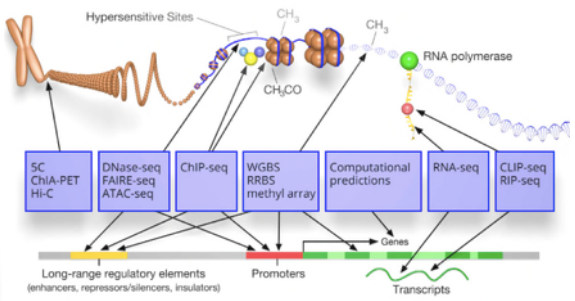
Input Feature extraction + Classification Output

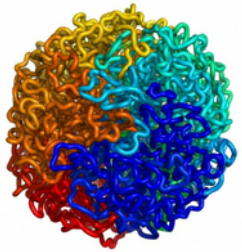
https://verhaert.com/wp-content/uploads/2018/05/Machine-vs-Deep-Learning-InputOutput.jpg

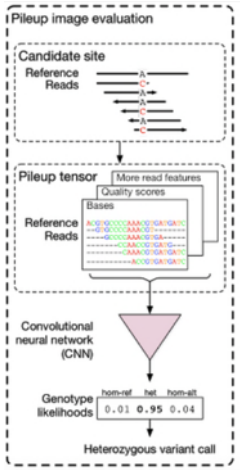
- Classify
- Explain
- Predict
- Establish Causality

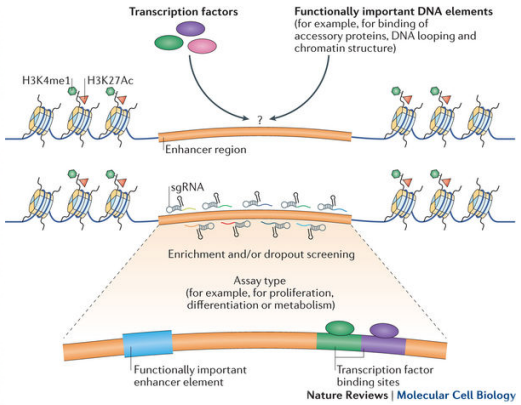
AI for genomics

- Human Genome
  - 2001
- Encode project
  - 2003
- GTEX
  - 2008
- Nucleosome 3D genome
  - ~2014
- Deep learning in genomics
  - Functional screens (CRISPR and parallel reporters)
  - ~2015
- Large scale deep sequencing of the human population
  - ~2016





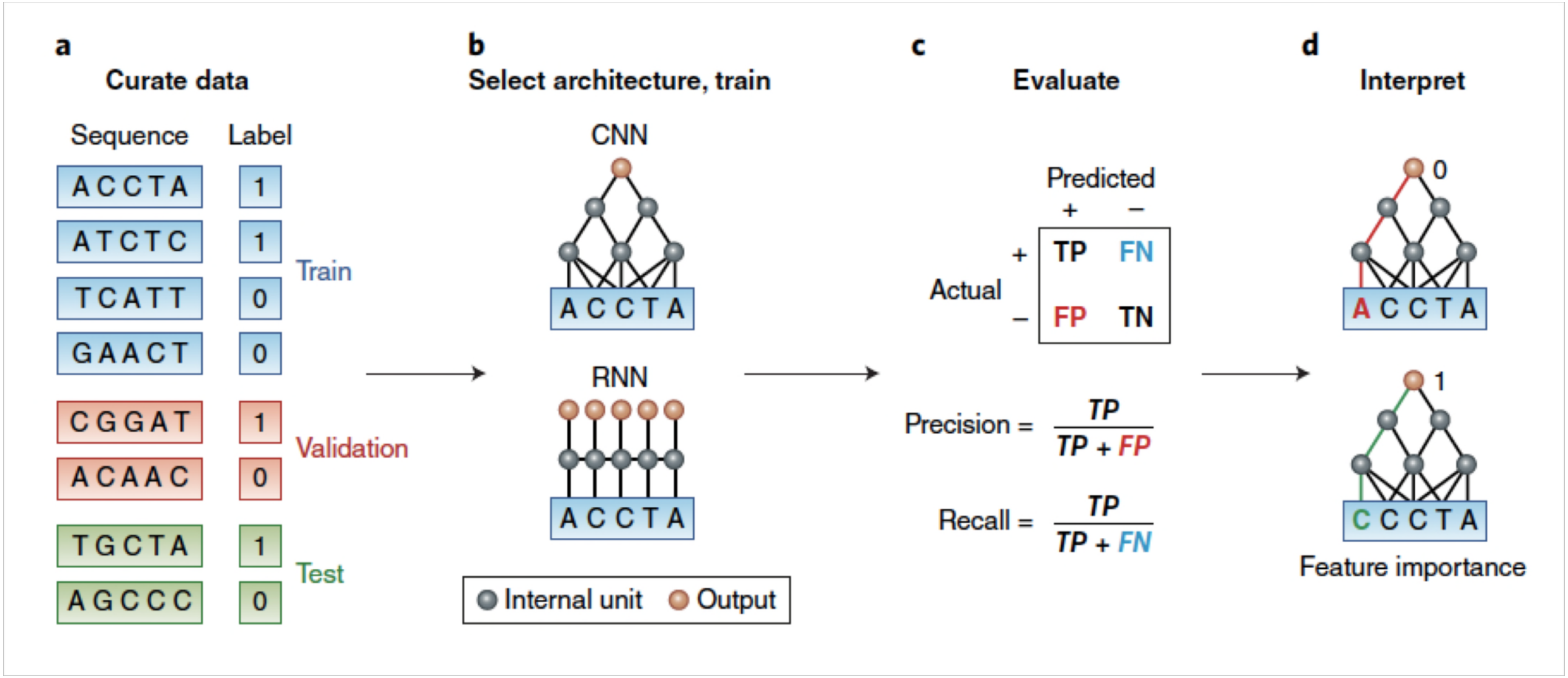




[gnomAD | genome Aggregation Database]

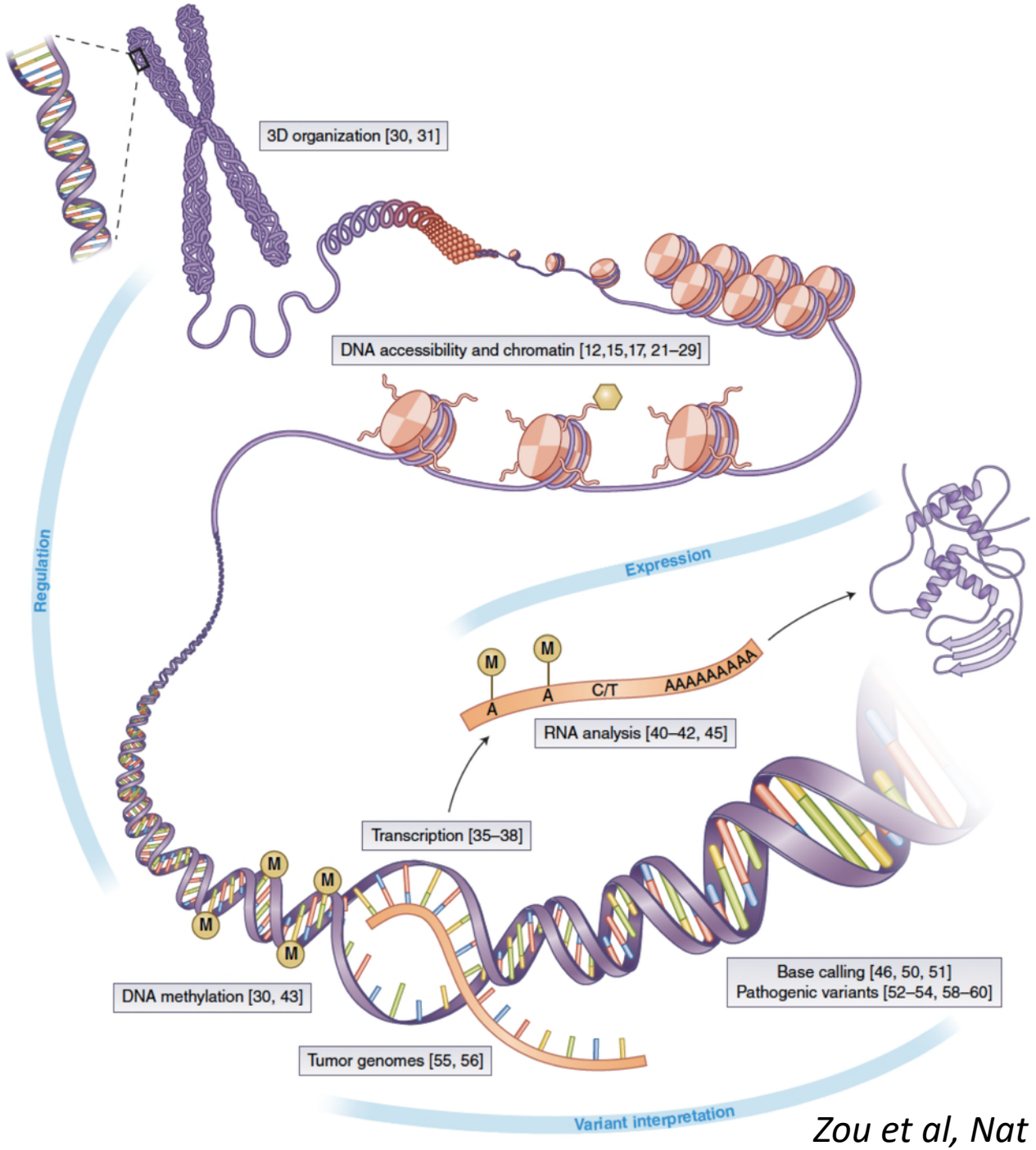
Telenti et al, Hum Mol Genet 2018

A primer on deep learning in genomics



Zou et al, Nat Genet 2019

A primer on deep learning in genomics

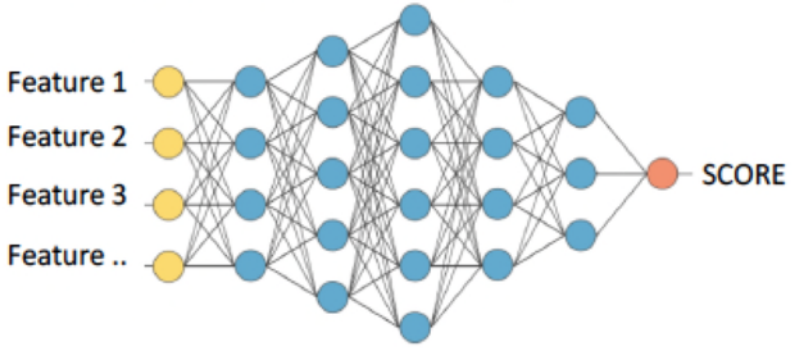


Zou et al, Nat Genet 2019

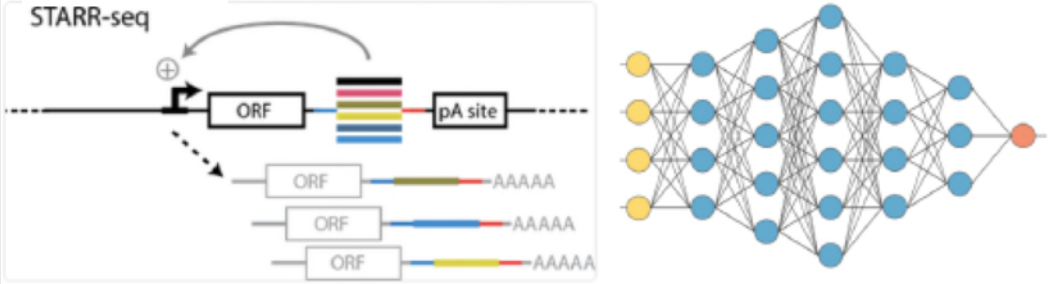
Two User Cases

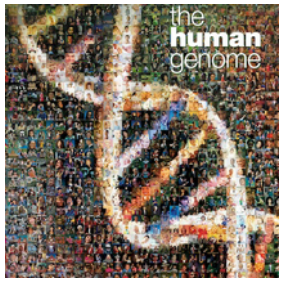
Scripps Research
Translational Institute

Step 1 Model of Essentiality

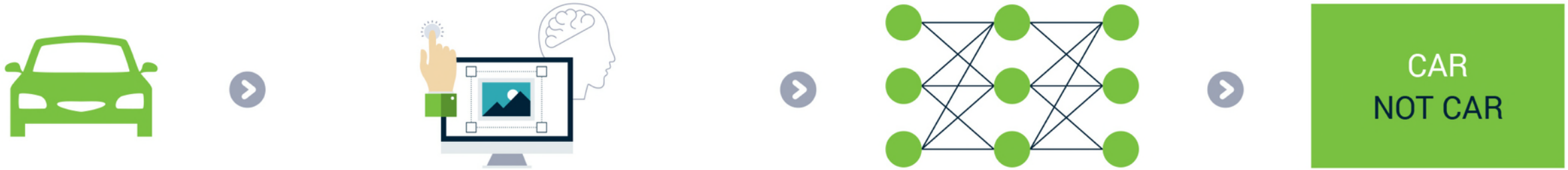


Step 2 Functional screen and retraining



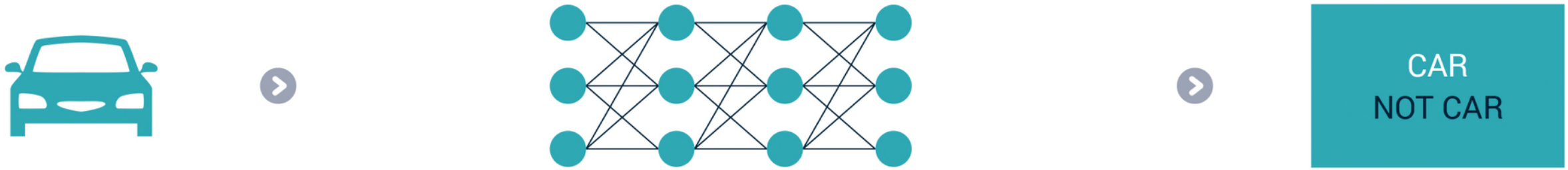


Machine Learning



Input Feature extraction Classification Output

Deep Learning

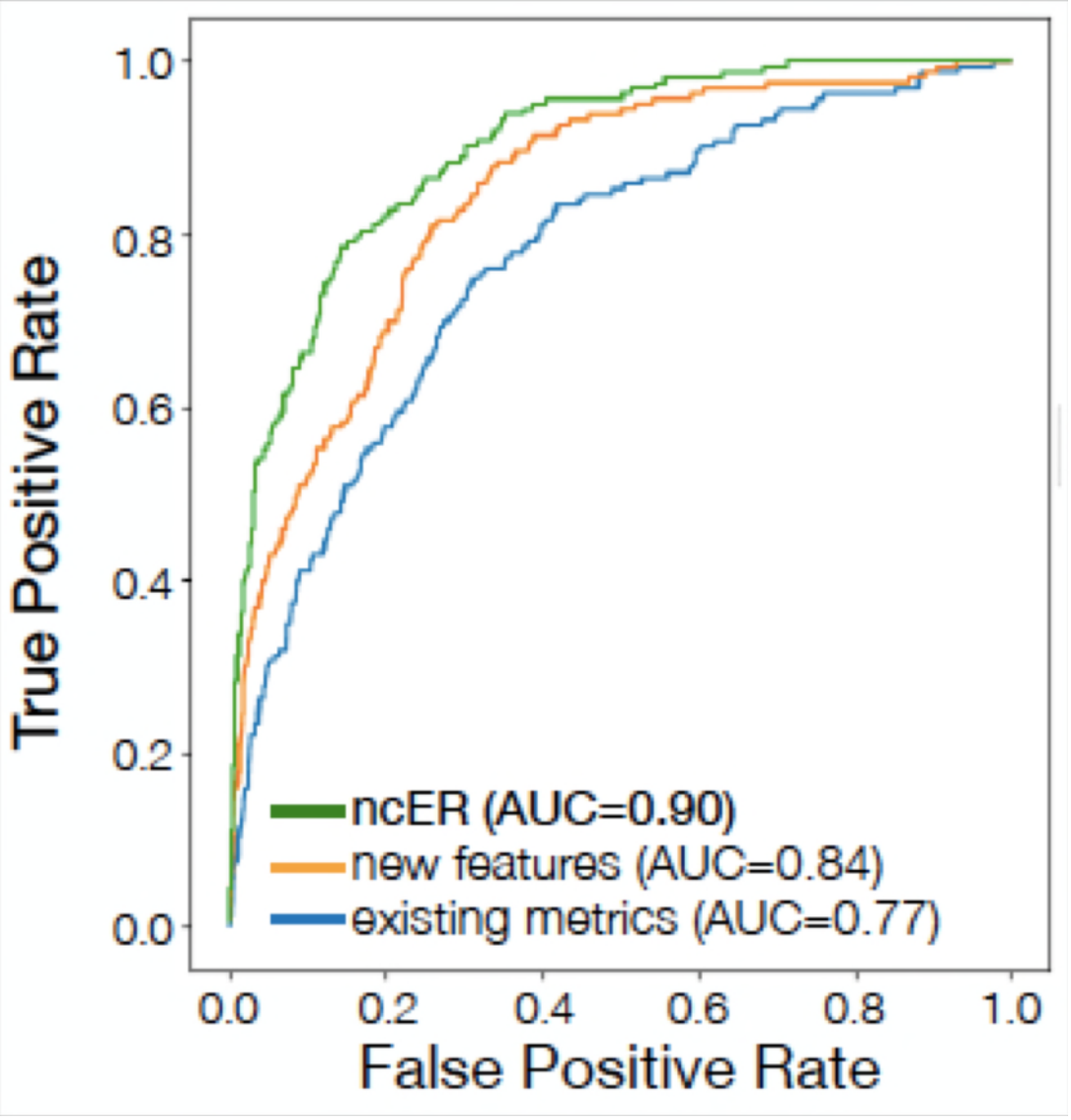


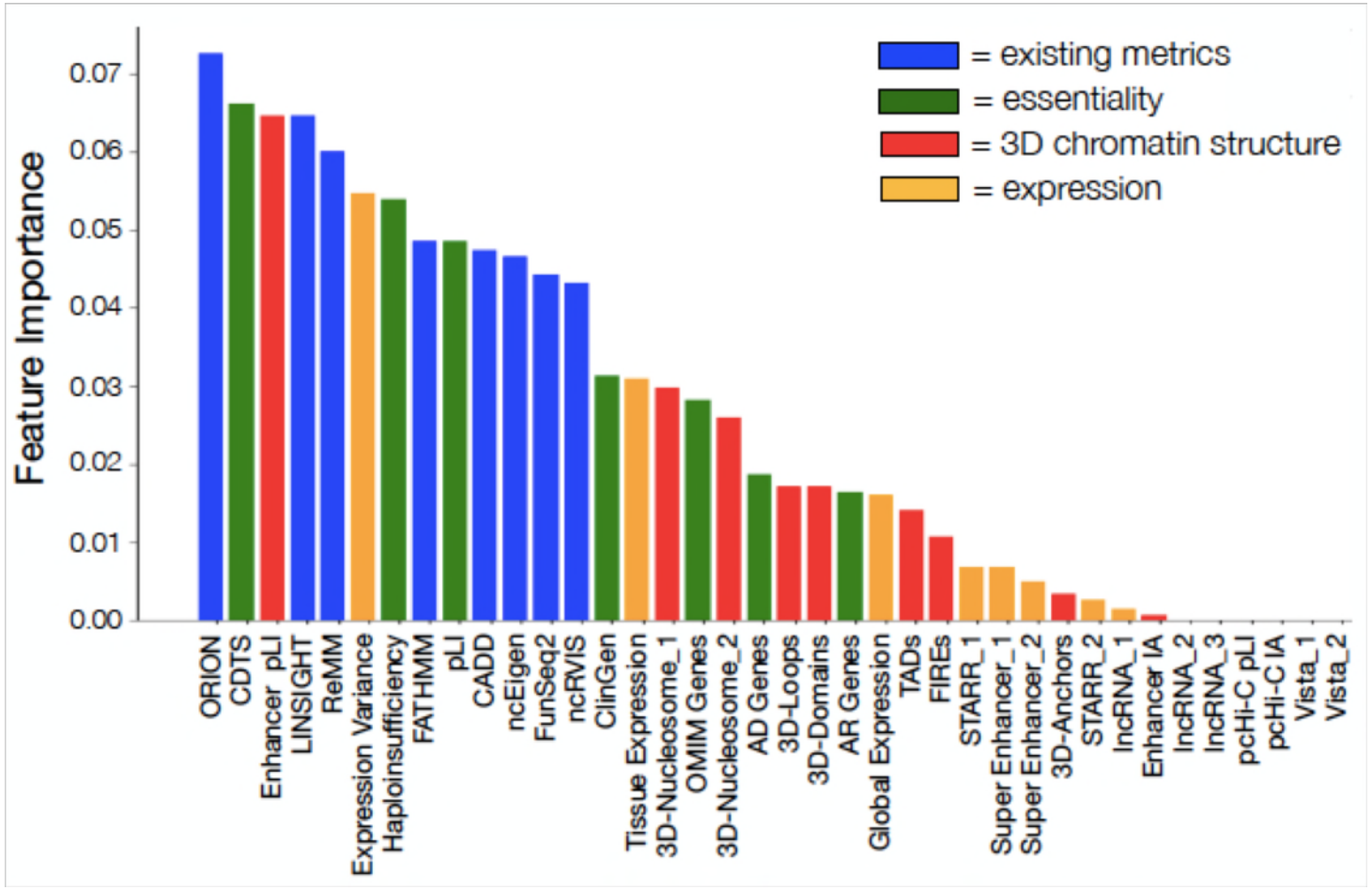
Input Feature extraction + Classification Output

https://verhaert.com/wp-content/uploads/2018/05/Machine-vs-Deep-Learning-InputOutput.jpg

Machine learning model of non-coding genome essentiality

Scripps Research
Translational Institute

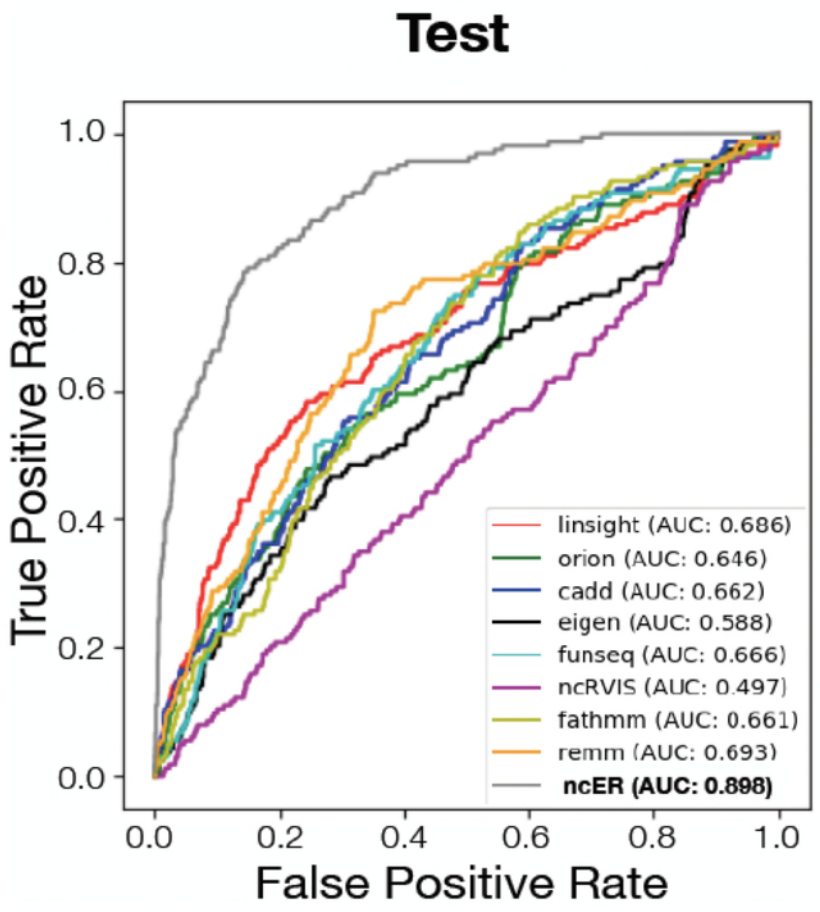




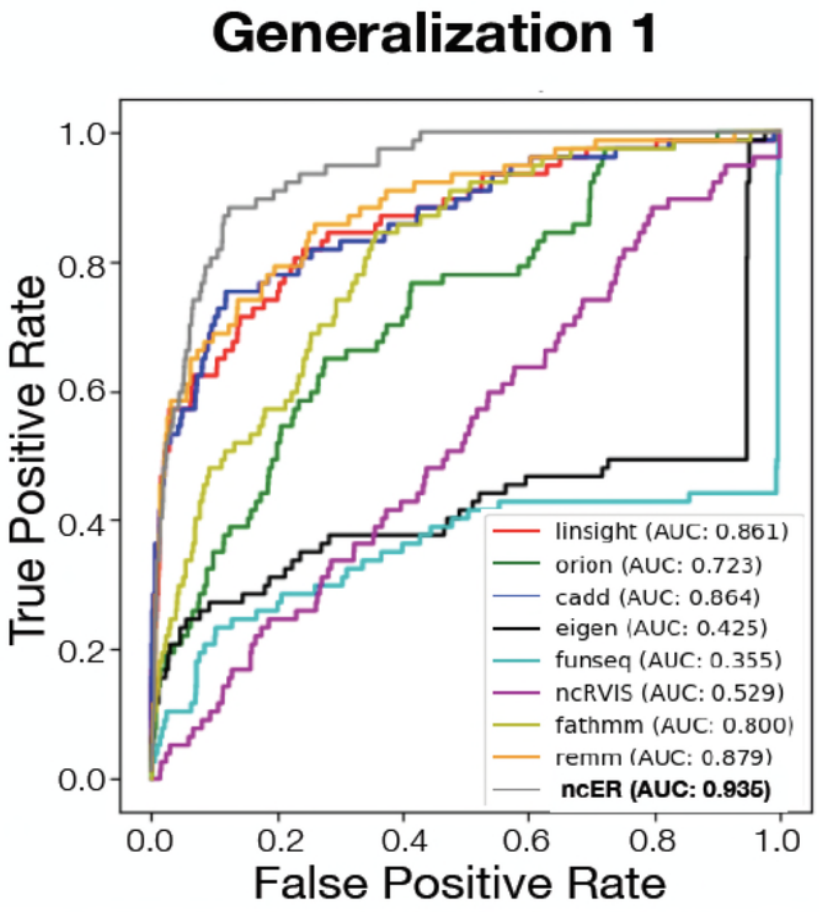
Wells…..di Iulio, BioRxiv 2018

Machine learning model of non-coding genome essentiality

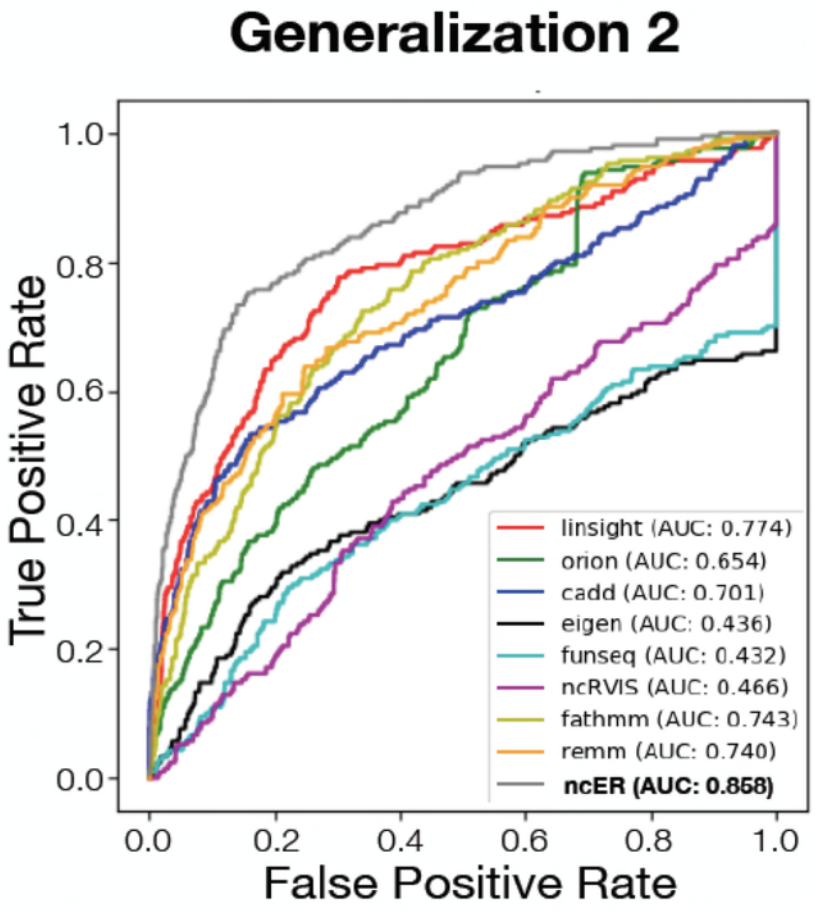
Scripps Research
Translational Institute



20% of data



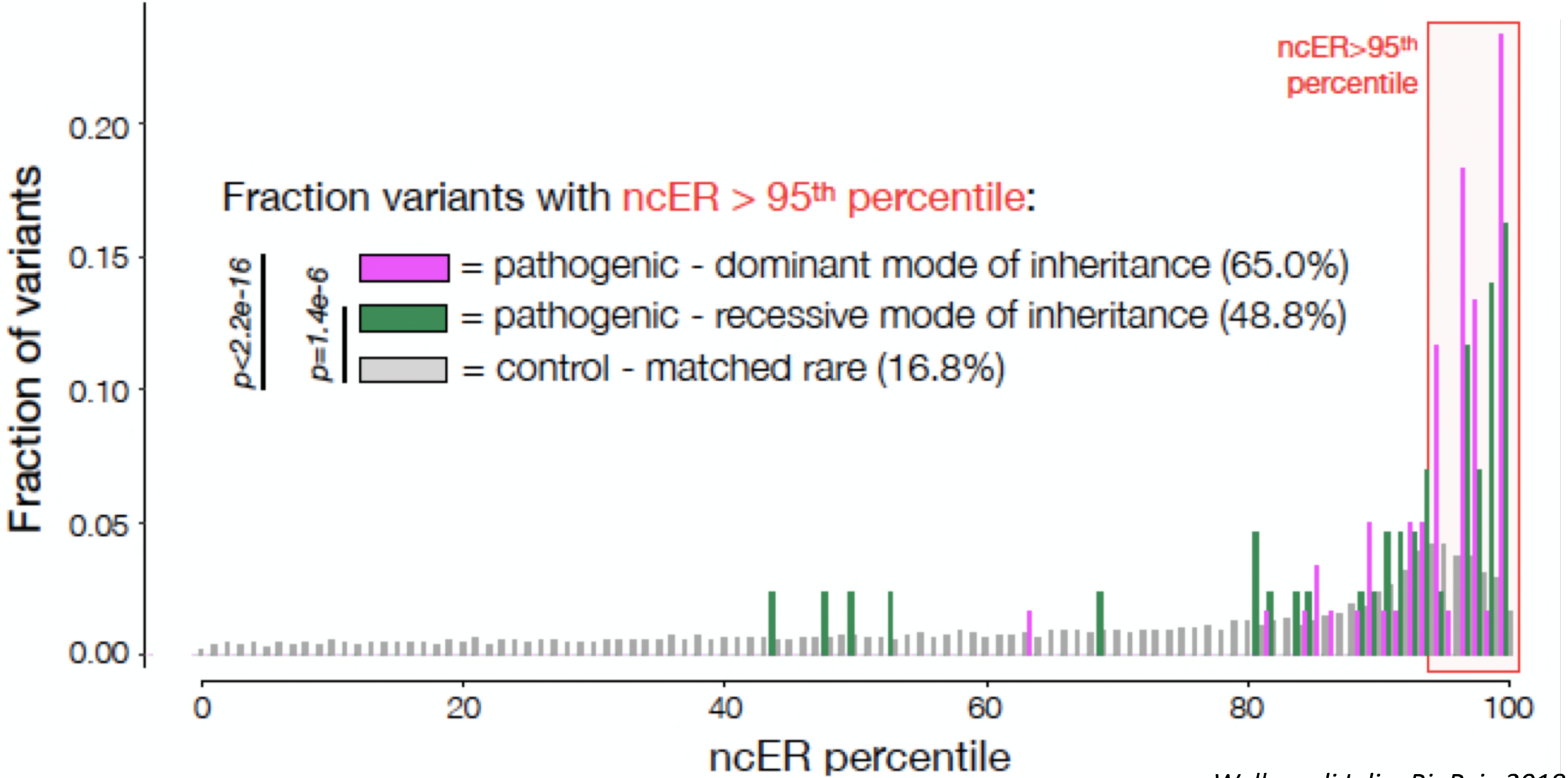
HGMD new set



ncRNA pathogenic variants

Curated Mendelian non-coding variants

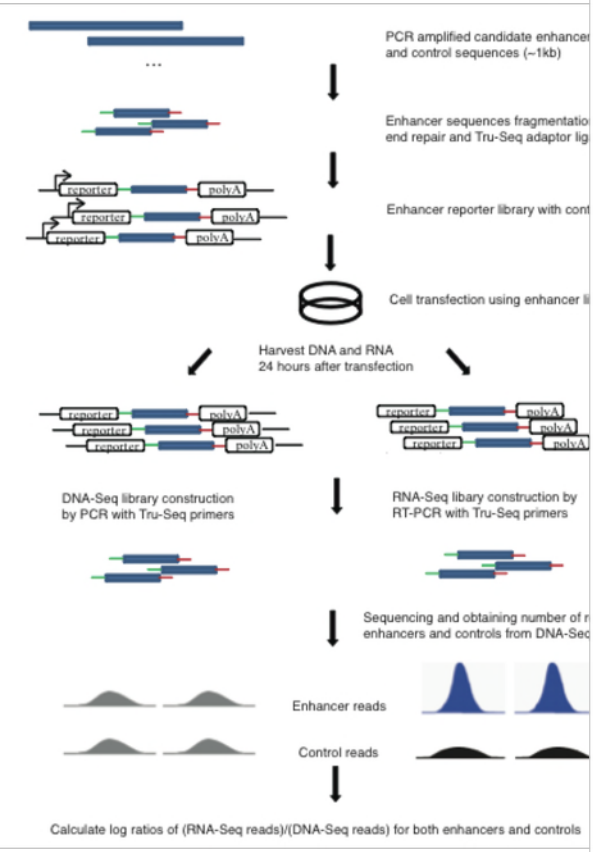
Scripps Research Translational Institute

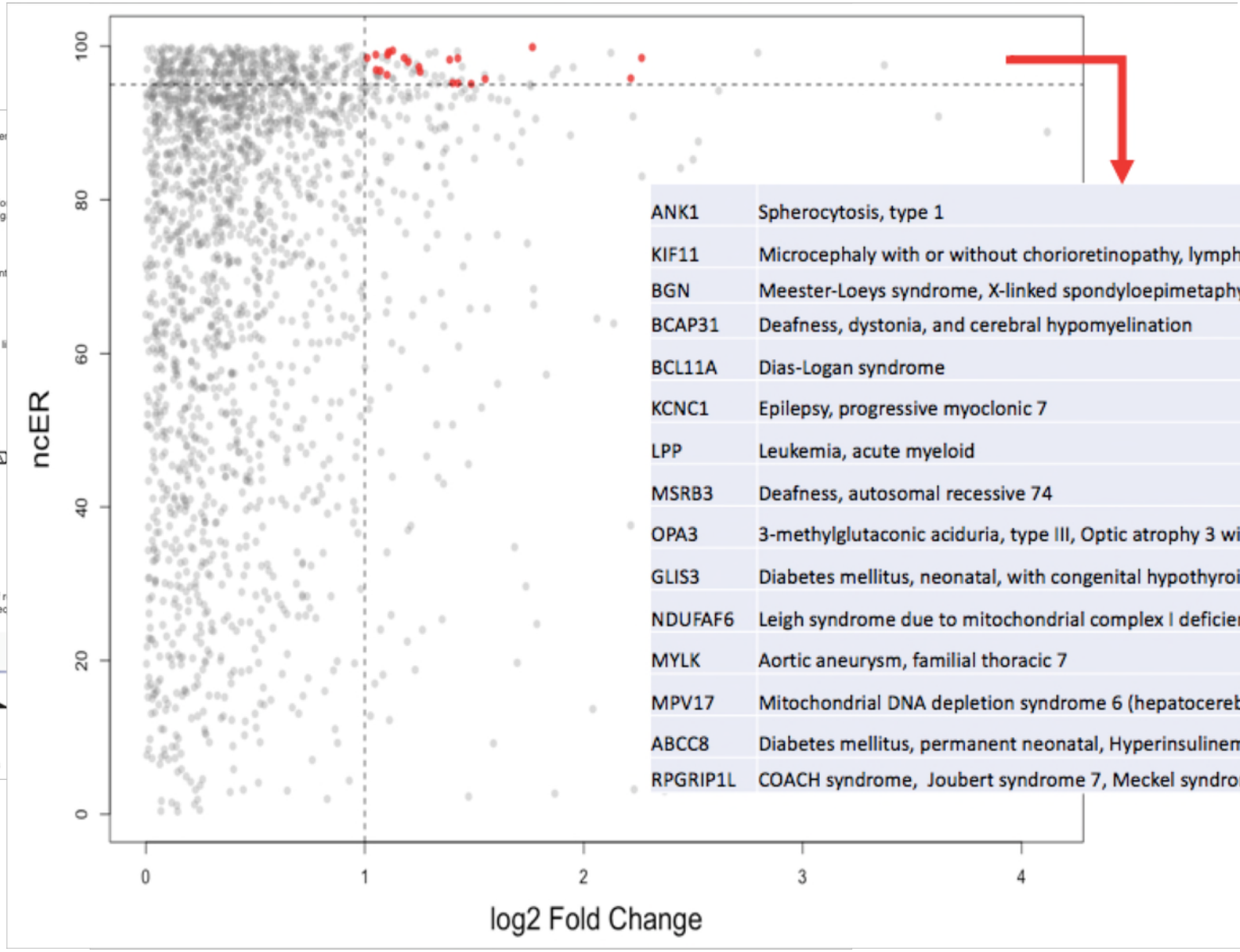


Wells…..di Iulio, BioRxiv 2018

Identification of essential regulatory elements

Scripps Research
Translational Institute



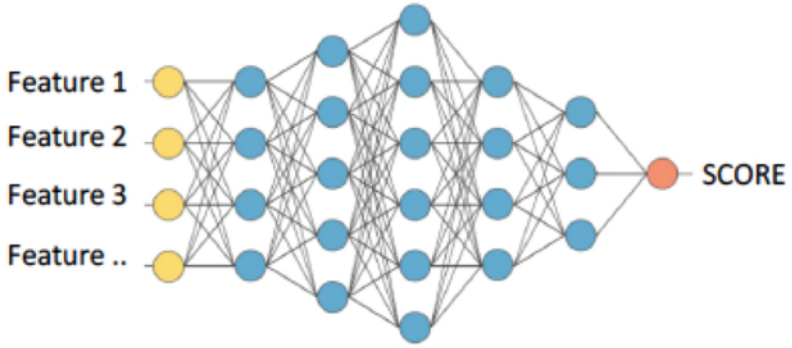


ANK1	Spherocytosis, type 1
KIF11	Microcephaly with or without chorioretinopathy, lymphedema, or mental retardation
BGN	Meester-Loeys syndrome, X-linked spondyloepimetaphyseal dysplasia
BCAP31	Deafness, dystonia, and cerebral hypomyelination
BCL11A	Dias-Logan syndrome
KCNC1	Epilepsy, progressive myoclonic 7
LPP	Leukemia, acute myeloid
MSRB3	Deafness, autosomal recessive 74
OPA3	3-methylglutaconic aciduria, type III, Optic atrophy 3 with cataract
GLIS3	Diabetes mellitus, neonatal, with congenital hypothyroidism
NDUFAF6	Leigh syndrome due to mitochondrial complex I deficiency
MYLK	Aortic aneurysm, familial thoracic 7
MPV17	Mitochondrial DNA depletion syndrome 6 (hepatocerebral type)
ABCC8	Diabetes mellitus, permanent neonatal, Hyperinsulinemic hypoglycemia, familial, 1
RPGRIP1L	COACH syndrome, Joubert syndrome 7, Meckel syndrome 5

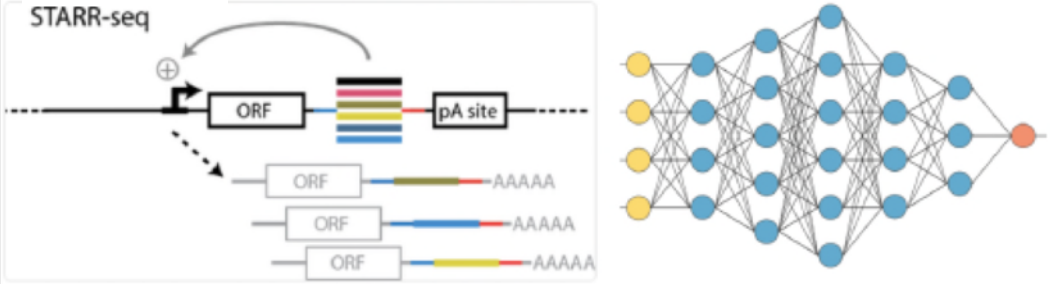
Two User Cases

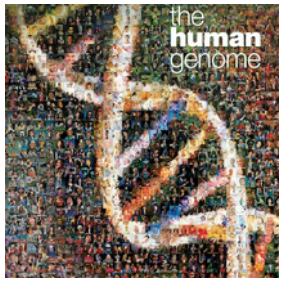
Scripps Research
Translational Institute

Step 1 Model of Essentiality

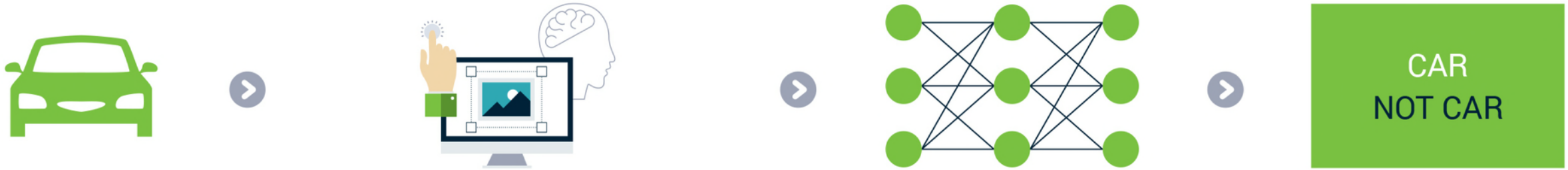


Step 2 Functional screen and retraining



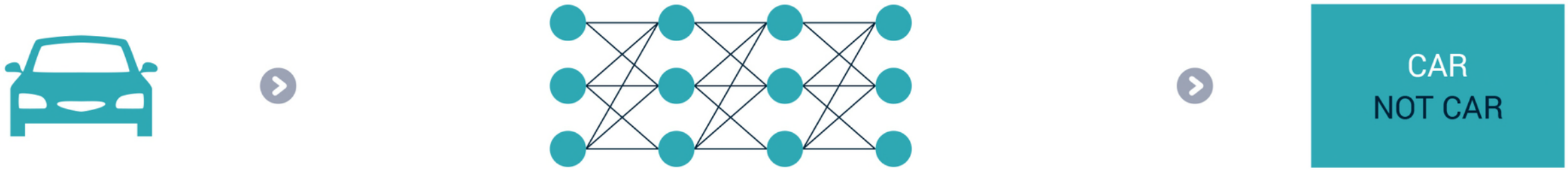


Machine Learning



Input Feature extraction Classification Output

Deep Learning

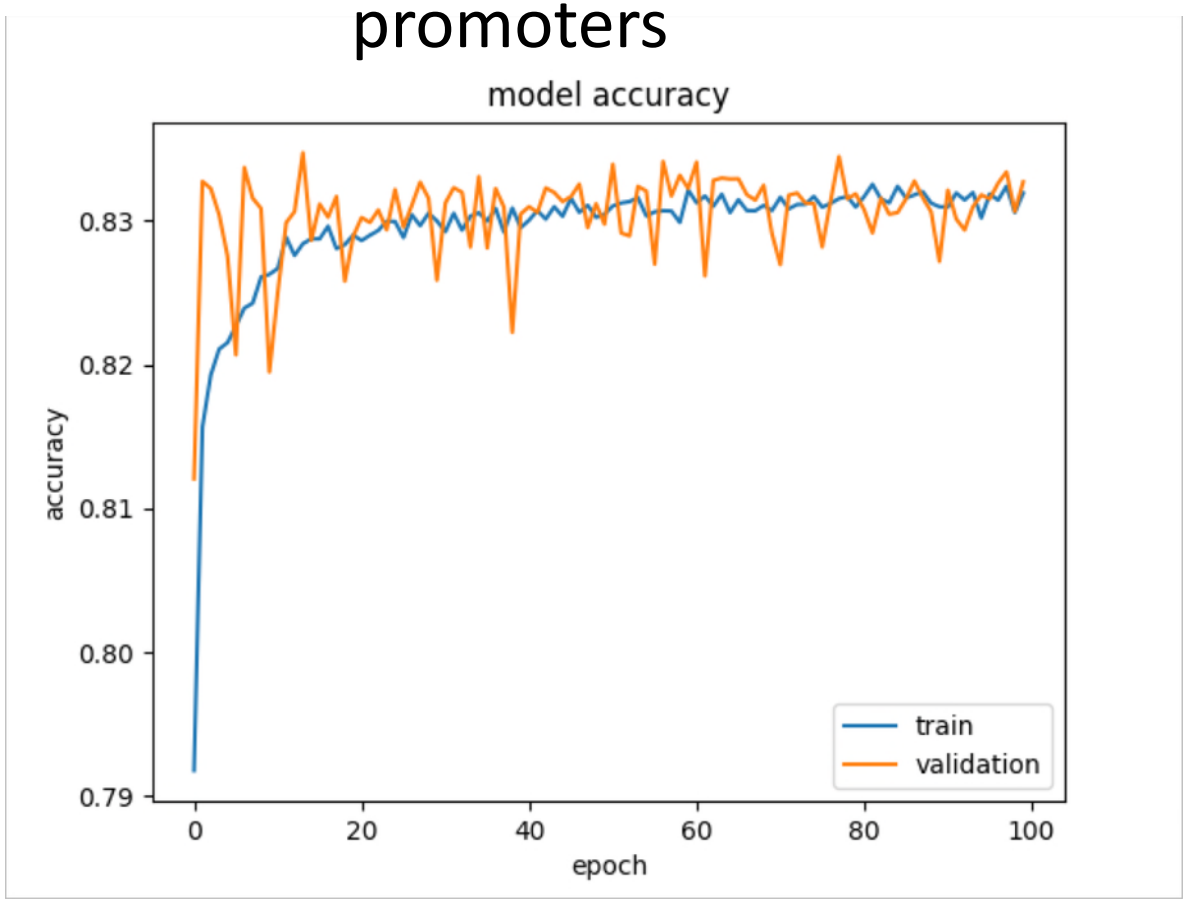


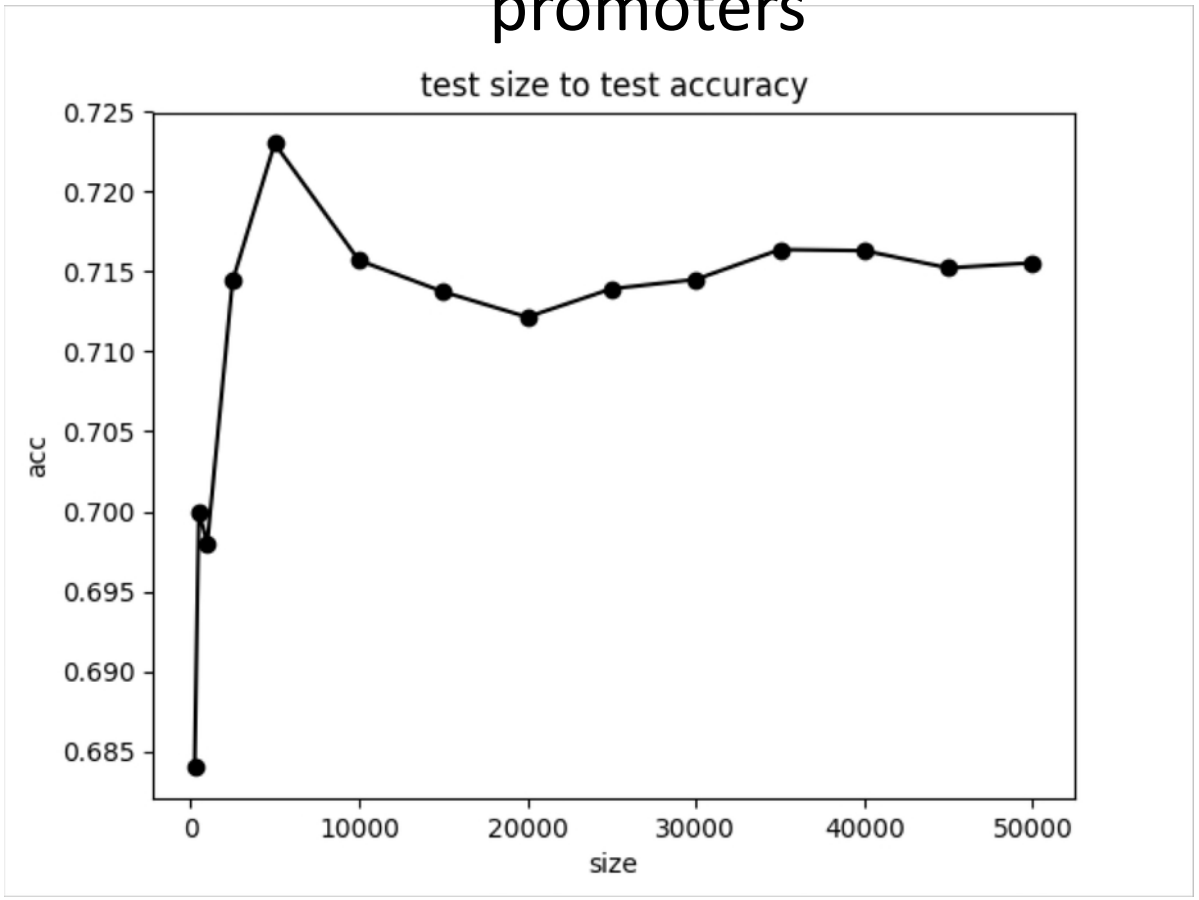
Input Feature extraction + Classification Output

https://verhaert.com/wp-content/uploads/2018/05/Machine-vs-Deep-Learning-InputOutput.jpg

Transfer learning

Training with human promoters → Deep Learning (CNN) → Transfer to mouse promoters





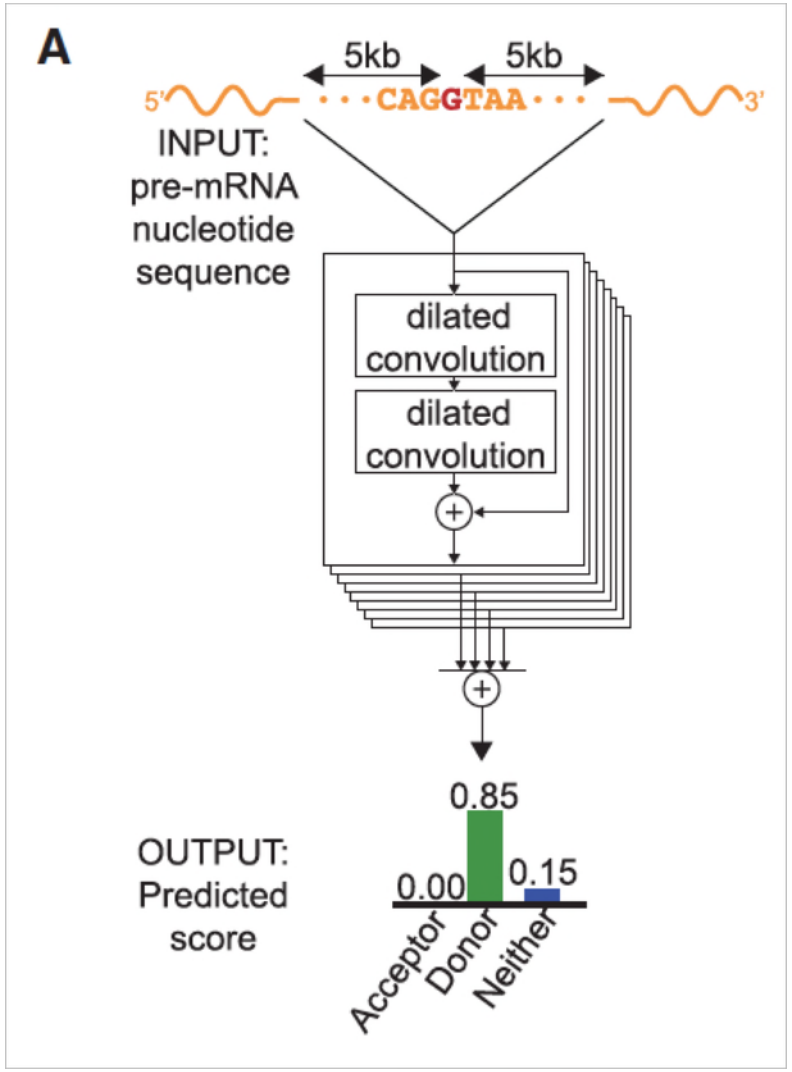
- Transfer learning from human to mouse and other mammals
- Goal is to provide alignment-free universal model for eukaryotic genomes

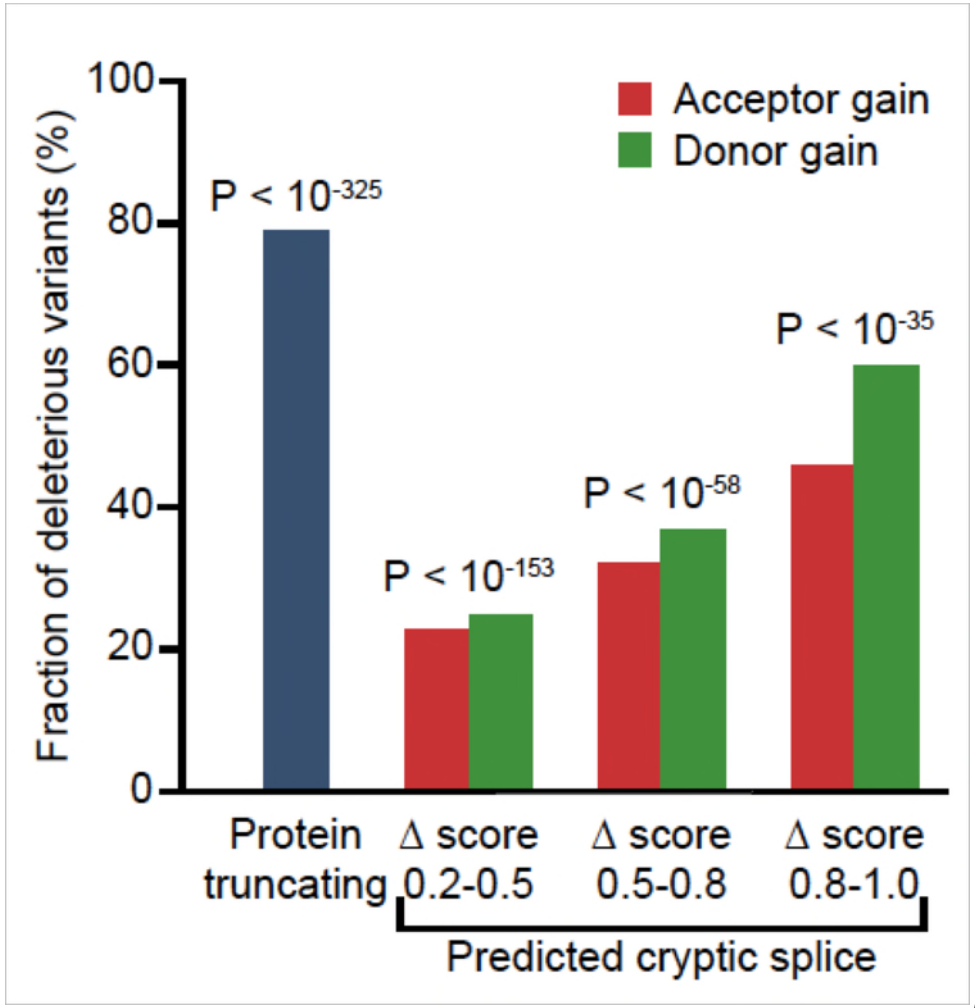
Gu, di Iulio

Notable progress

- New biology
- Better processivity of data

Predicting Splicing from Primary Sequence with Deep Learning



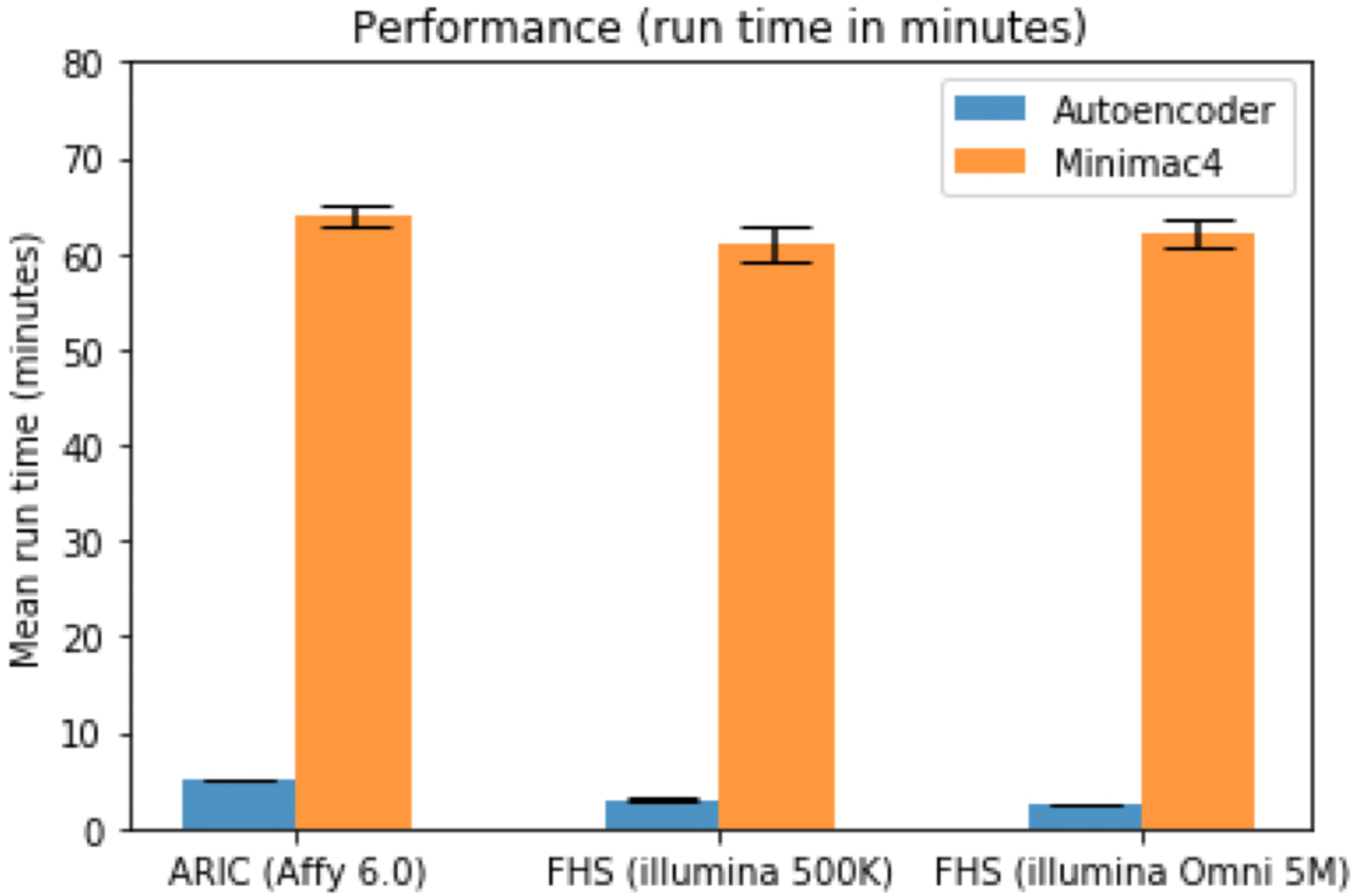


- A deep neural network models mRNA splicing
- Accurately predicts noncoding cryptic splice mutations
- Estimates that ~10% of pathogenic mutations in patients with rare genetic disorders are caused by this mechanism

Jaganathan et al., 2019, Cell 176, 1–14

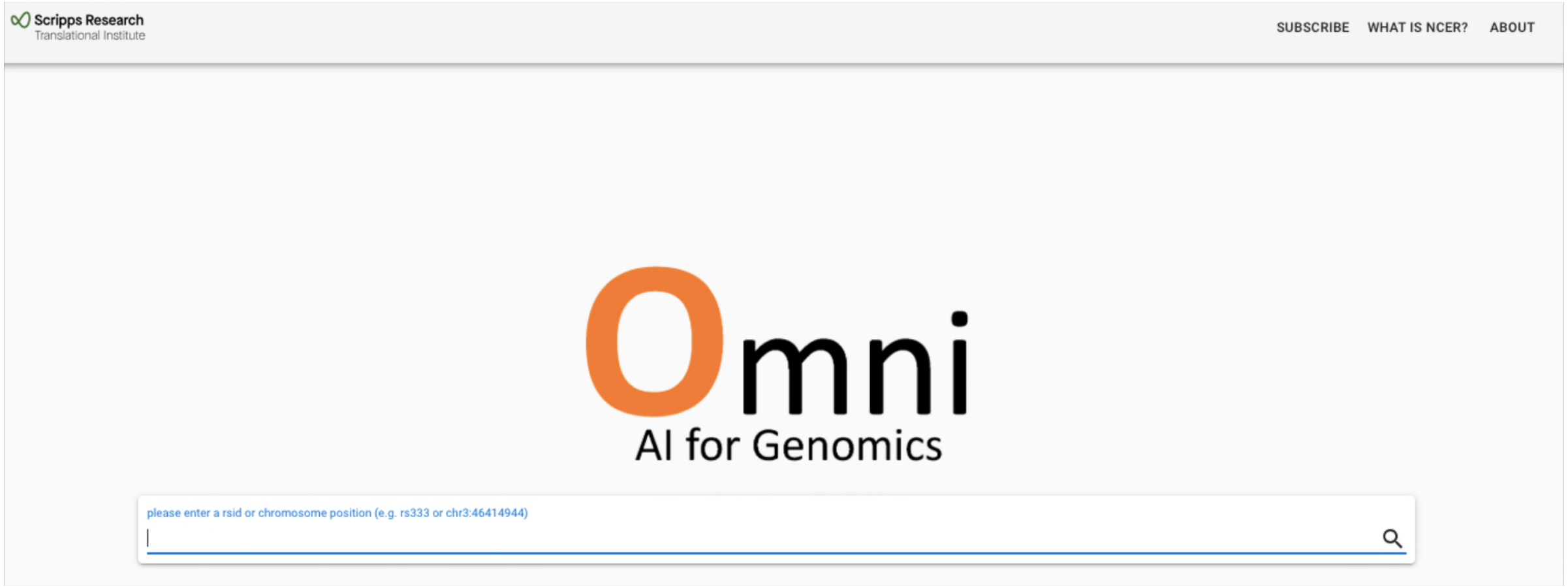
Imputation run time: Deep learning approach versus common statistic methodology

Scripps Research
Translational Institute



Raquel Dias & Ali Torkamani, Scripps

Where to find the new AI genome-wide scores?



omni.telentilab.com

Li Yin

Conclusions and perspectives

Scripps Research Translational Institute

"Application of deep learning to genomic datasets is primed to revolutionize genome analysis" (*Editorial Nat Genet 2019*).

- How to design deep learning systems that support medical decisions (for example, genome interpretation)?
- How to avoid biases in training sets and how to interpret predictions?
- There is a need for iterative experimentation, in which deep learning predictions can be validated by functional laboratory tests or by formal clinical assessment.

Acknowledgements

Julia di Iulio (Scripps)
Li Yin (Scripps)
Alex Wells (Stanford)
Zijing Gu (Scripps)
Shang-Fu Chen (Scripps)
Raquel Dias (Scripps)
Ali Torkamani (Scripps)
Kyle Farh (Illumina)

Scripps Research Translational Institute

NIH National Center for Advancing Translational Sciences