Stanford
MEDICINE

Integrating AI-driven Genomics and Image Analysis for Precision Medicine

HoJoon Lee

K-PAI April 22, 2025

Table of Contents

Stanford MEDICINE

Beginning of deep learning

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

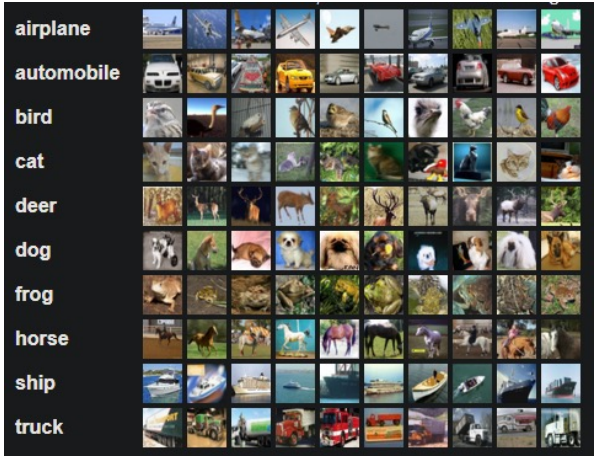
Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

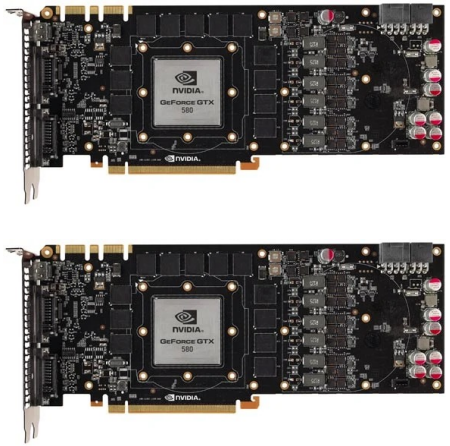
Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

AlexNet, 2012



ImageNet, Fei-Fei Li at Stanford



Nvidia GTX 580 GPUs

Stanford MEDICINE

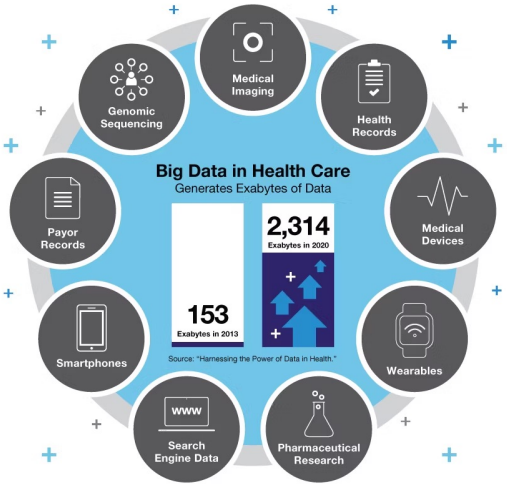
AI in biomedical field

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Transformer, 2017

Big biomedical data*

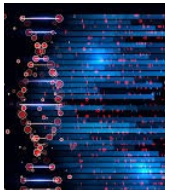


Nvidia H100

*https://www.micron.com/about/blog/applications/data-center/big-data-can-revolutionize-health-care

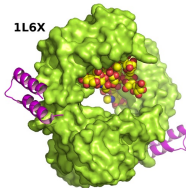
Stanford MEDICINE

Big biomedical data



> YouTube

*Genomic Data

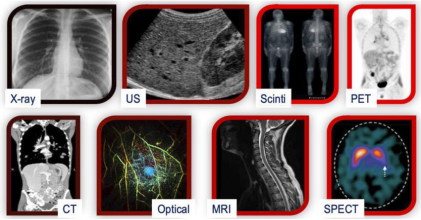


→ Google DeepMind INTRODUCING AlphaFold-3

Proteomic Data

→ LLM

Electronic health records (EHRs)



Medical Images

Wearable Device

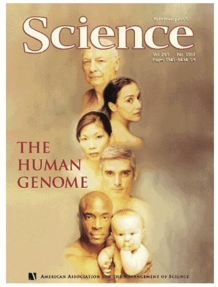
MEDICAL BILL

Insurance records

*3billion.io/blog/big-data-among-big-data-genome-data

Stanford MEDICINE

Big genomics data



Human genome project

Multiple genomics/proteomics data of >11,000 samples from 33 cancer types

100,000 genomes club

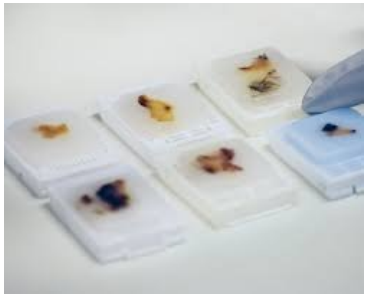
Genomics England

FINNGEN

All of Us
THE FUTURE OF HEALTH BEGINS WITH YOU
The Precision Medicine Initiative

Stanford MEDICINE

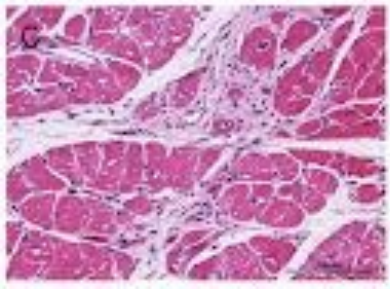
Histopathological images



Formalin-Fixed Paraffin-Embedded (FFPE) tissue blocks

Blocks for every patients

Staining



Hematoxylin and Eosin (H&E) images

Digitized whole slide images (WSI)

Early phase

Stanford MEDICINE

What can we do with big data?

Diagnosis

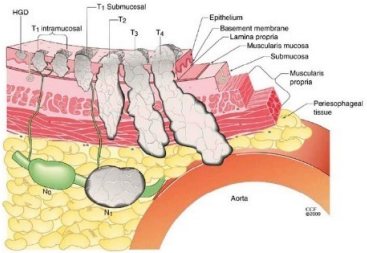
Prognosis

Therapeutics

Stanford MEDICINE

Traditional approach

Diagnosis

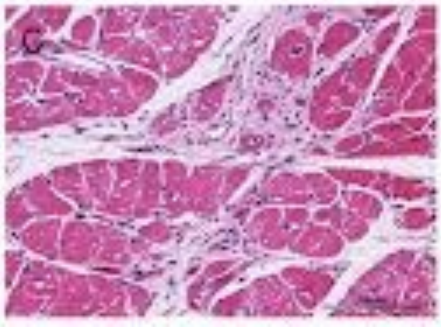


Prognosis

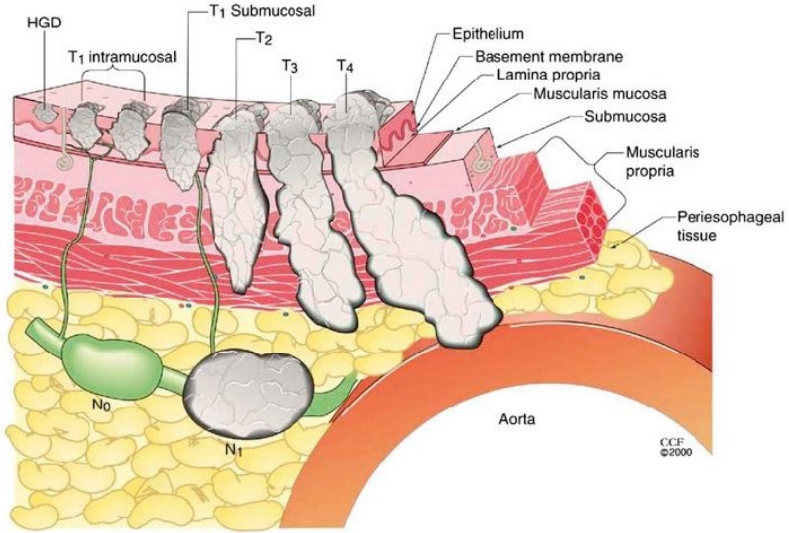
Therapeutics

Stanford MEDICINE

Cancer diagnosis by H&E images



Hematoxylin and Eosin (H&E) images



TNM staging:
Stage 1, 2, 3 and 4

It has been instrumental for medical diagnosis

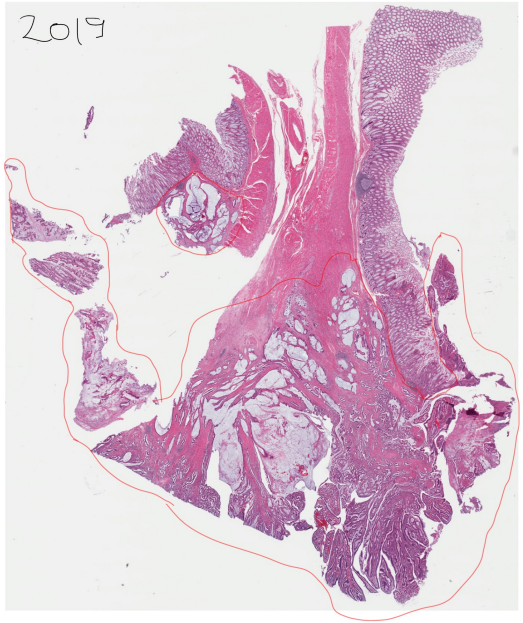
Stanford MEDICINE

Prognosis by clinical stage

	Stage at Diagnosis	5-Year Relative Survival (%)
	All Stages	9.2
Stage 1 & 2	Localized	33.8
Stage 3	Regional	19.8
Stage 4	Distant	4.2
	Unstaged	11.1

Stanford MEDICINE

Current limits on pathology



Annotation by pathologists

- Labor intense task: not scalable
- Subjectivity and Variability: not robust
- No annotations for other cell types such as immune cells

Clinical staging:

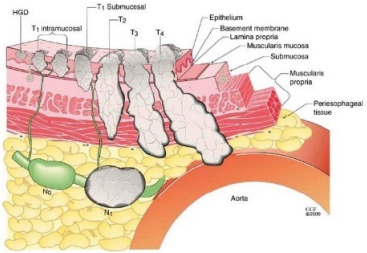
- Discrepancies between clinical stage and true extent of a disease

Build deep learning models to identify cell types on H&E images

Stanford MEDICINE

Precision Medicine by genomic info

Diagnosis



Prognosis

\+

Genomic/Proteomic Profiling

Stanford MEDICINE

What can we do with genomic data?

Diagnosis

- Cancer Susceptibility: BRCA1 and BRCA2
- Liquid biopsies: circulating tumor DNA in blood

Prognosis

Therapeutics

Stanford MEDICINE

What can we do with genomic data?

Diagnosis

- Cancer Susceptibility: BRCA1 and BRCA2
- Liquid biopsies: circulating tumor DNA in blood

Prognosis

- BRAF V600E mutations in melanoma tumors
- High tumor mutation burden (TMB)

Therapeutics

Stanford MEDICINE

What can we do with genomic data?

Diagnosis

- Cancer Susceptibility: BRCA1 and BRCA2
- Liquid biopsies: circulating tumor DNA in blood

Prognosis

- BRAF V600E mutations in melanoma tumors
- High tumor mutation burden (TMB)

Therapeutics

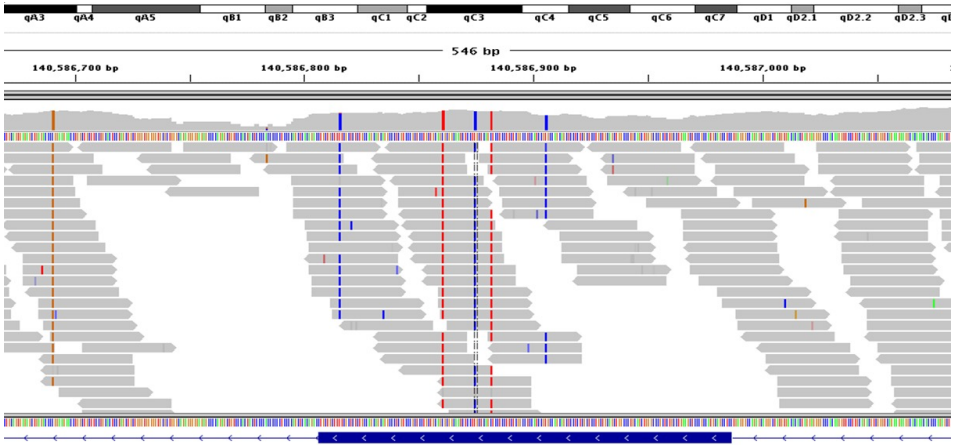
- Vemurafenib: BRAF V600E mutations in melanoma tumors
- Herceptin: HER2 mutation

Stanford MEDICINE

AI in genomics analysis

The primary goal of genomic analysis:

- Identify genomic alterations in cancer tissues from sequencing data



Google DeepVariant

Stanford MEDICINE

Table of Contents

Stanford MEDICINE

How to treat colon advanced cancers?

Caner patients with advanced stages

- Surgery
- Chemotherapy
- Radiotherapy
- Targeted therapy
- Immunotherapy

Stanford MEDICINE

How to treat colon advanced cancers?

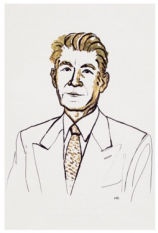
Caner patients with advanced stages

- ~~Surgery~~
- ~~Chemotherapy~~
- ~~Radiotherapy~~
- ~~Targeted therapy~~
- Immunotherapy

The Nobel Prize in Physiology or Medicine 2018

Ill. Niklas Elmehed. © Nobel Media

James P. Allison



Ill. Niklas Elmehed. © Nobel Media

Tasuku Honjo

The response rate to immune checkpoint inhibitor: 20%

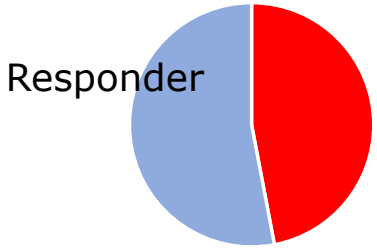
Who will response??

Stanford MEDICINE

Markers for responders

- Microsatellite Instable (MSI) or high tumor mutation burden (TMB)

2020



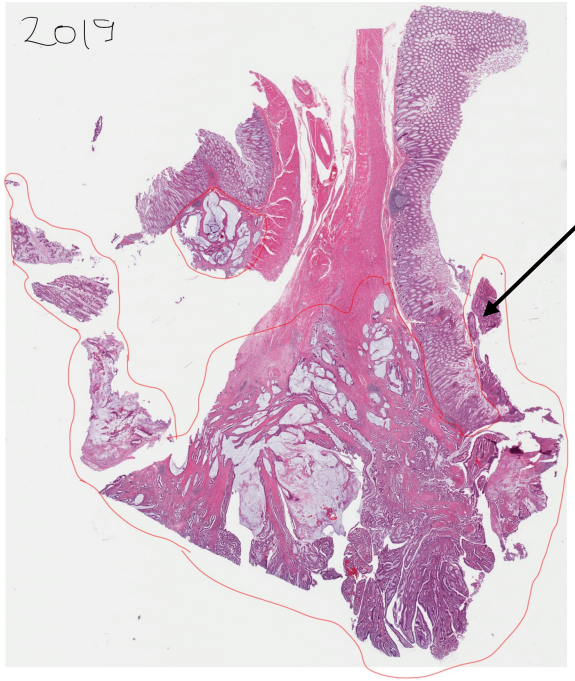
MSI-H cancers

Better marker or indicator?

- Tumor infiltrating lymphocytes (TILs)

Stanford MEDICINE

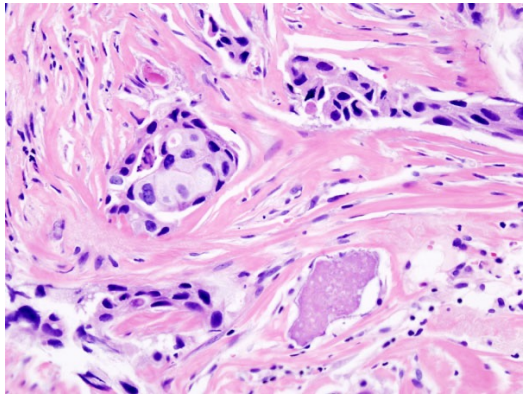
How to assess TILs?



- Locate the immune cells in tumor regions
- Urgent need of AI applications

Stanford MEDICINE

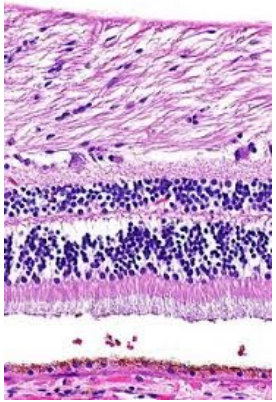
Challenges in AI application



256 x 256
ImageNet

Gigapixel (50, 000 × 80,000) for a tissue biopsy (1.5 cm × 2.5cm)

Hard to distinguish between different cells



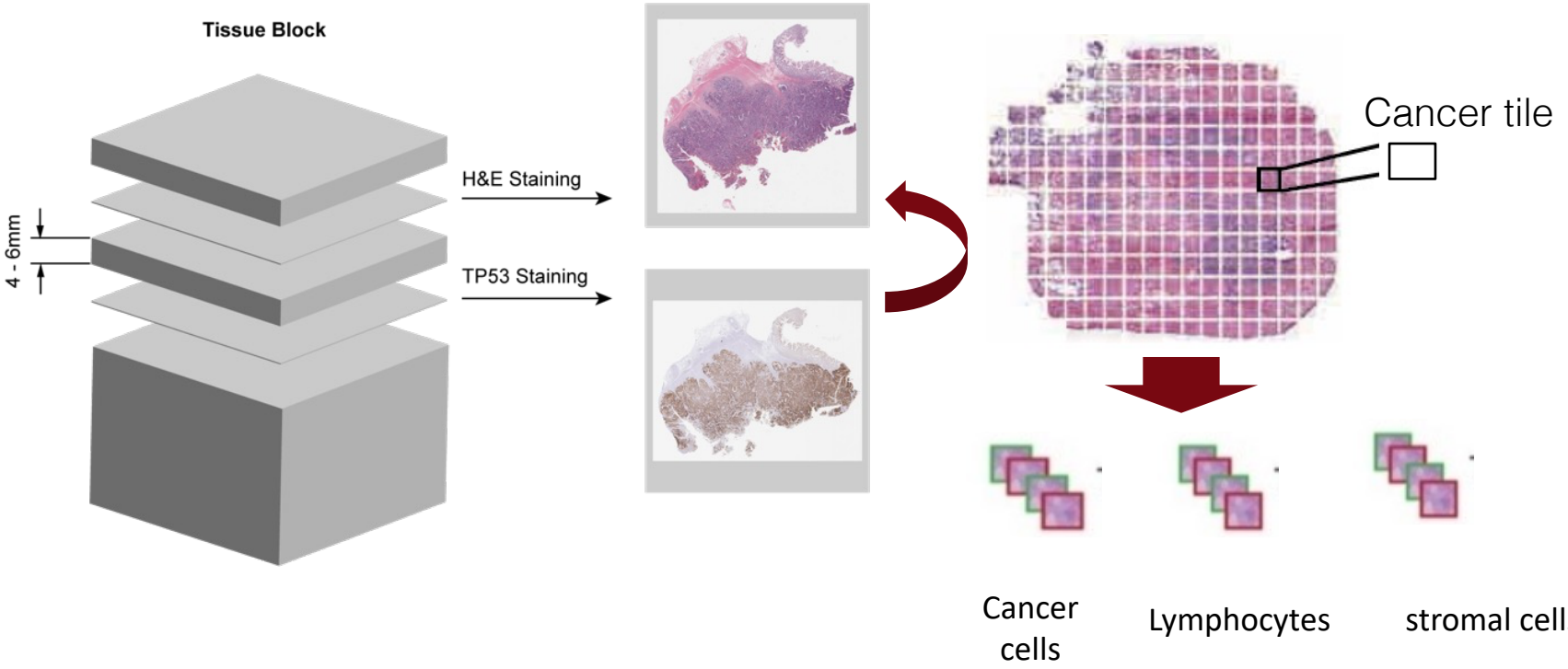
easy to distinguish

Lunit

Modella AI

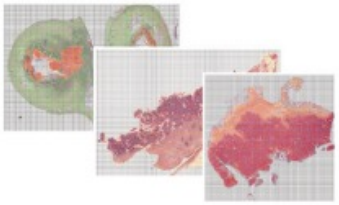
Stanford MEDICINE

Our solution: molecular staining



Stanford MEDICINE

Training images for deep learning model



Non-Cancer Tiles (n=8782)

Uncertain Tiles (n=23275)

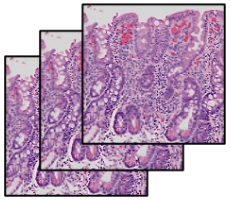
Cancer Tiles (n=21939)

Training Convolutional Neural Network (CNN)

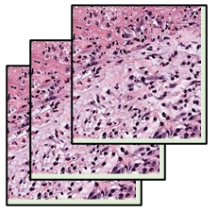
Molecular labeling is scalable and robust

~20-50 histology images: >10K images of each cell type

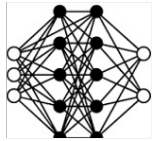
Traditional approach



>10K gastric cancer images



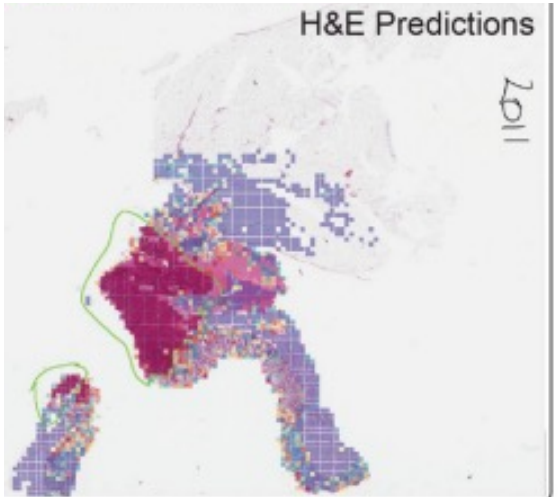
>10K normal tissue images

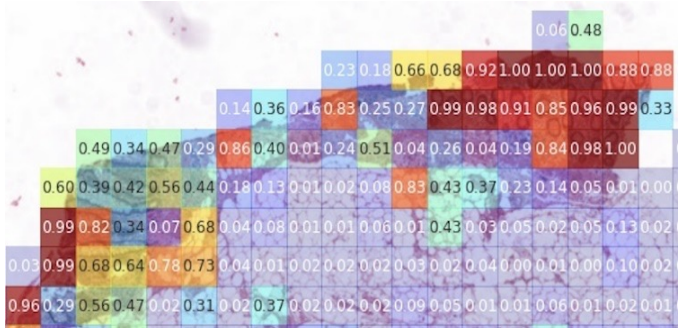


The number of histology images per each category >10K

Stanford MEDICINE

Deep learning model for cancer cells





ROC AUC
- 0.84

npj | Precision Oncology
www.nature.com/npjprecisiononcology

ARTICLE OPEN

Check for updates

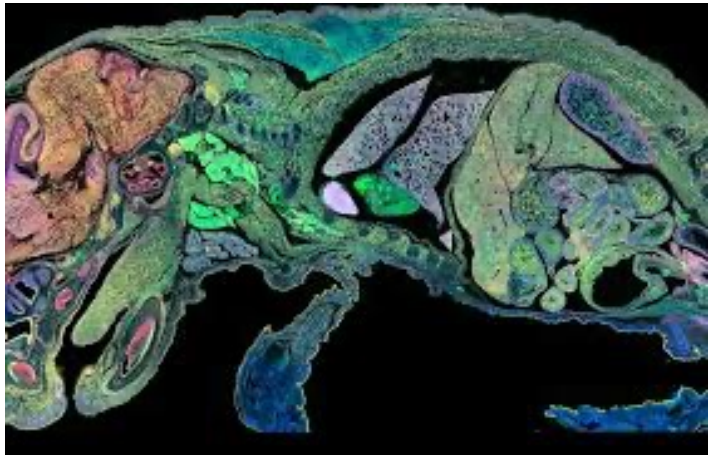
A deep learning model for molecular label transfer that enables cancer cell identification from histopathology images

Andrew Su[1,6], HoJoon Lee[2,6], Xiao Tan[1,6], Carlos J. Suarez[3], Noemi Andor[2,5], Quan Nguyen[1✉] and Hanlee P. Ji[2,4✉]

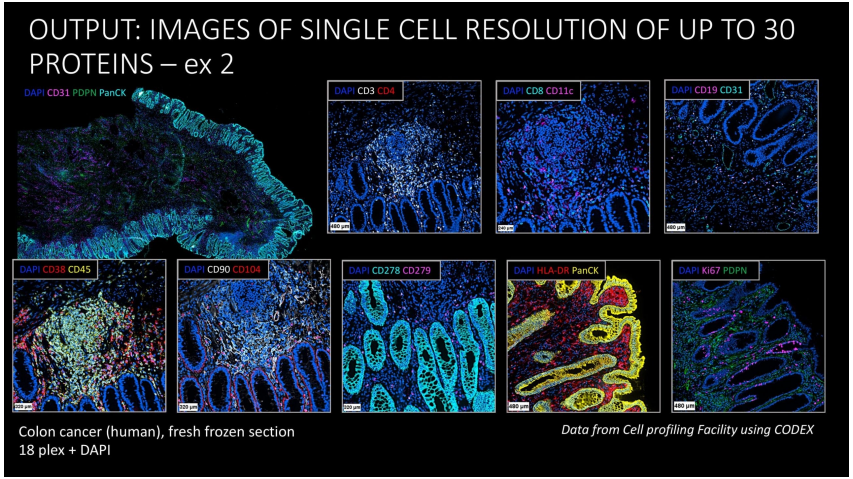
Deep-learning classification systems have the potential to improve cancer diagnosis. However, development of these computational approaches so far depends on prior pathological annotations and large training datasets. The manual annotation is low-resolution, time-consuming, highly variable and subject to observer variance. To address this issue, we developed a method, H&E Molecular neural network (HEMnet). HEMnet utilizes immunohistochemistry as an initial molecular label for cancer cells on a H&E image and trains a cancer classifier on the overlapping clinical histopathological images. Using this molecular transfer method, HEMnet successfully generated and labeled 21,939 tumor and 8782 normal tiles from ten whole-slide images for model training. After building the model, HEMnet accurately identified colorectal cancer regions, which achieved 0.84 and 0.73 of ROC AUC values compared to p53 staining and pathological annotations, respectively. Our validation study using histopathology images from TCGA samples accurately estimated tumor purity, which showed a significant correlation (regression coefficient of 0.8) with the estimation based on genomic sequencing data. Thus, HEMnet contributes to addressing two main challenges in cancer deep-learning analysis, namely the need to have a large number of images for training and the dependence on manual labeling by a pathologist. HEMnet also predicts cancer cells at a much higher resolution compared to manual histopathologic evaluation. Overall, our method provides a path towards a fully automated delineation of any type of tumor so long as there is a cancer-oriented molecular stain available for subsequent learning. Software, tutorials and interactive tools are available at:https://github.com/BiomedicalMachineLearning/HEMnet

npj Precision Oncology (2022)6:14 ; https://doi.org/10.1038/s41698-022-00252-0

Technologies for other cell types



Xenium 10X genomics



CODEX, multiplexed single-cell imaging technology

Annotating H&E images with many different cell types including immune cells

Stanford MEDICINE

Clinical decision with pathological images

Immune cells in tumor sites

Yes

Immune-check point inhibitor

Reactivating immune response

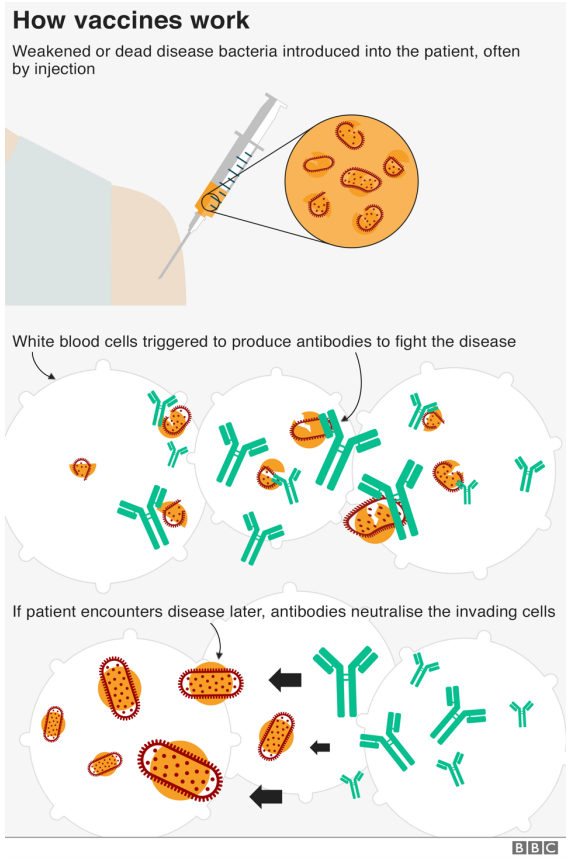
No

- **Cancer vaccines**
- T-cell therapy

Enhancing immune response by education

Stanford MEDICINE

How vaccines work?

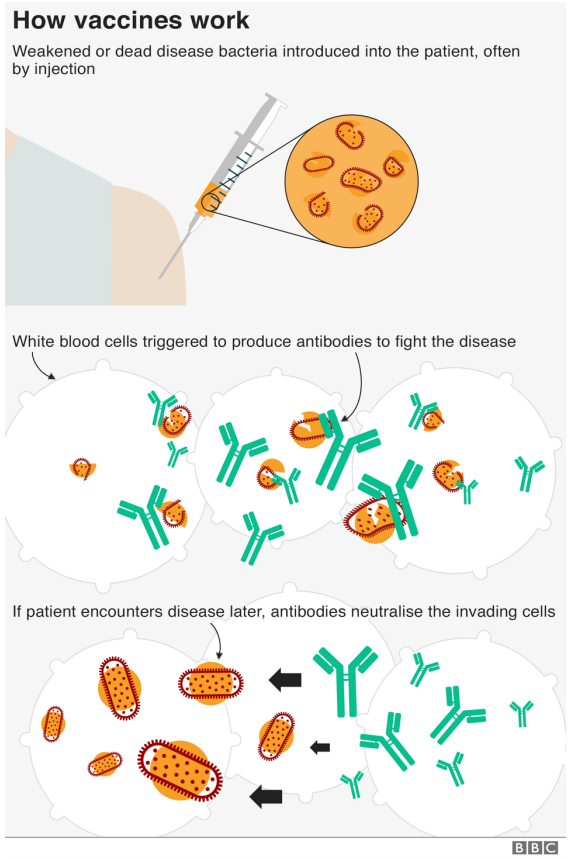


Need targets!

https://www.bbc.co.uk/news/world-48186856

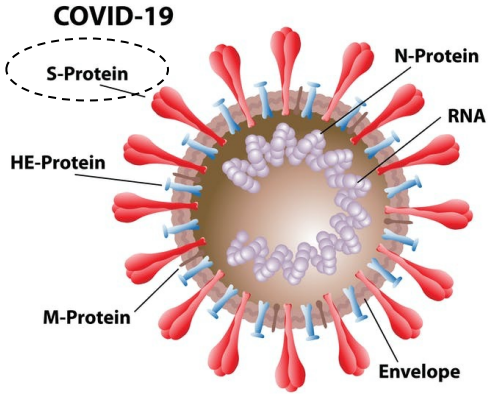
Stanford MEDICINE

How vaccines work?



https://www.bbc.co.uk/news/world-48186856

Need targets!

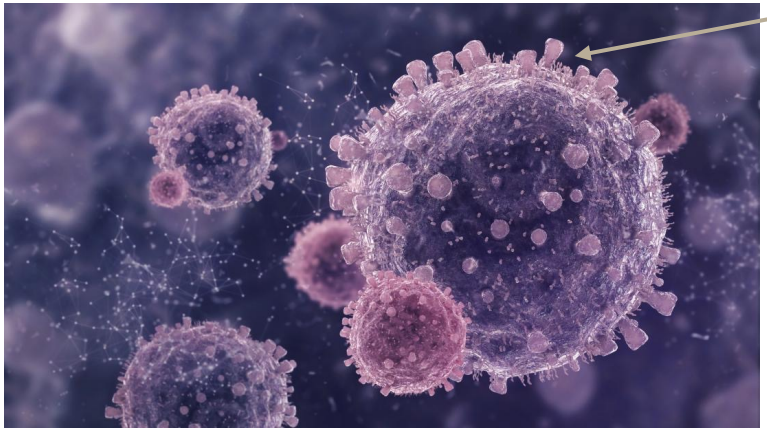


https://theconversation.com/covid-vaccines-focus-on-the-spike-protein-but-heres-another-target-150315

Stanford MEDICINE

Targets for cancer cells?

Cancer cells



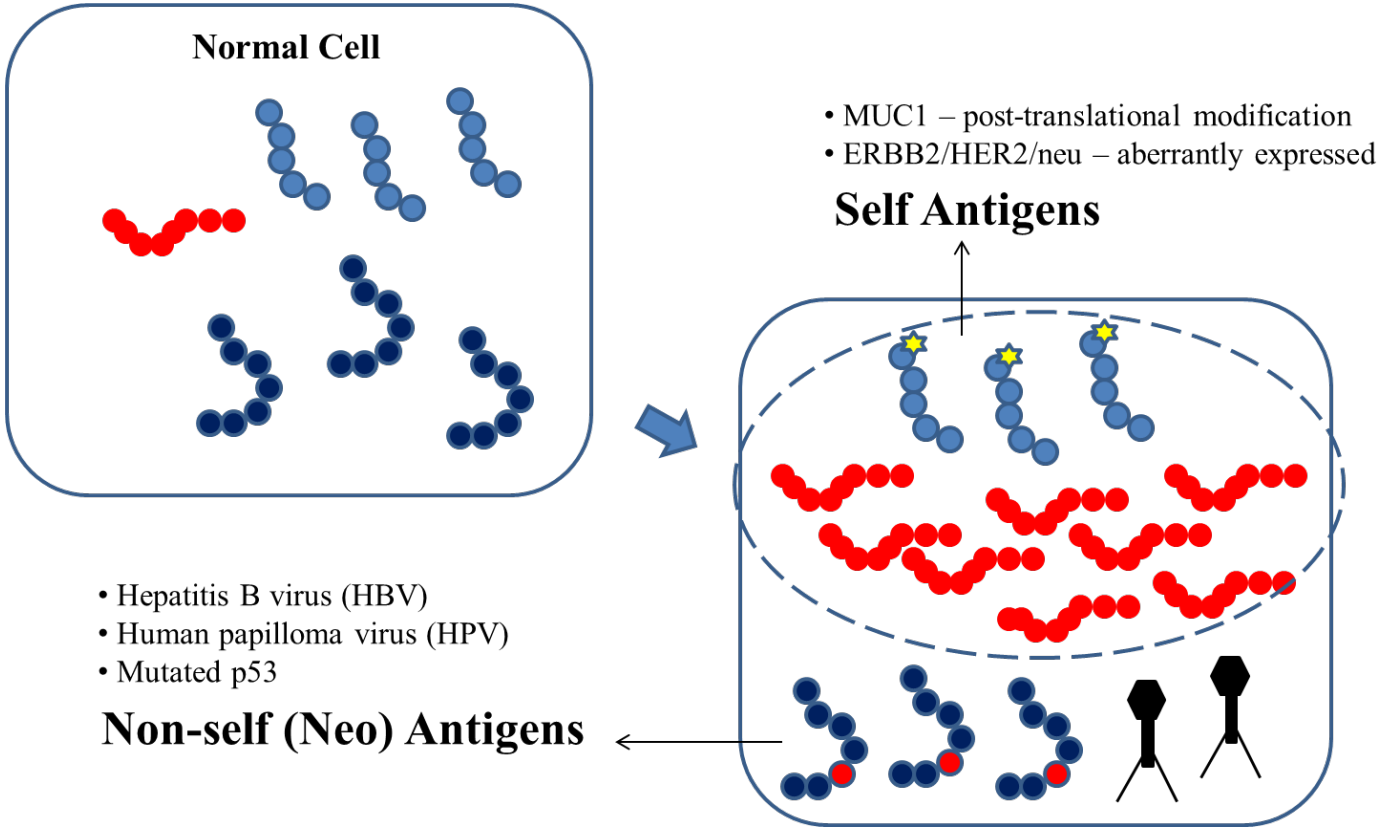
https://www.medicalnewstoday.com/articles/244845

Non-self molecules

(Neoantigens)

Stanford MEDICINE

What is neo-antigen?

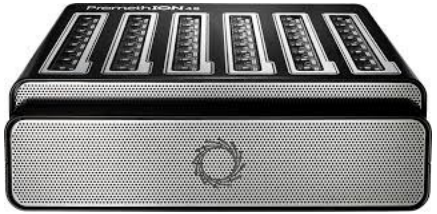


Stanford MEDICINE

Source of (endogenous) neoantigens

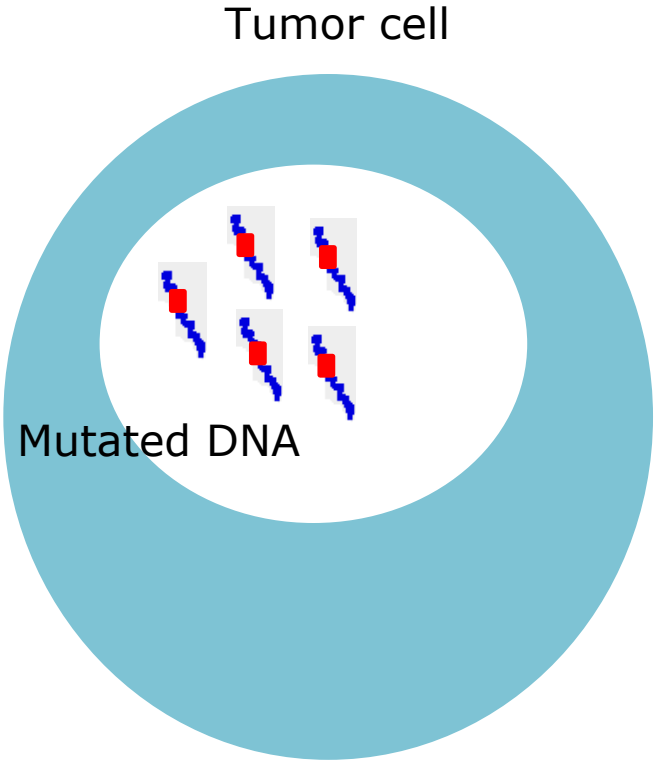
Genetic alterations

- Point mutations
  - silent, ***missense***, nonsense
- Insertion/deletion (Indels)
- Splicing variants
- Copy number variations (CNVs)
  - Amplifications/deletions of genomic regions
- Chromosomal rearrangements
  - Gene fusions



Stanford MEDICINE

Which mutations are good targets?

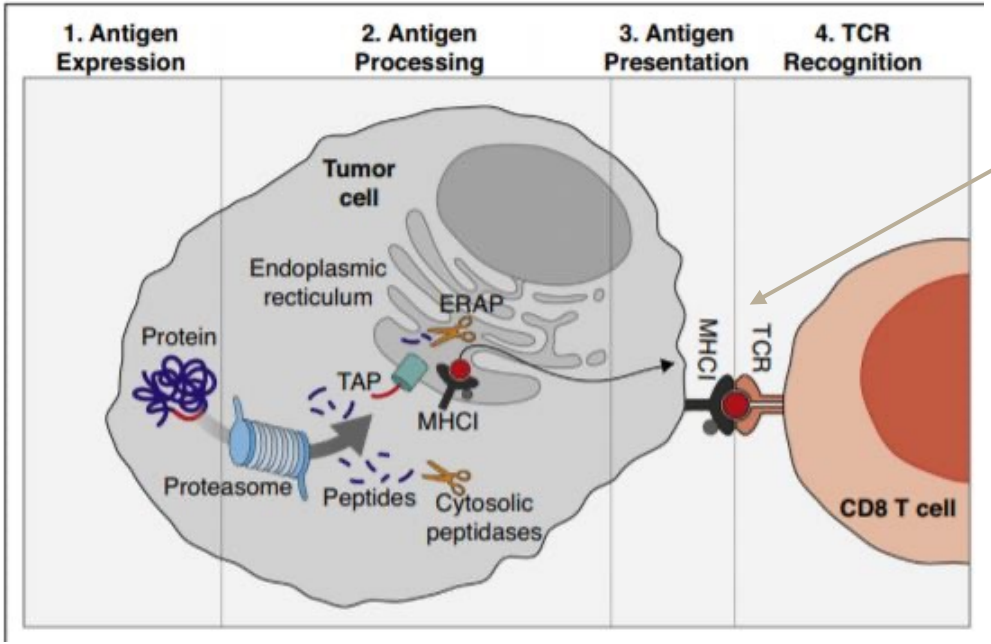


Colorectal Cancer
: ~70 mutations / patient

Which mutations should be in vaccine?

Stanford MEDICINE

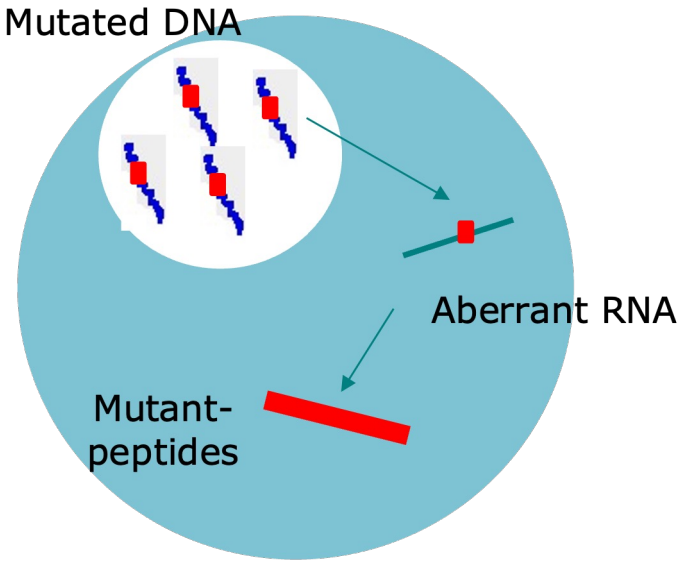
Biological process of presenting neoantigen



Which mutations can produce **peptides** that will be **presented** on the cell surface?

Stanford MEDICINE

Expressed mutations

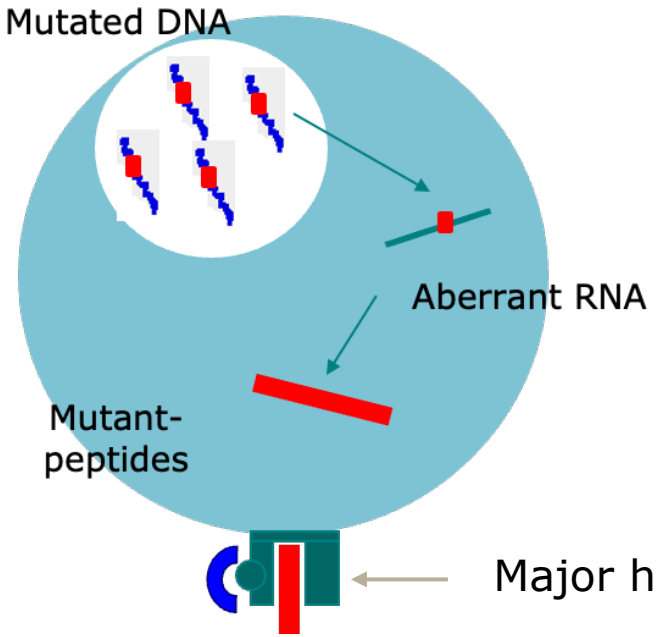


Which mutations in RNAseq?

Mutations produce mutant peptides

Stanford MEDICINE

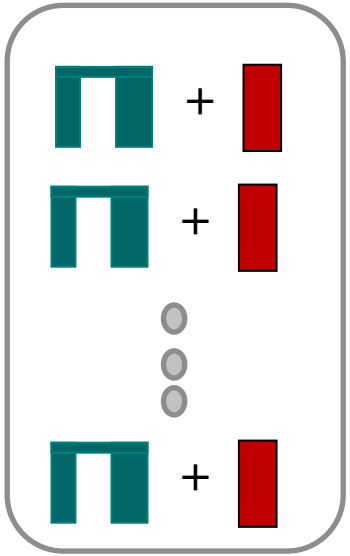
Which mutations will be on cell surface?



- MHC highly variable across people
- 3 different alleles / person
  - HLA-A, HLA-B, HLA-C
  - 3 ~ 6 MHCs per person

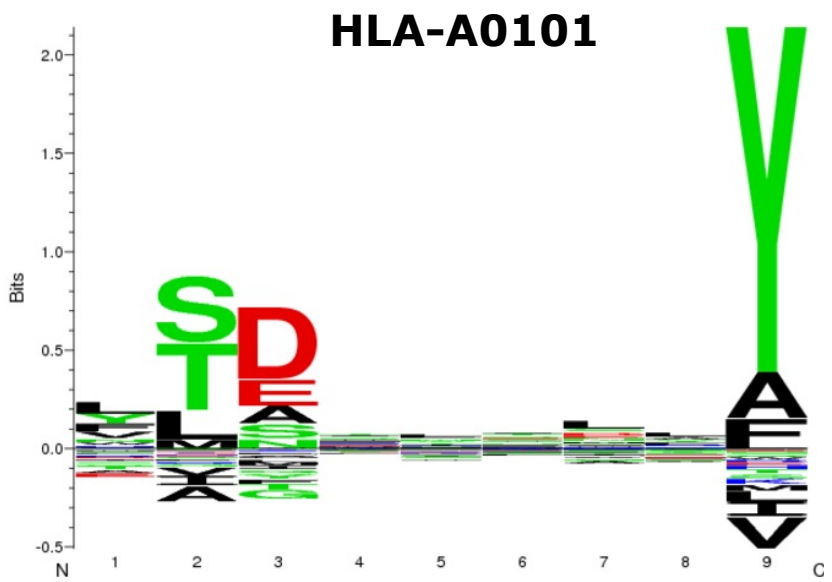
Stanford MEDICINE

Prediction of binding affinity



Experiments data

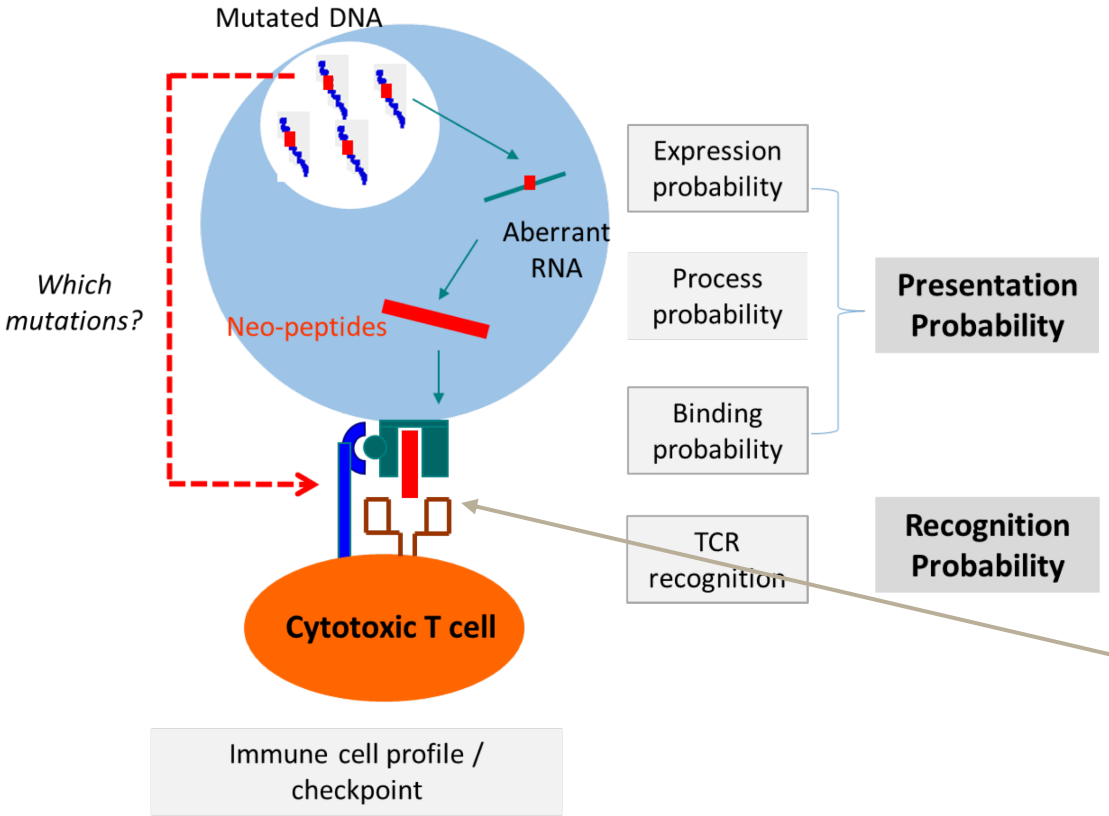
artificial neural networks (ANNs)



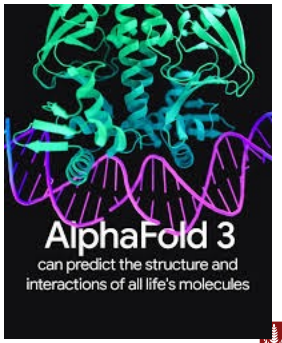
Predict binding affinity

Stanford MEDICINE

Analysis pipeline for good targets



We need only WES & RNAseq



Stanford MEDICINE

Summary

Genomic and image data is able to guide to make better clinical decision

Presence of tumor infiltrating lymphocytes (TILs) → Checkpoint inhibitors

+

Identification of neoantigens → Cancer vaccine

Amount of immune cells → Cell Therapy

Stanford MEDICINE

Table of Contents

Stanford MEDICINE

Cancer stages at diagnosis

	Breast	Lung	Colon
Localized	65.9	29.1	33.4
Regional/distant	32.3	65.6	61

How to detect cancer earlier?

How to treat them effectively?

https://progressreport.cancer.gov/diagnosis/stage
Estimation based on 2021

Stanford MEDICINE

Future perspectives

GRAIL

natera™

Genomic data

Tumor DNA

Immune DNA

Prevention
(by early detection)

Best treatment for patients
(Precision Medicine)

LLM on EHR

Wearable Devices

New Drug

Molecular data:
Protein, Antibody, small molecules

DATA IS KING!!