National Human Genome Research Institute

Machine Learning in Genomics: Tools, Resources, Clinical Applications and Ethics

National Institutes of Health

Genomics in the Machine Learning Space

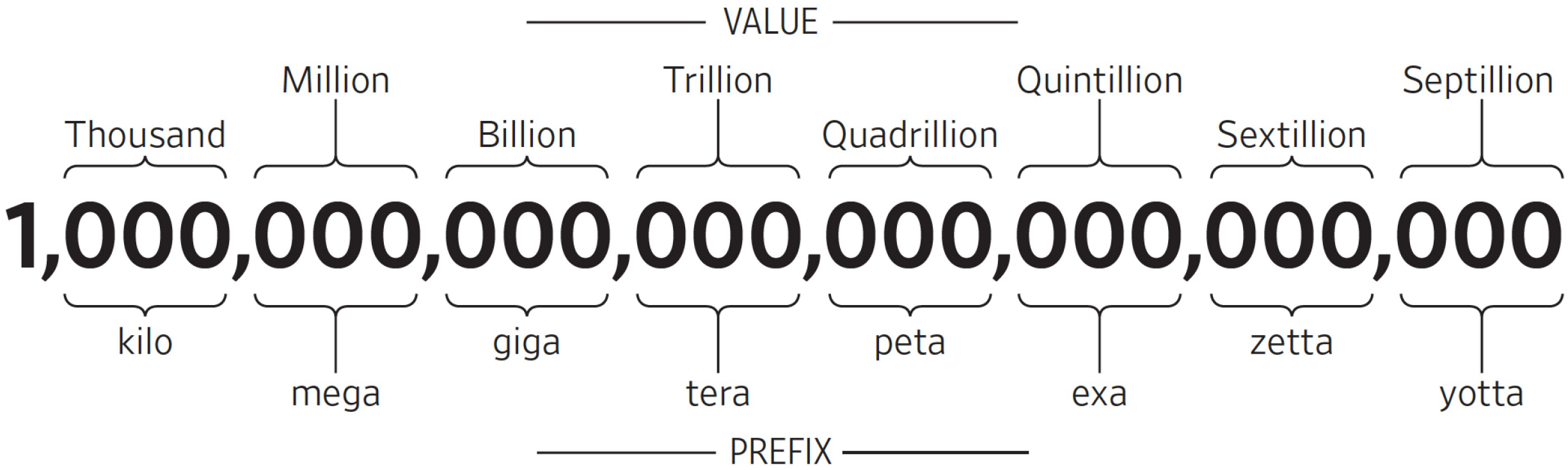
@erictopol

Scripps Research
Science Changing Life

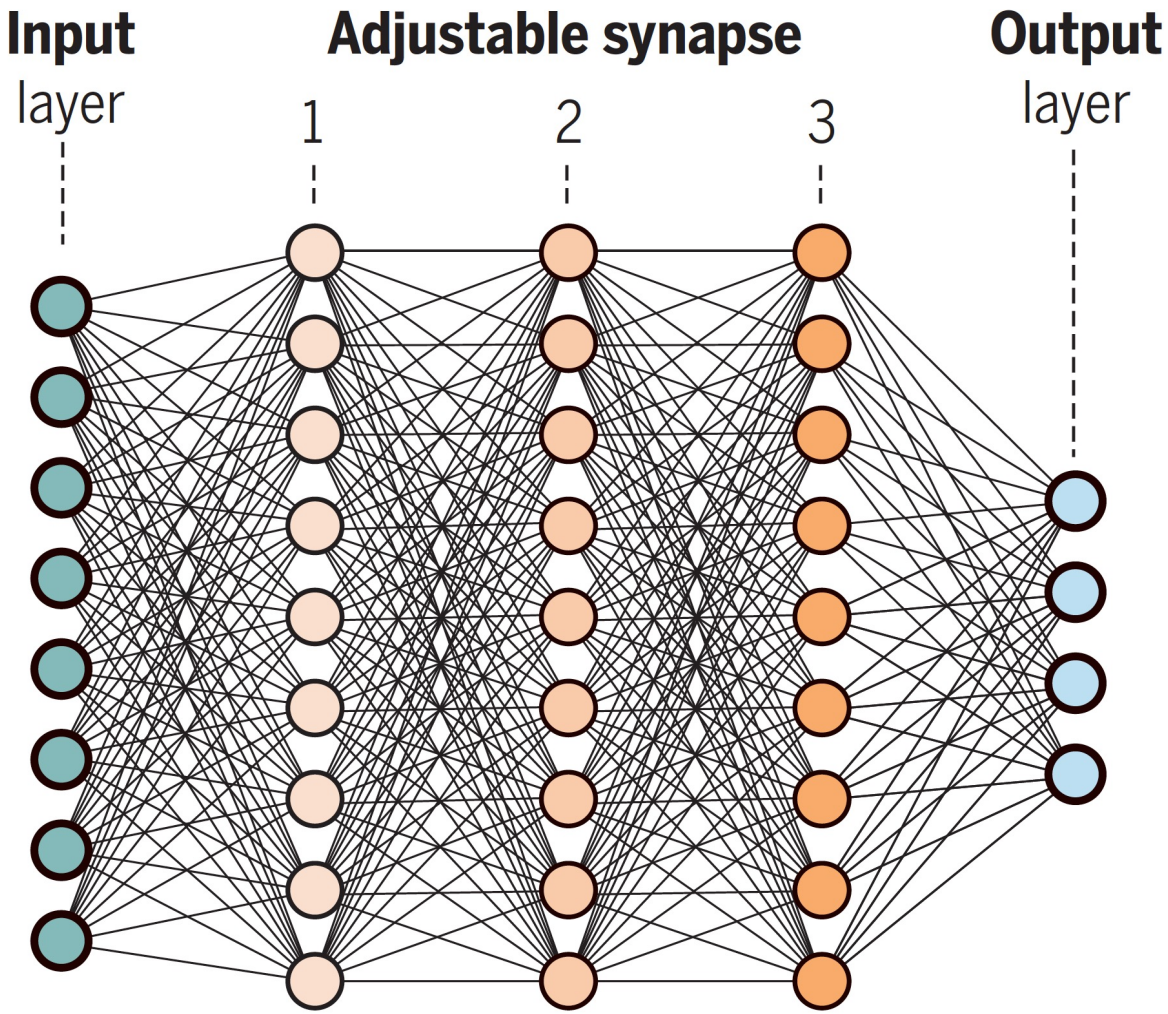
Scripps Research
Translational Institute


Microhabitats save mammals, but not birds, from warming pp. 553 & 633
Gut microbiota modulate immunotherapy pp. 573, 595, & 602
Physically distanced quantum gates pp. 576 & 614
Science
$15
5 FEBRUARY 2021
sciencemag.org
AAAS
SPECIAL ISSUE
HUMAN GENOME AT
20


The international journal of science / 11 February 2021
nature
GENOME REVOLUTION
How 20 years of the human genome sequence have helped reshape science
Coronavirus
The power and pitfalls of rapid tests for COVID-19
Into the unknown
Quantum technology offers boost to the hunt for dark matter
The hear and now
Sounding out how the middle ear evolved in mammals
Vol. 590, No. 7845
nature.com

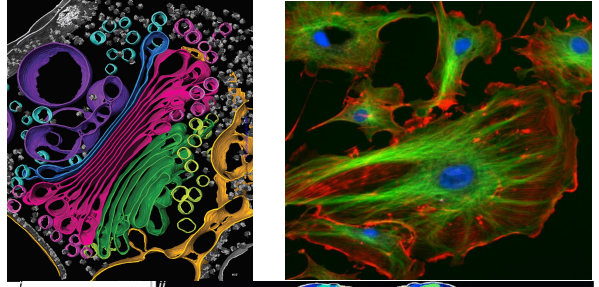

VALUE
Thousand
Million
Billion
Trillion
Quadrillion
Quintillion
Sextillion
Septillion
1,000,000,000,000,000,000,000,000,000
kilo
mega
giga
tera
peta
exa
zetta
yotta
PREFIX

The inside story of how Google used artificial intelligence to tra nra nsform one of its most popular services — and how its approach to A.I. is poised to reinvent computing itself BY GID EO N LEWIS-KRAUS Illustration by Jessica Svendsen

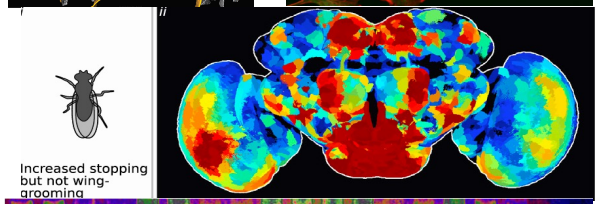


Science 15 FEBRUARY 2019

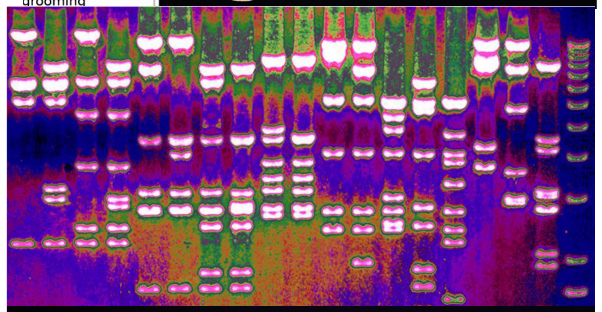
Imaging



Machine vision



Genomics



Protein structure and engineering

Drug discovery

Deep Neural Networks for Genomics

A Short History

2015

Convolutional networks
(DeepBind, DeepSEA, Basset)

ERIC AND WENDY
SCHMIDT CENTER
AT BROAD INSTITUTE

BROAD
INSTITUTE

Mila

ellis
European Laboratory for Learning and Intelligent Systems

Microsoft

BIG DATA INSTITUTE

UNIVERSITY OF OXFORD

The Alan Turing Institute

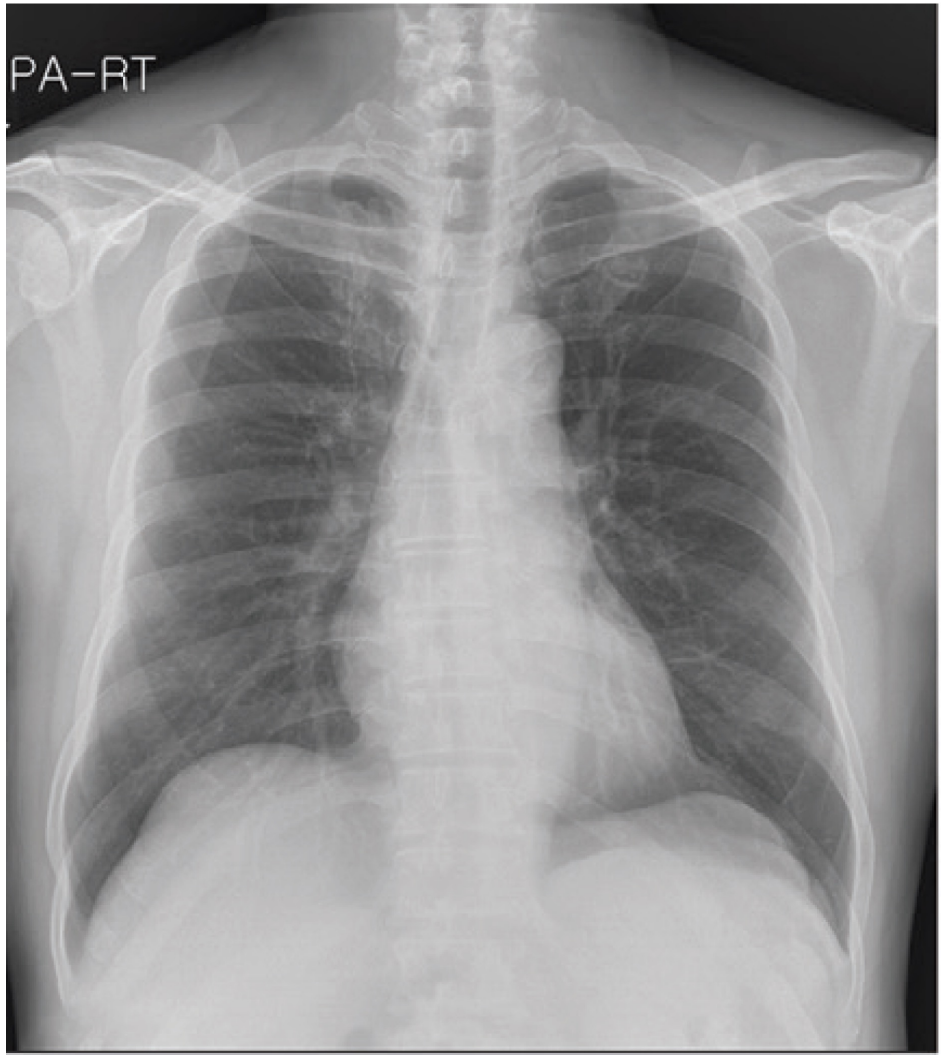
Google DeepMind

NOVARTIS

Genentech

AstraZeneca

Medical Images are Simple Compared with Genomics



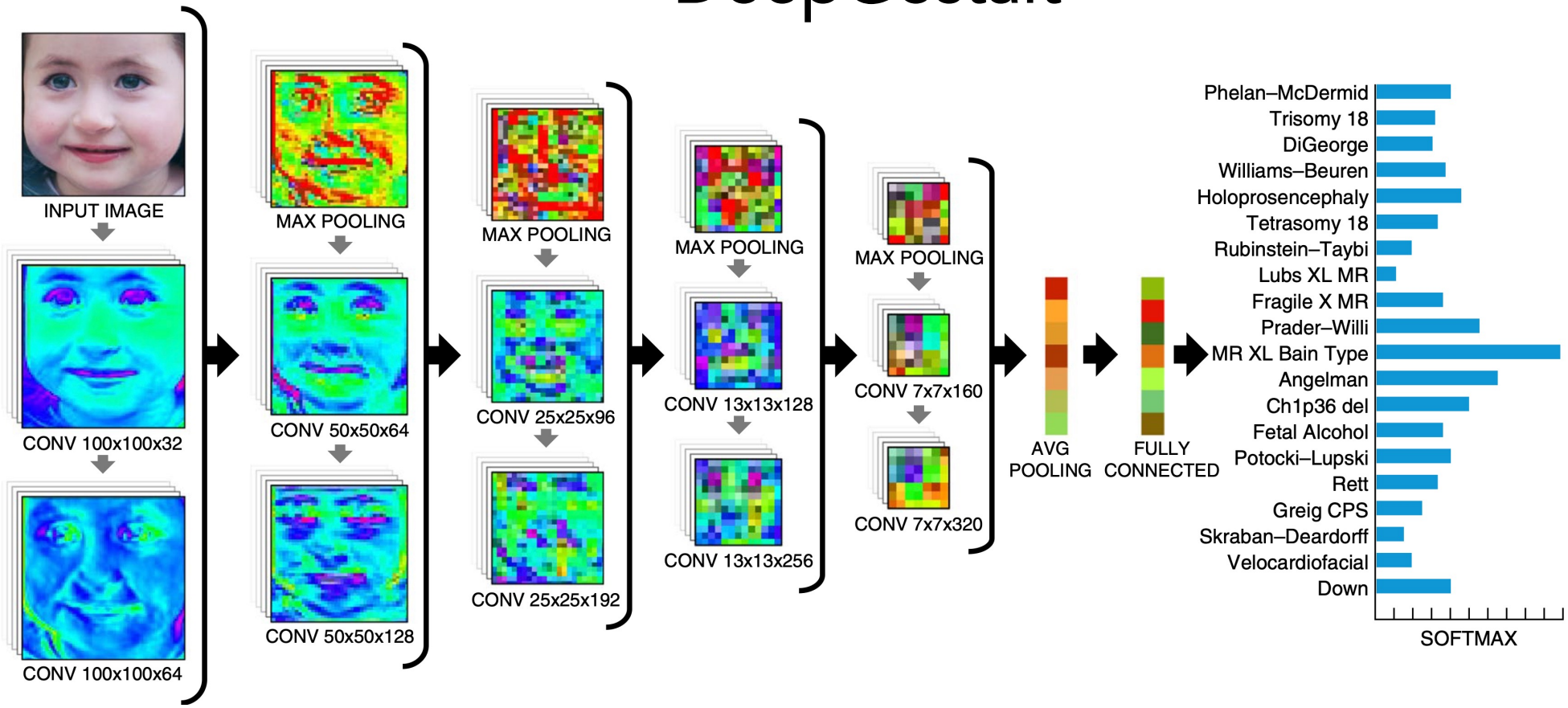

```
1441 cttgaaaacc attcttcgta agggtggtcg cactattgcc tttggaggct gtgtgttctc
1501 ttatgttggt tgccataaca agtgtgccta ttgggttcca cgtgctagcg ctaacatagg
1561 ttgtaaccat acaggtgttg ttggagaagg ttccgaaggt cttaatgaca accttcttga
1621 aatactccaa aaagagaaag tcaacatcaa tattgttggt gactttaaac ttaatgaaga
1681 gatcgccatt attttggcat ctttttctgc ttccacaagt gcttttgtgg aaactgtgaa
1741 aggtttggat tataaagcat tcaaacaaat tgttgaatcc tgtggtaatt ttaaagttac
1801 aaaaggaaaa gctaaaaaag gtgcctggaa tattggtgaa cagaaatcaa tactgagtcc
1861 tctttatgca tttgcatcag aggctgctcg tgttgtacga tcaattttct cccgcactct
1921 tgaaactgct caaaattctg tgcgtgtttt acagaaggcc gctataacaa tactagatgg
1981 aatttcacag tattcactga gactcattga tgctatgatg ttcacatctg atttggctac
2041 taacaatcta gttgtaatgg cctacattac aggtggtgtt gttcagttga cttcgcagtg
2101 gctaactaac atctttggca ctgtttatga aaaactcaaa cccgtccttg attggcttga
2161 agagaagttt aaggaaggtg tagagtttct tagagacggt tgggaaattg ttaaatttat
2221 ctcaacctgt gcttgtgaaa ttgtcggtgg acaaattgtc acctgtgcaa aggaaattaa
2281 ggagagtgtt cagacattct ttaagcttgt aaataaattt ttggctttgt gtgctgactc
2341 tatcattatt ggtggagcta aacttaaagc cttgaattta ggtgaaacat ttgtcacgca
2401 ctcaaaggga ttgtacagaa agtgtgttaa atccagagaa gaaactggcc tactcatgcc
2461 tctaaaagcc ccaaaagaaa ttatcttctt agagggagaa acacttccca cagaagtgtt
2521 aacagaggaa gttgtcttga aaactggtga tttacaacca ttagaacaac ctactagtga
2581 agctgttgaa gctccattgg ttggtacacc agtttgtatt aacgggctta tgttgctcga
2641 aatcaaagac acagaaaagt actgtgccct tgcacctaat atgatggtaa caaacaatac
2701 cttcacactc aaaggcggtg caccaacaaa ggttactttt ggtgatgaca ctgtgataga
2761 agtgcaaggt tacaagagtg tgaatatcac ttttgaactt gatgaaagga ttgataaagt
2821 acttaatgag aagtgctctg cctatacagt tgaactcggt acagaagtaa atgagttcgc
2881 ctgtgttgtg gcagatgctg tcataaaaac tttgcaacca gtatctgaat tacttacacc
2941 actgggcatt gatttagatg agtggagtat ggctacatac tacttatttg atgagtctgg
3001 tgagtttaaa ttggcttcac atatgtattg ttctttctac cctccagatg aggatgaaga
3061 agaaggtgat tgtgaagaag aagagtttga gccatcaact caatatgagt atggtactga
```

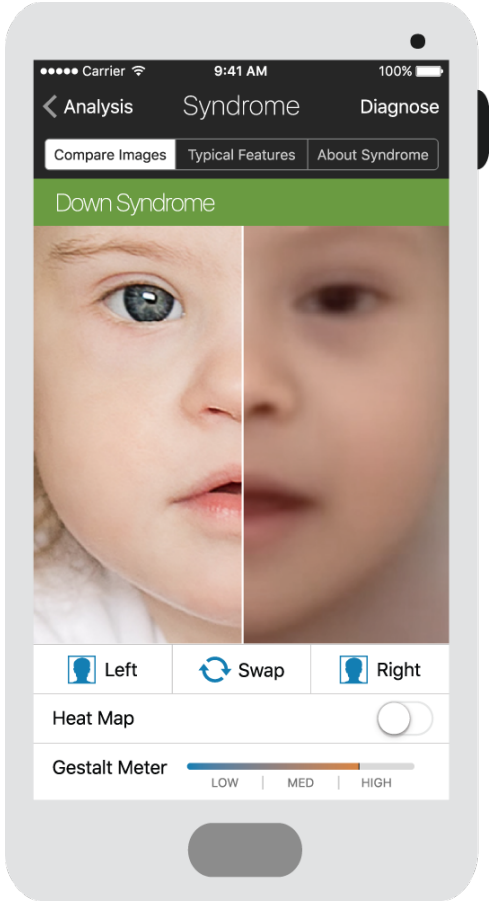
nature medicine

Identifying facial phenotypes of genetic disorders using deep learning

FACE2GENE
Smart Phenotyping. Better Genetics.

DeepGestalt



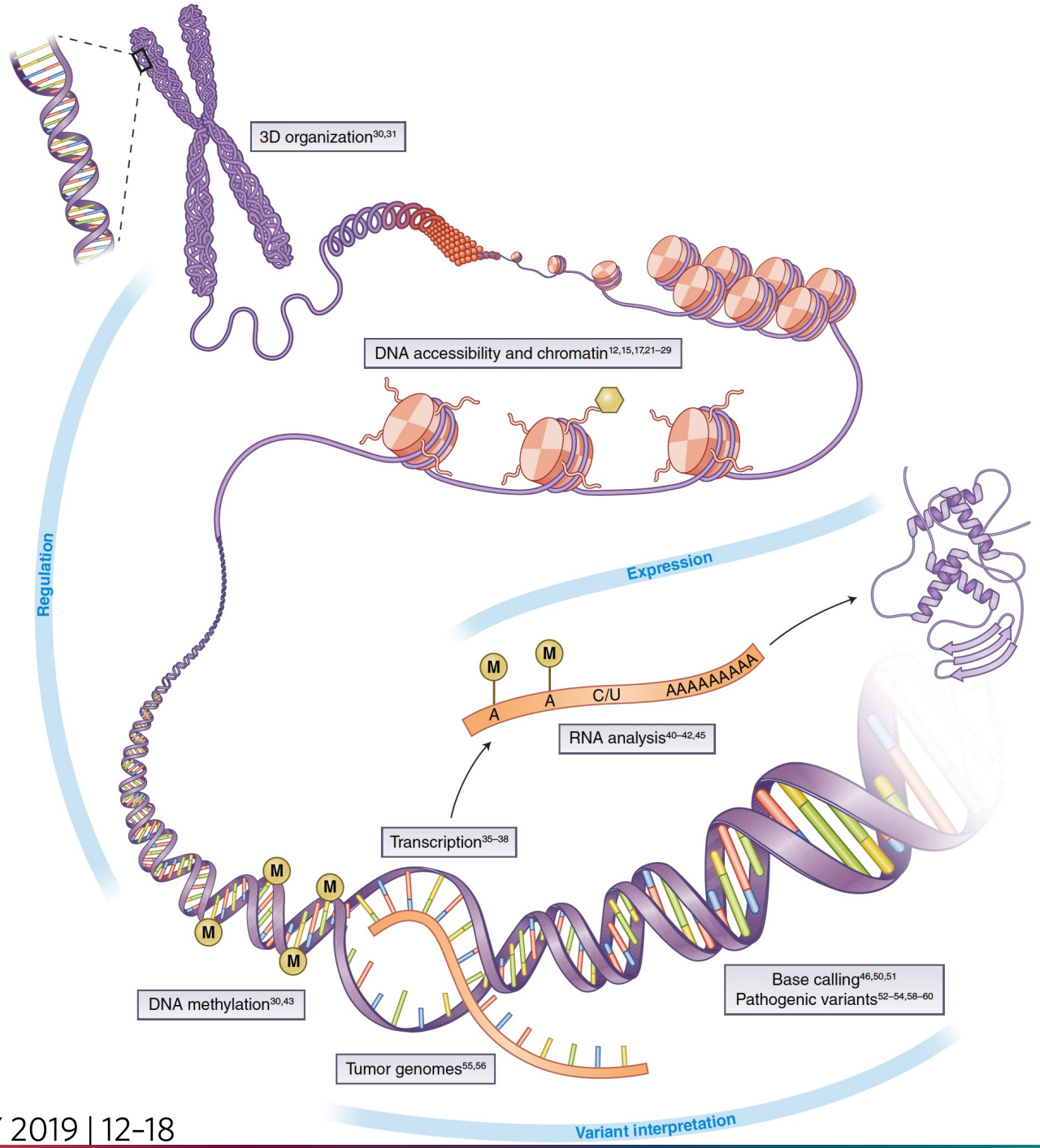


>91% accuracy for > 200 syndromes

Gurovich Y et al, January 2019

nature genetics

Applications of Deep Learning in Genomics



Base calling

Pathogenic variants

Tumor genomes

DNA methylation

RNA analysis

Transcription

DNA accessiblity and chromatin

3D organization

Special challenges:
Functional genomics
The regulatory genome

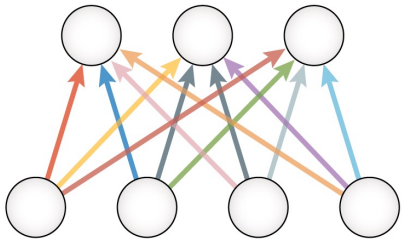
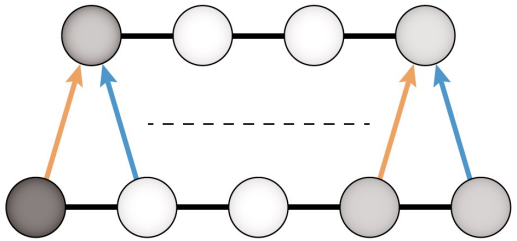
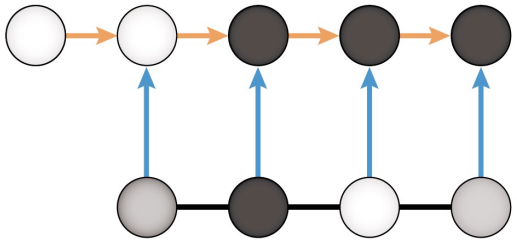
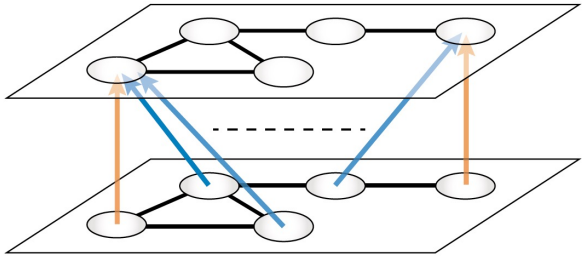
Zou J et al

NATURE GENETICS | VOL 51 | JANUARY 2019 | 12–18

nature REVIEWS GENETICS

Deep learning: new computational modelling techniques for genomics

4 Major Classes of Neural Networks Used in Genomics

Network type	Fully connected	Convolutional	Recurrent	Graph convolutional
Output Parameters Input				
Invariance	–	Translation	Time	Node index permutation
Input example	Predefined features such as number of *k*-mer matches, total conservation	• DNA sequence • Amino acid sequence • Image	• DNA sequence • Amino acid sequence • Time series measurements	• Protein–protein interaction network • Citation network • Protein structure

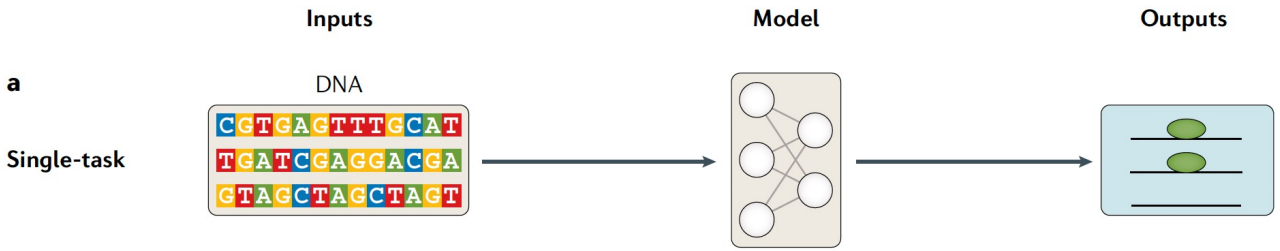
Feed forward

→ Parameterized information flow — Link ○ Node in the neural network (scalar or tensor)

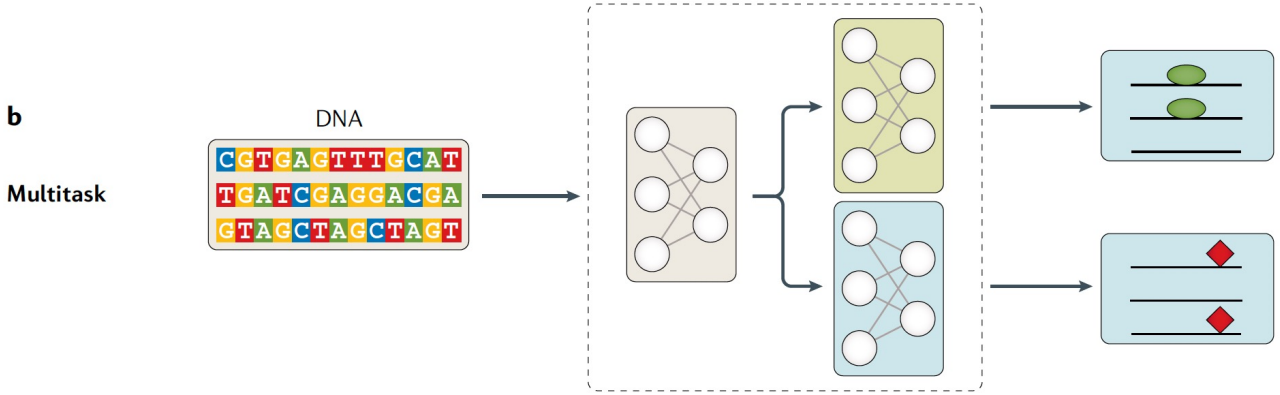
Eraslan G et al, July 2019

nature REVIEWS GENETICS

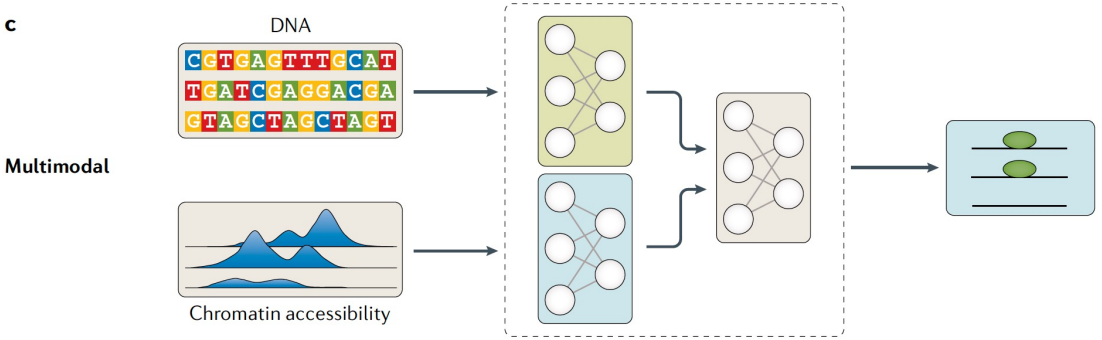
Deep learning: new computational modelling techniques for genomics



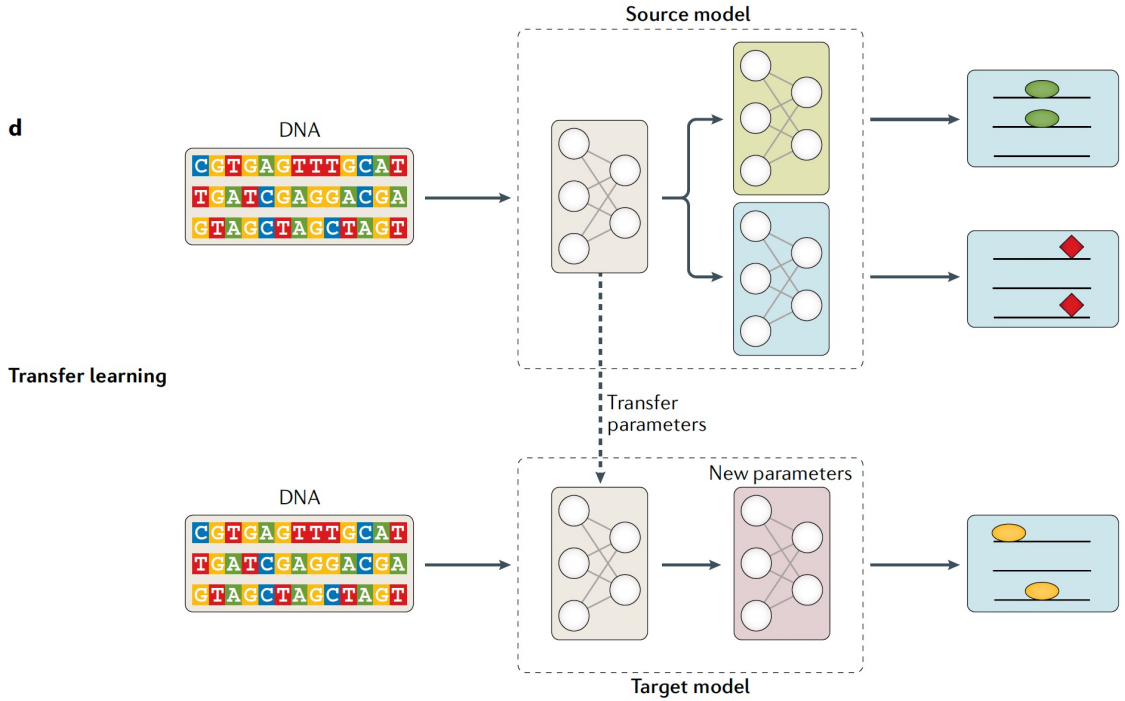
Binding of a single transcription factor



Simultaneous binding of 2 transcription factors



Inputs w/ DNA sequence and chromatin accessibility

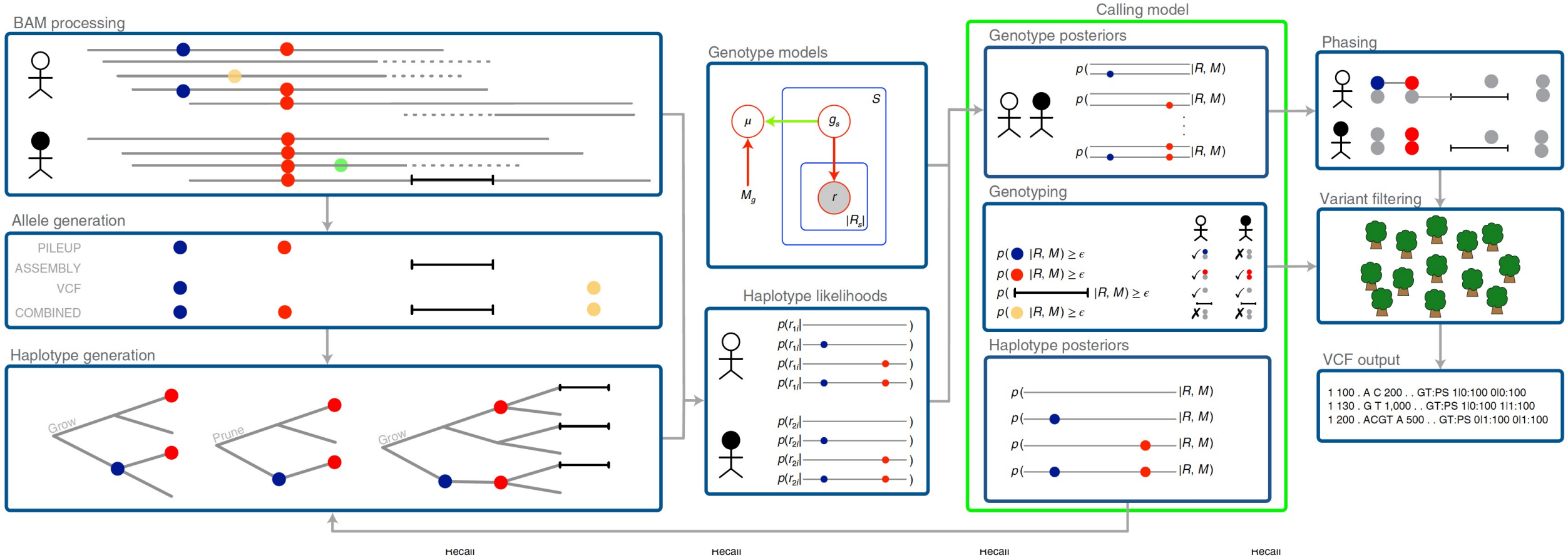


Eraslan G et al, July 2019

nature biotechnology

A unified haplotype-based method for accurate and comprehensive variant calling

Octopus higher sensitivity and specificity that prior variant callers



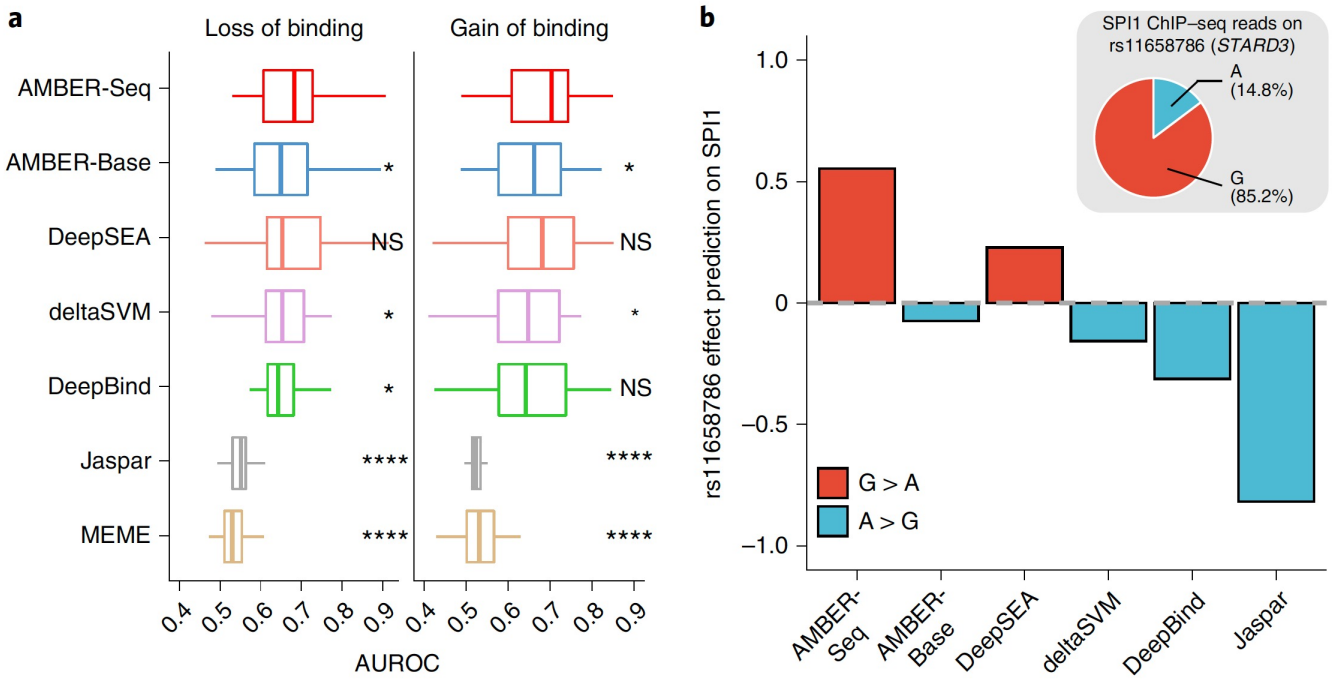
Bayesian model, combines sequence data, prior, to phase-called genotypes

Cooke D et al, March 29, 2021

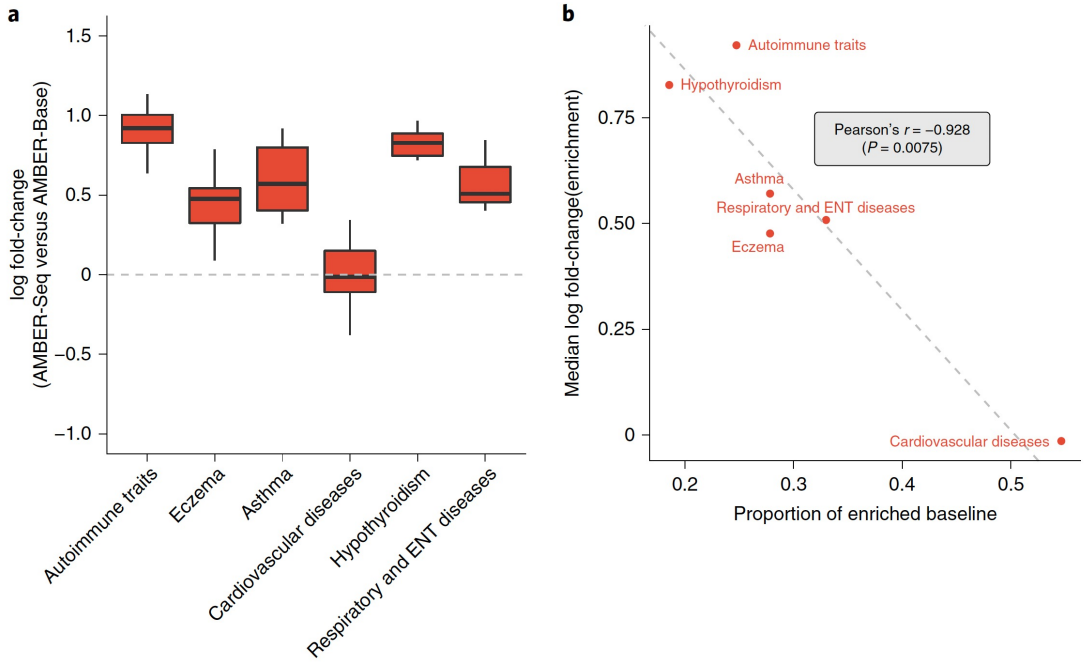
nature machine intelligence

An automated framework for efficiently designing deep convolutional neural networks in genomics

Variant Prediction



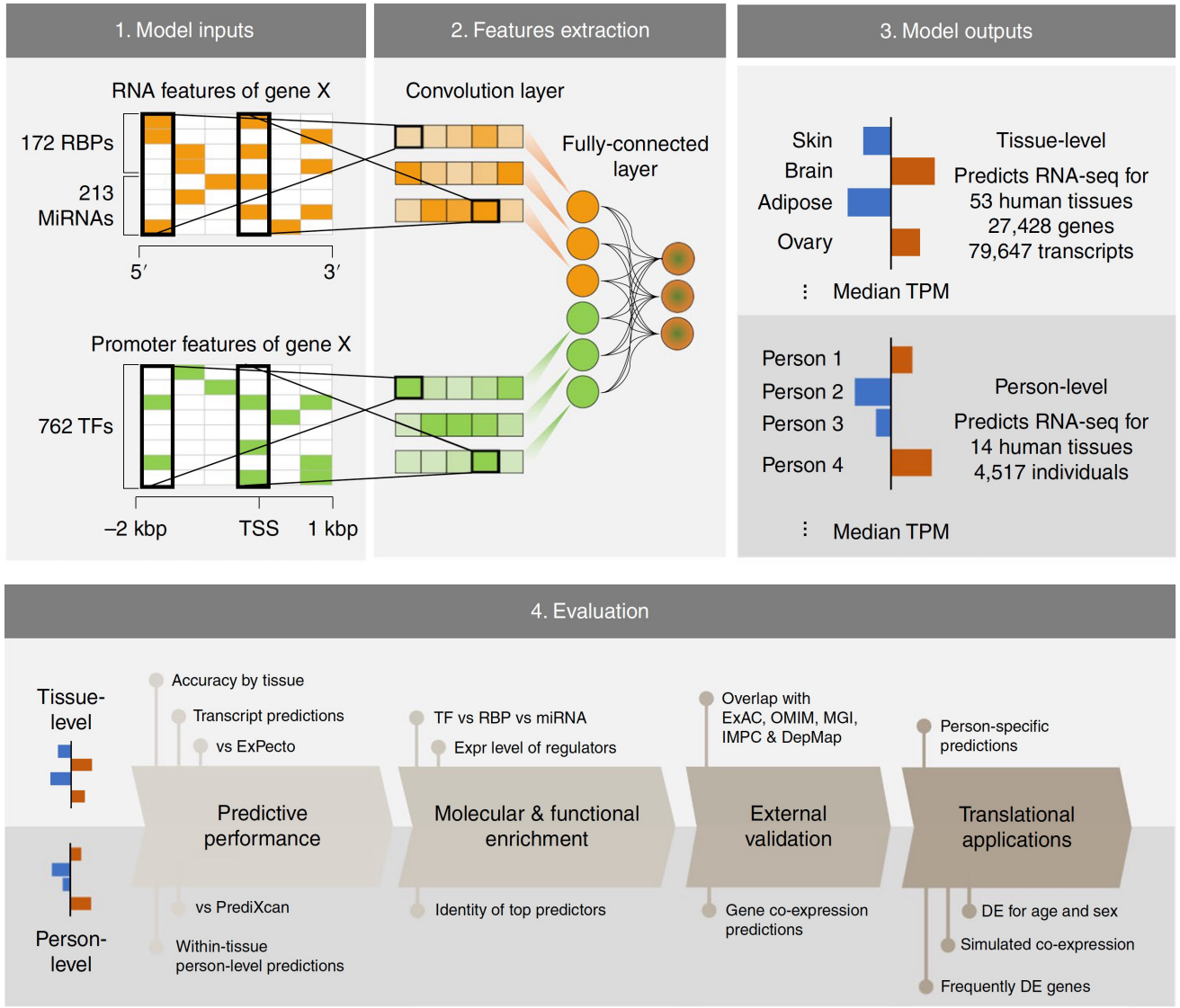
Heritability Enrichment in GWAS



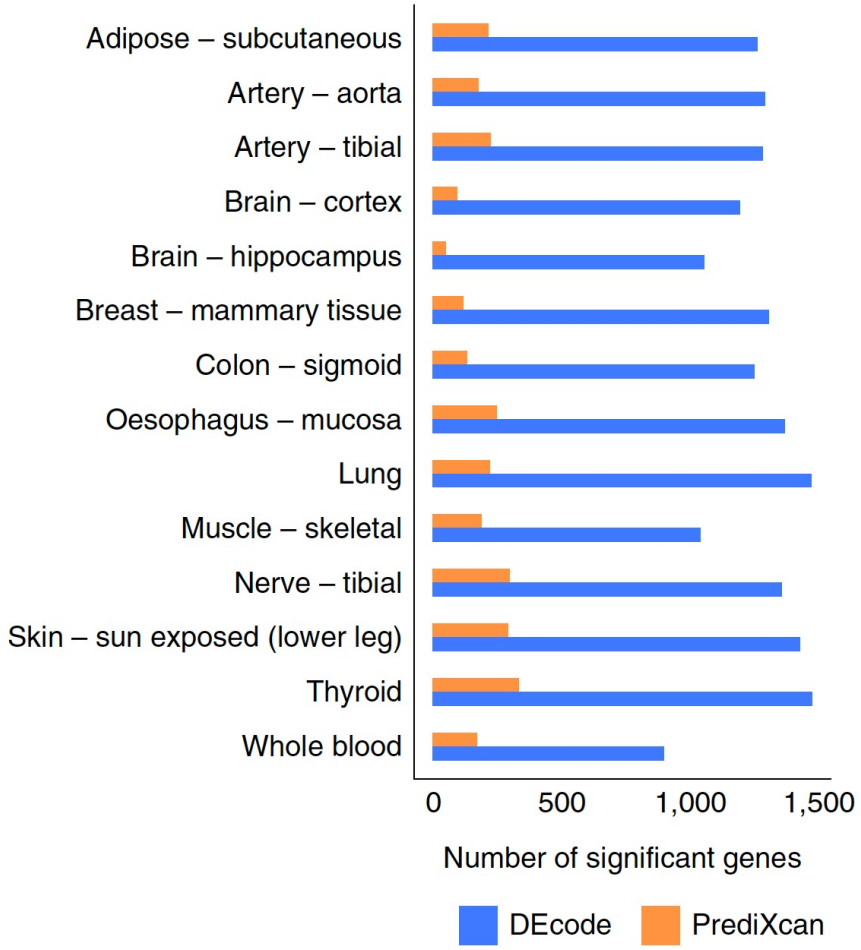
Zhang Z, March 2021

nature machine intelligence

Deep learning decodes the principles of differential gene expression



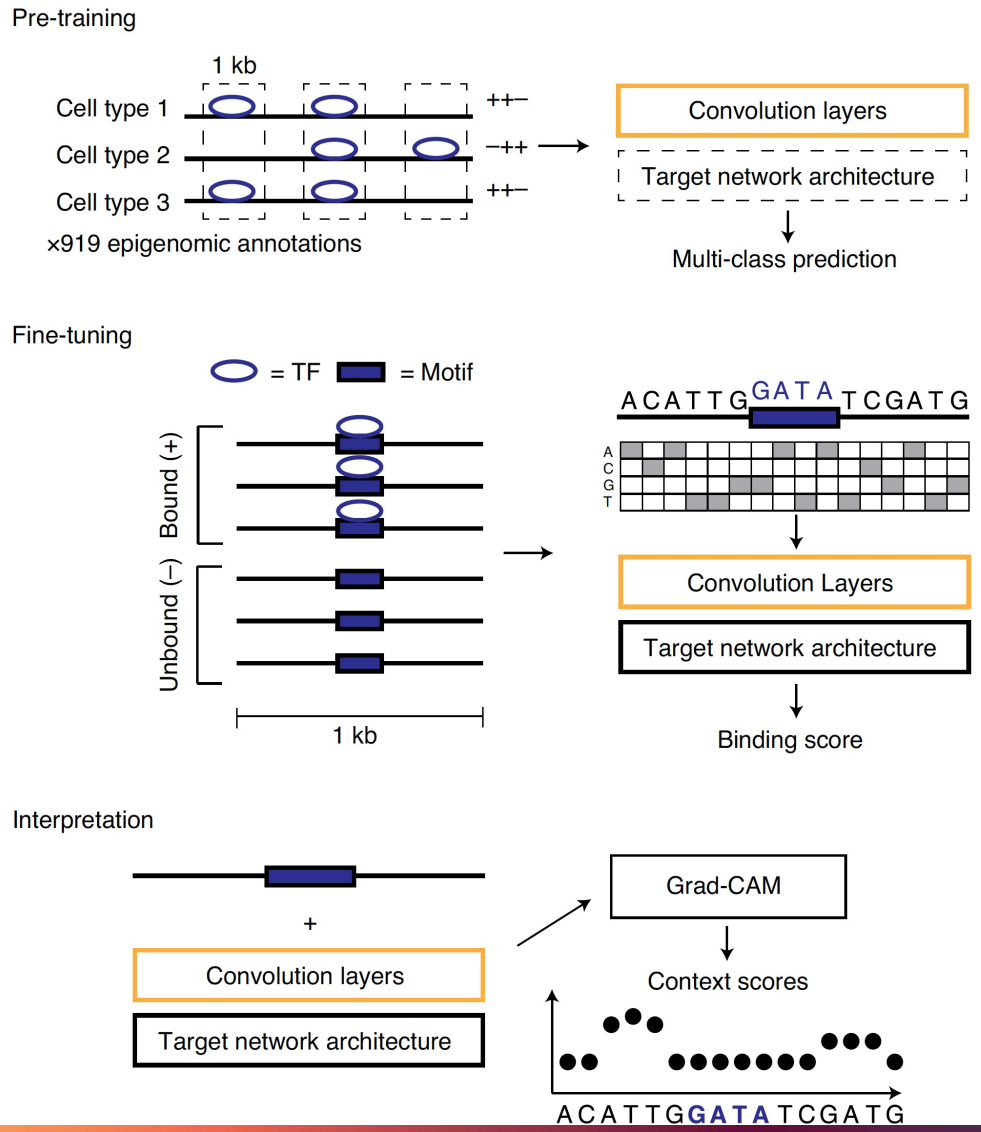
DECode



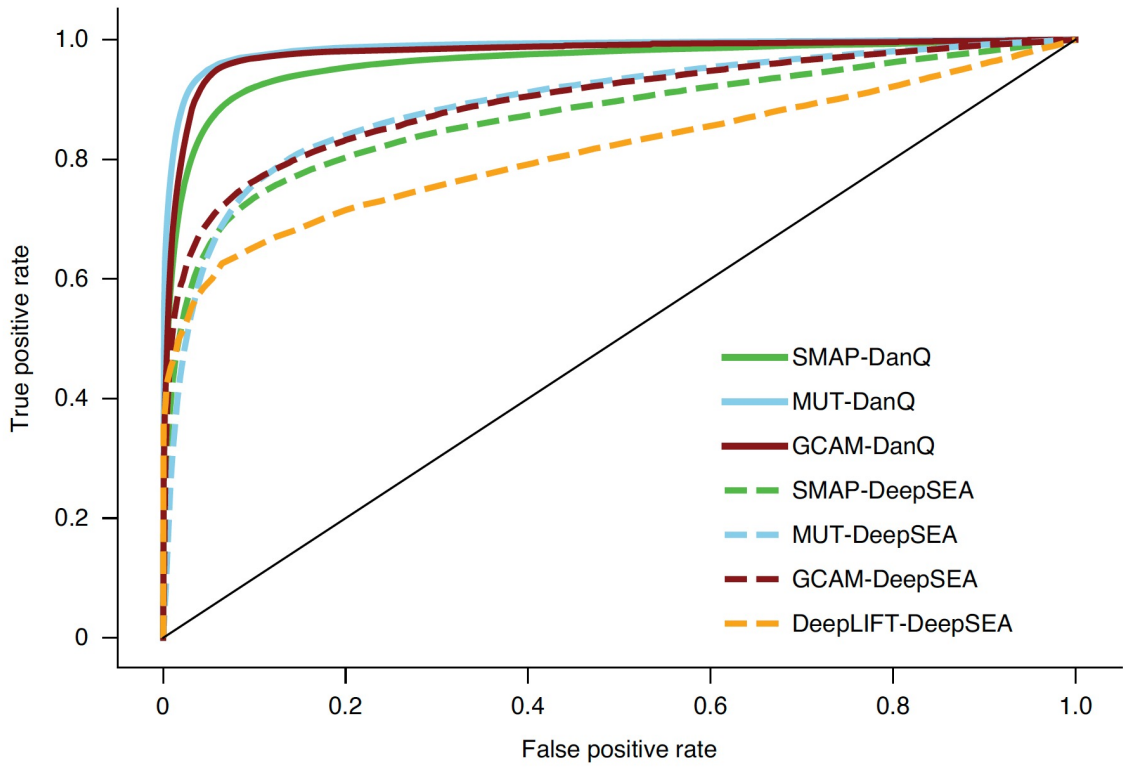
Tasaki S et al , January 2020

nature machine intelligence

Deep neural networks identify sequence context features predictive of transcription factor binding

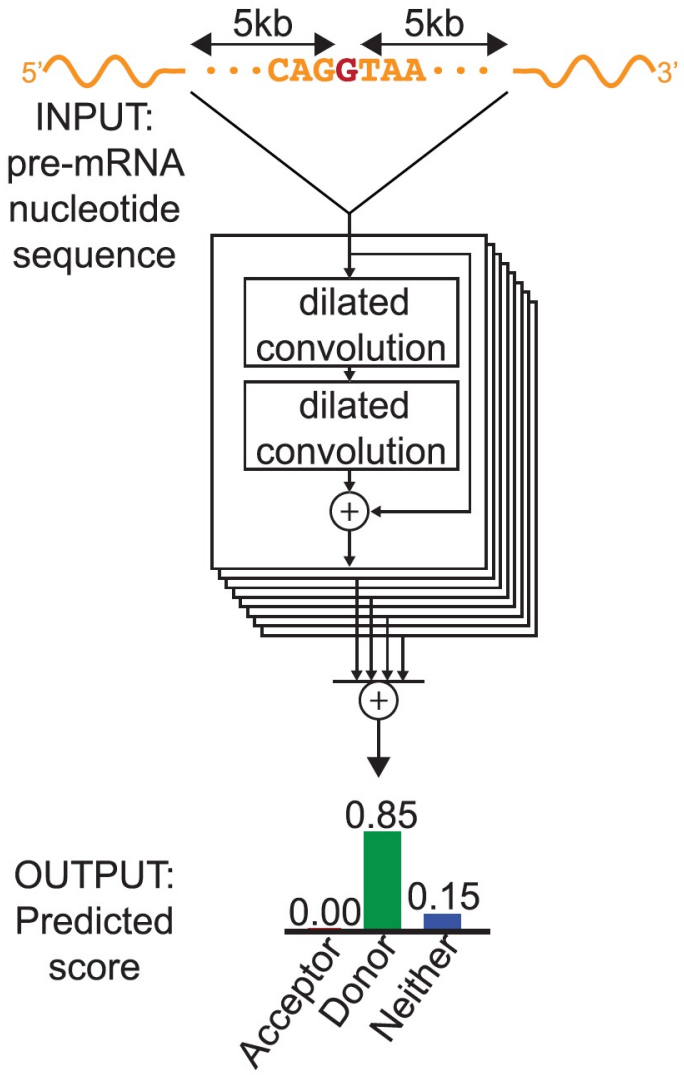


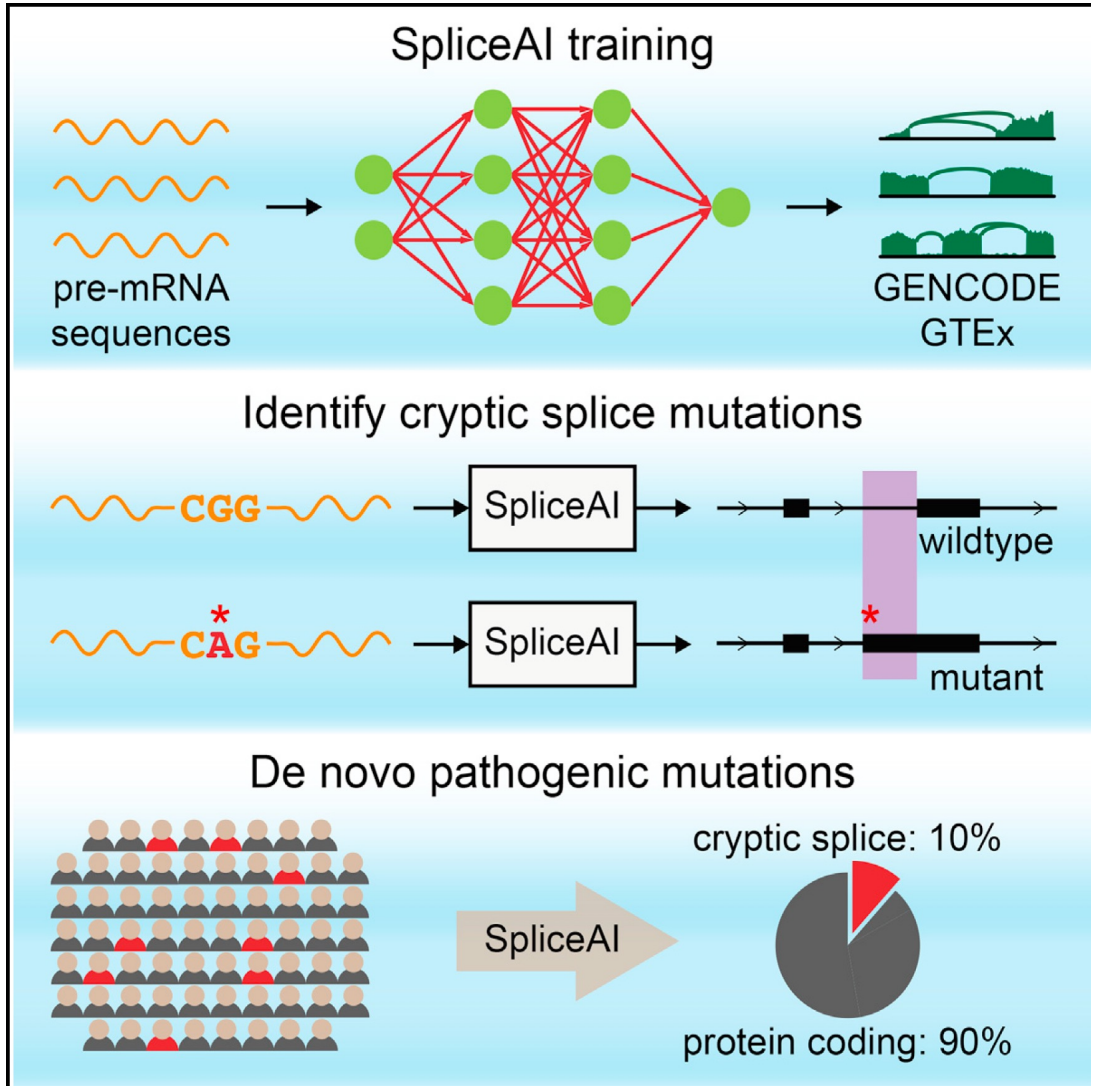
AgentBind



Zheng A et al , January 2021

Predicting Splicing from Primary Sequence with Deep Learning

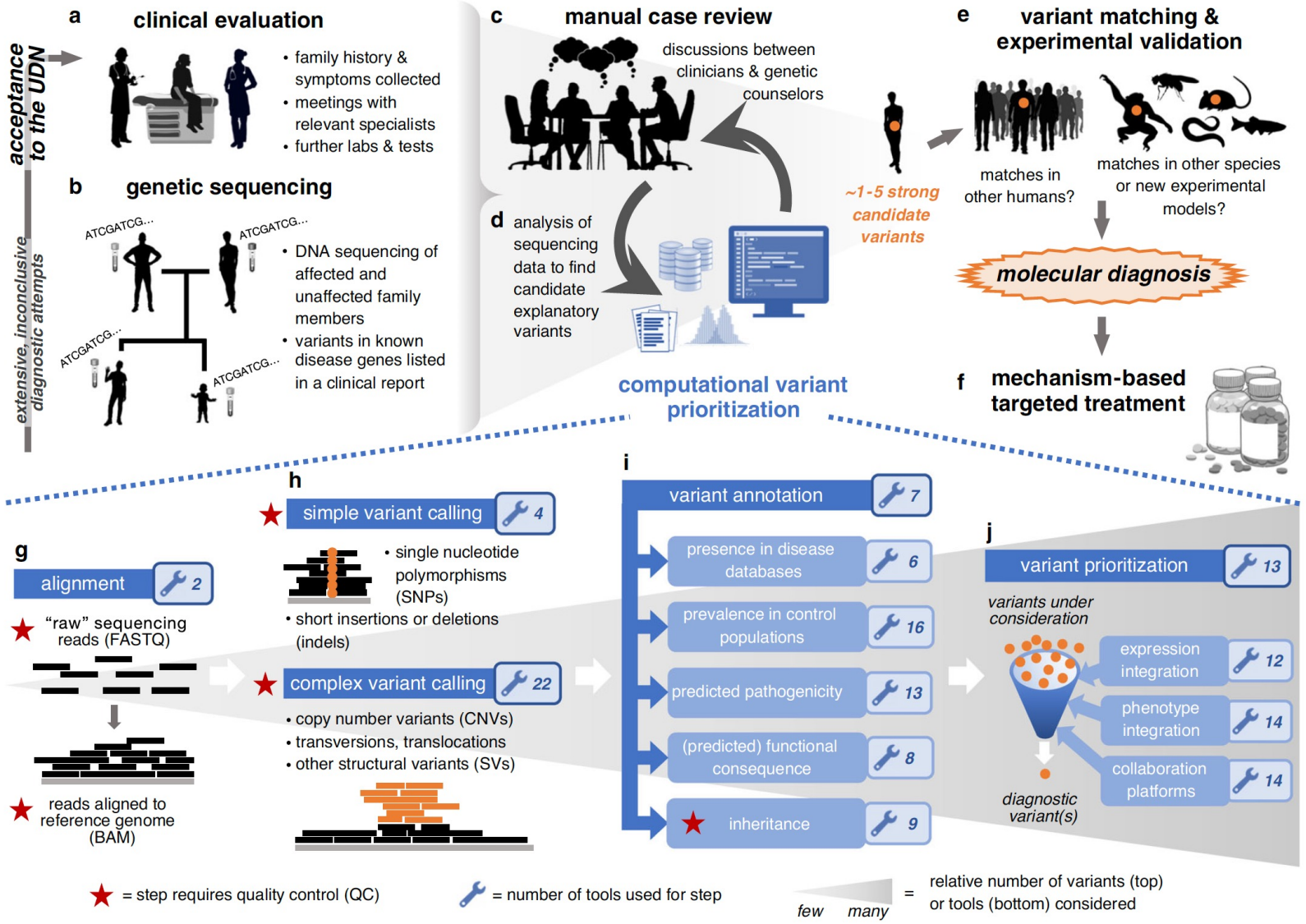




Jaganathan et al, January 2019

Genetics in Medicine

Commonalities across computational workflows for uncovering explanatory variants in undiagnosed cases

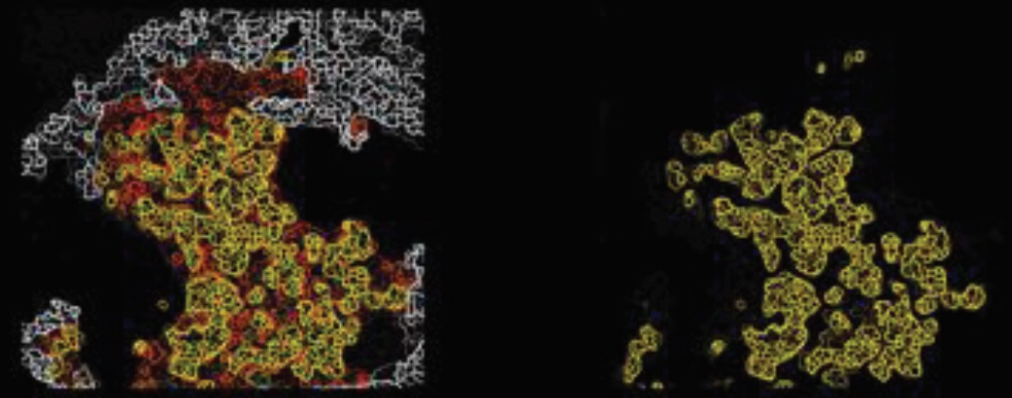


Kobren S et al, February 2021

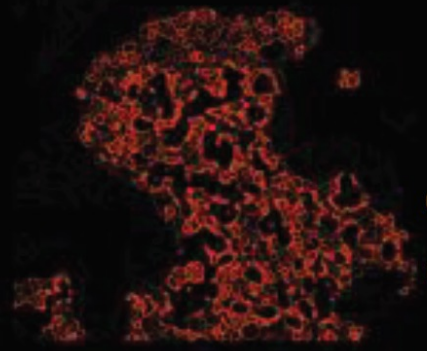
OUTLOOK
Cannabis

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



CANCER DECONSTRUCTED



Fluorescent labelling reveals the changing cellular environment of early-stage metastasis

PAGES 589 & 603

ARCHAEOLOGY

HANDLE WITH CARE

Ancient remains need safeguards from sequencing

PAGE 581

ELECTRONICS

FIRING ON ALL CYLINDERS

A microprocessor made from carbon-nanotube transistors

PAGES 588 & 595

REGENERATIVE BIOLOGY

DIVISION OF LABOUR

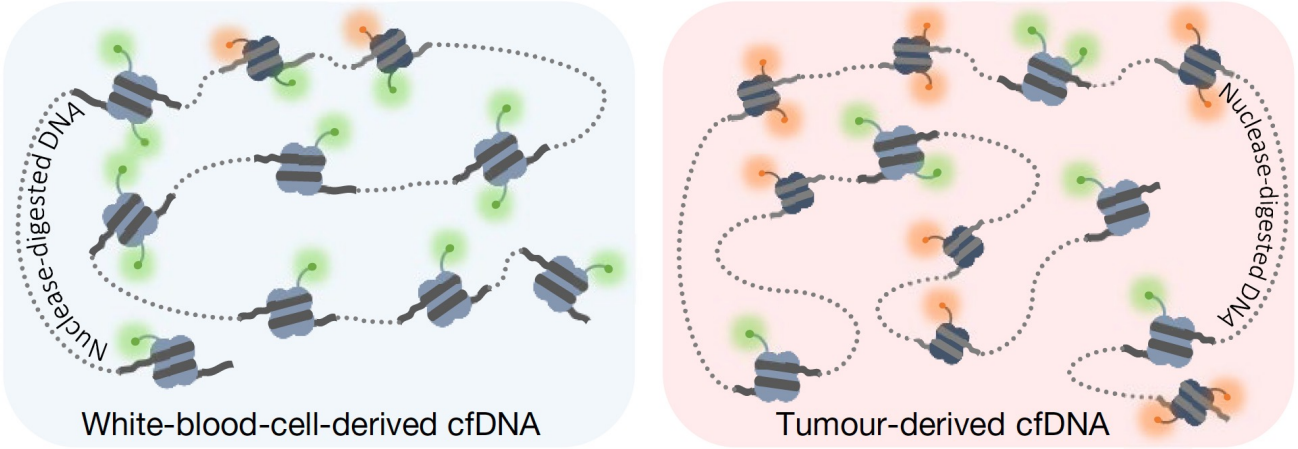
The molecular cues that prompt fission in flatworms

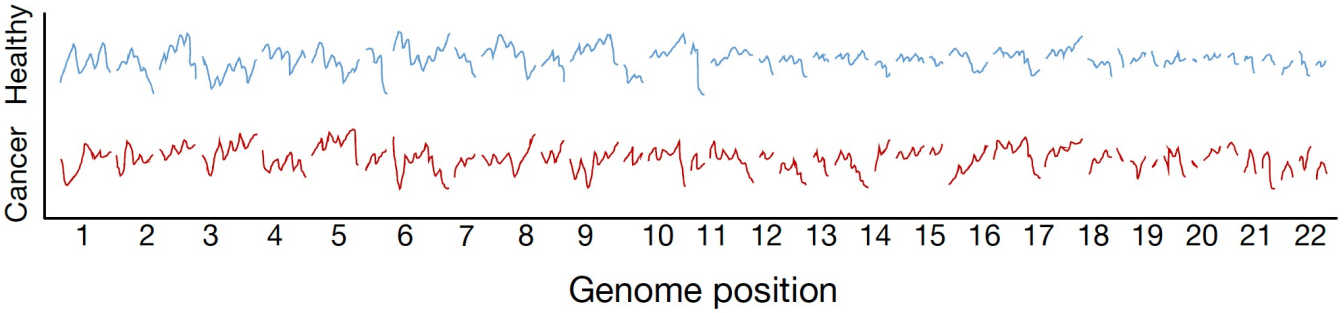
PAGES 593 & 655

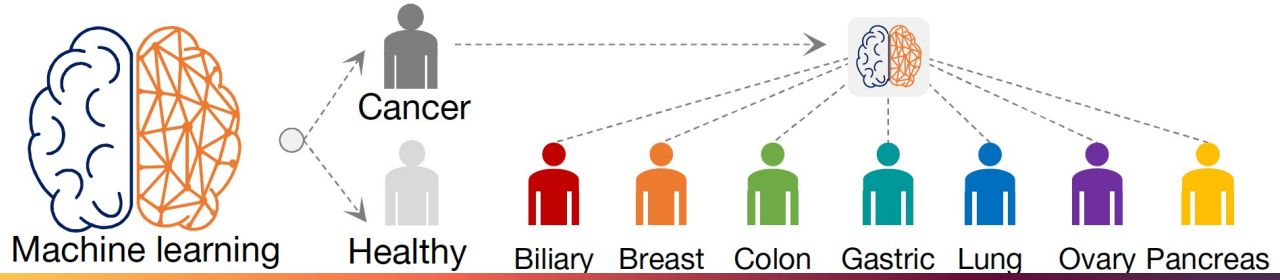
NATURE.COM
29 August 2019
Vol. 572, No. 7771

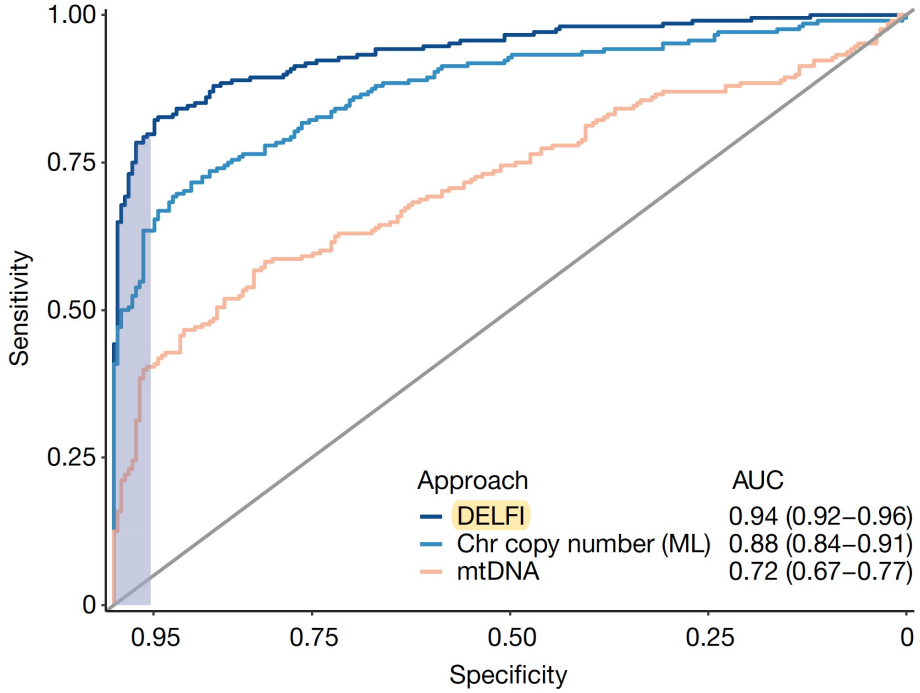
Genome-wide cell-free DNA fragmentation in patients with cancer

Noninvasive cancer screening (DELFI)







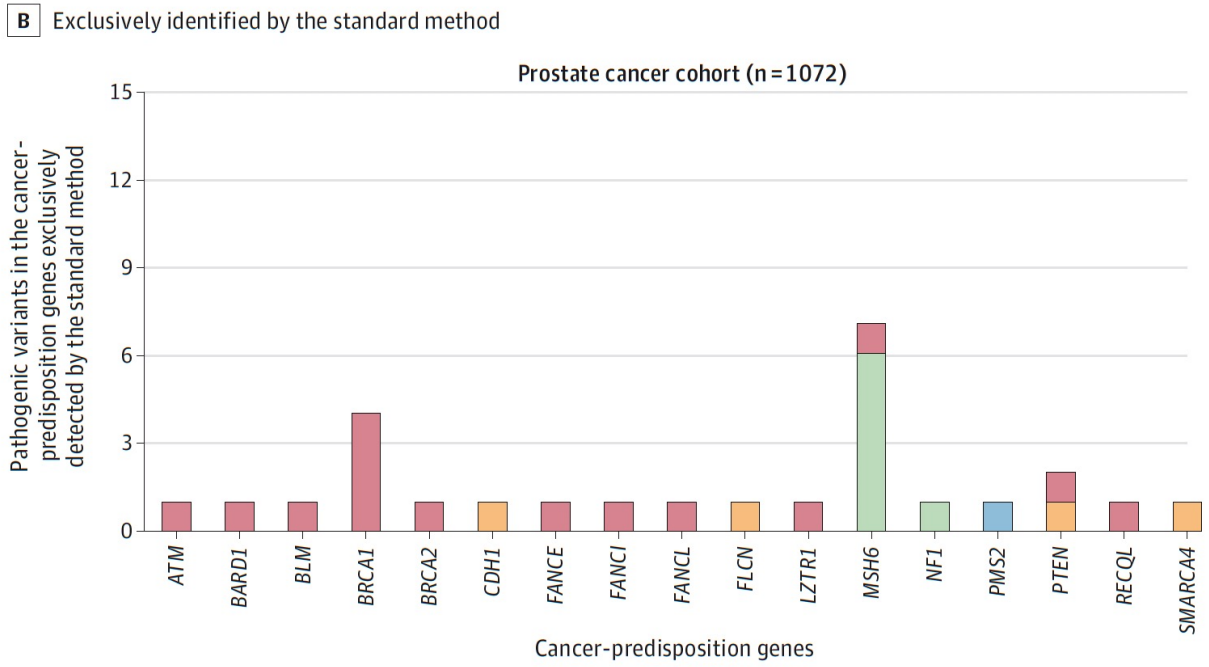


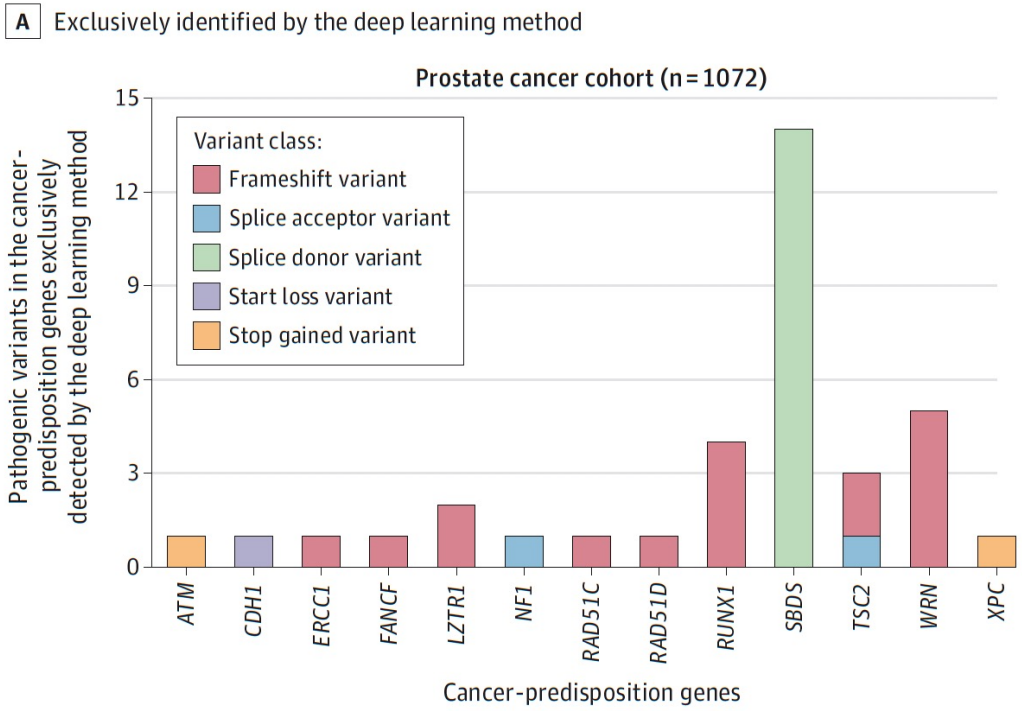
nature

Cristiano S et al, January 2019

JAMA

Detection of Pathogenic Variants With Germline Genetic Testing Using Deep Learning vs Standard Methods in Patients With Prostate Cancer and Melanoma



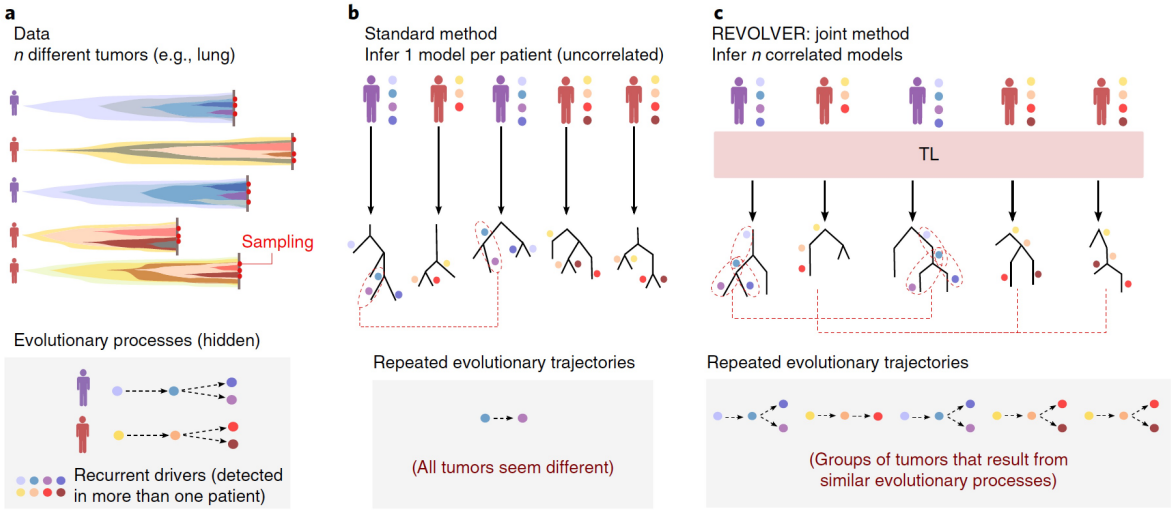


AlDubayan S et al, November 2020

nature methods
ARTICLES
https://doi.org/10.1038/s41592-018-0108-x

Detecting repeated cancer evolution from multi-region tumor sequencing data

Giulio Caravagna[1,2*], Ylenia Giarratano[2,3], Daniele Ramazzotti[4], Ian Tomlinson[5], Trevor A. Graham[6], Guido Sanguinetti[2*] and Andrea Sottoriva[1*]



DAILY EXPRESS
THE WORLD'S GREATEST NEWSPAPER express.co.uk
WEATHER: DRY AND SUNNY
SATURDAY SEPTEMBER 1, 2018 90p

BREXIT DEAL 'DONE AND DUSTED' IN WEEKS
SEE PAGE 7

EXCLUSIVE Experts hail 'exciting' medical breakthrough

ROBOT WAR ON CANCER

- **Artificial Intelligence predicts tumour growth**
- **Scientists hope to stay one step ahead of disease**

By **Giles Sheldrick**
Chief Reporter

A COMPUTER tool that uses artificial intelligence could save the lives of thousands of cancer patients.

The machine, designed in Britain, can learn to predict how tumours will grow, evolve and spread, scientists revealed last night.

That will enable doctors to tackle the disease earlier and tailor drug treatment to each individual.

The technology has the potential to forecast whether a

how likely it is to respond to treatment and what drug combinations might work.

The new technique, which has been shown to work in tests on historic tissue samples, could be in use in cancer clinics within a few years.

The work is being carried out by a team led by the Institute of Cancer Research in London.

The institute's Dr Andrea Sottoriva said: "It's an exciting

What would Poldark say, Demelza?
THE LEADING LADIES WHO STEAL A KISS IN NEW MOVIE SEE PAGE 3

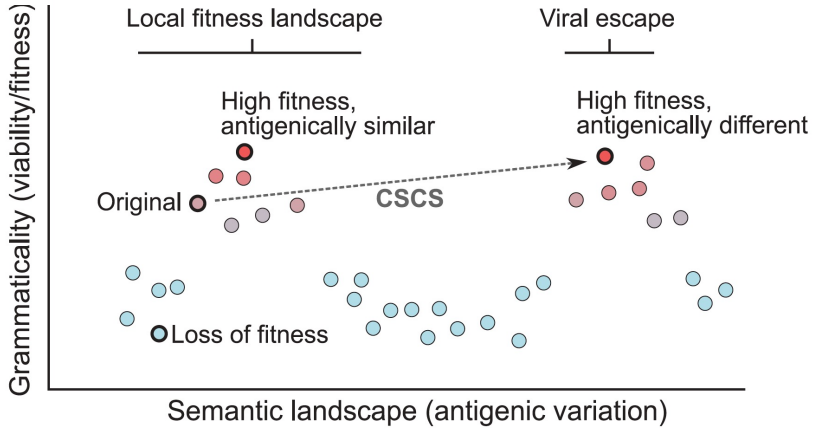
Ultracheap DNA sequencing could help end the pandemic — and reshape the future of health care.

By Jon Gertner

nyt mag 28 March 2021

The full genome of the B.1.1.7 coronavirus variant.

Learning the language of viral evolution and escape

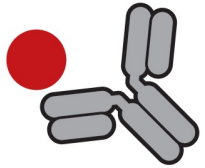


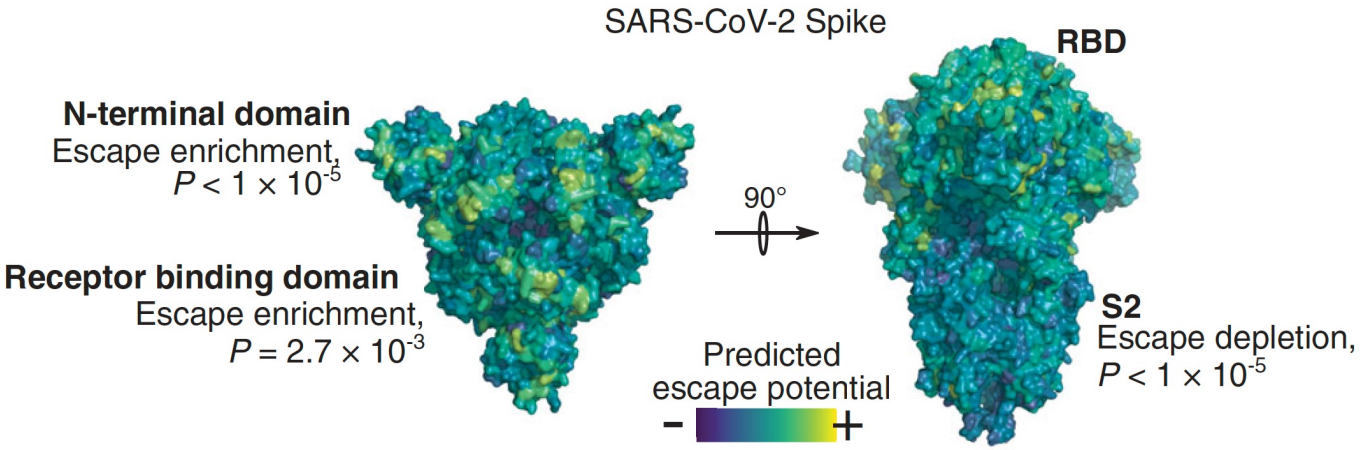

```
       Original: australian dead   in bali
Semantically closest: aussie     dead   in bali
     CSCS change: australian ballet in bali

       Original: blast off     of apollo 8
Semantically closest: blast off     of apollo 13
     CSCS change: blast victims of apollo 8
```

Viral protein sequences

VLSAKAA	Fitness	Immune semantics
VLSIKAA	High	Similar
VLSYKAA	High	Distant ← Immune escape
VLSKKAA	Low	

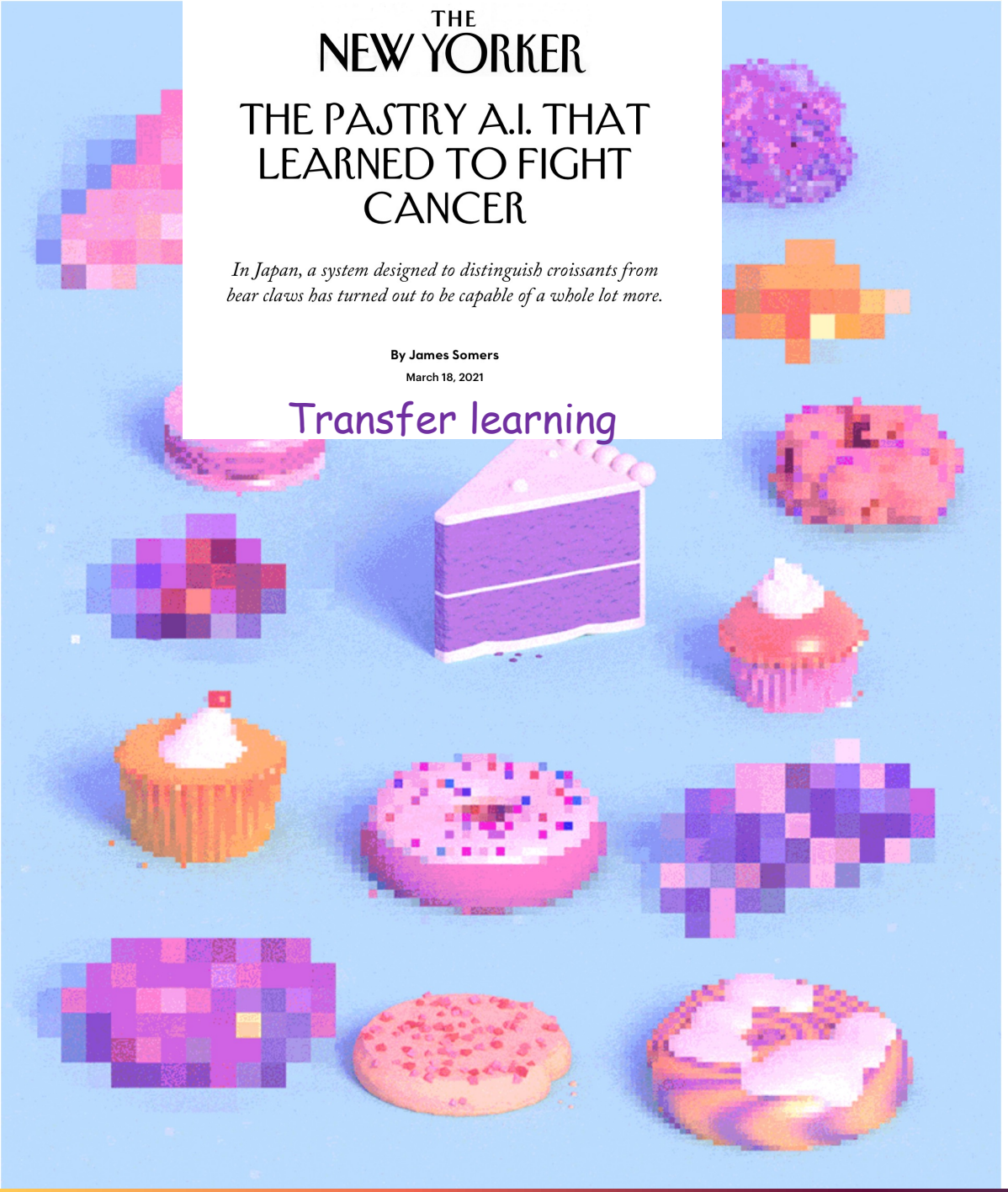
Predict immune escape



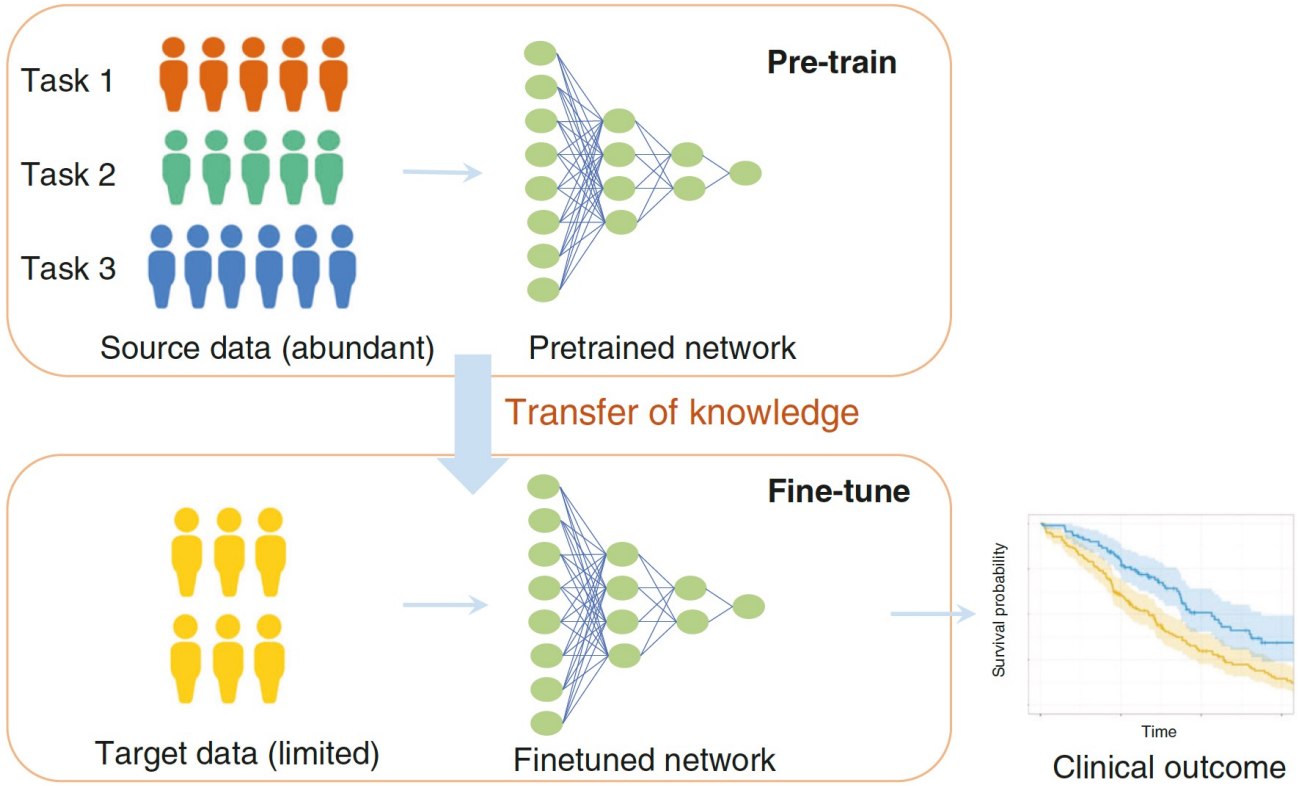


Science

14 January 2021



Meta-learning for genomics to reduce the amount of data for models

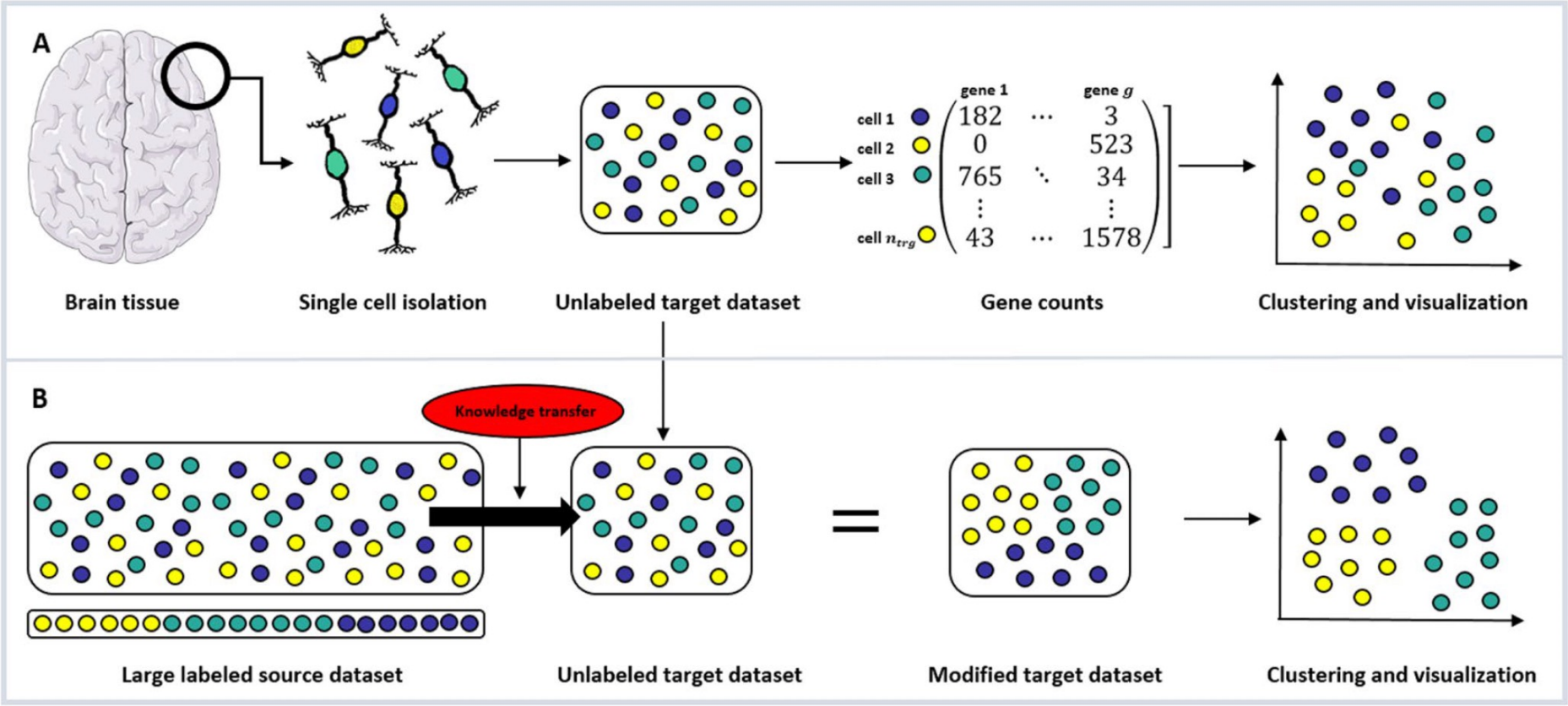


BJC British Journal of Cancer

29 March 2021

SCIENTIFIC REPORTS

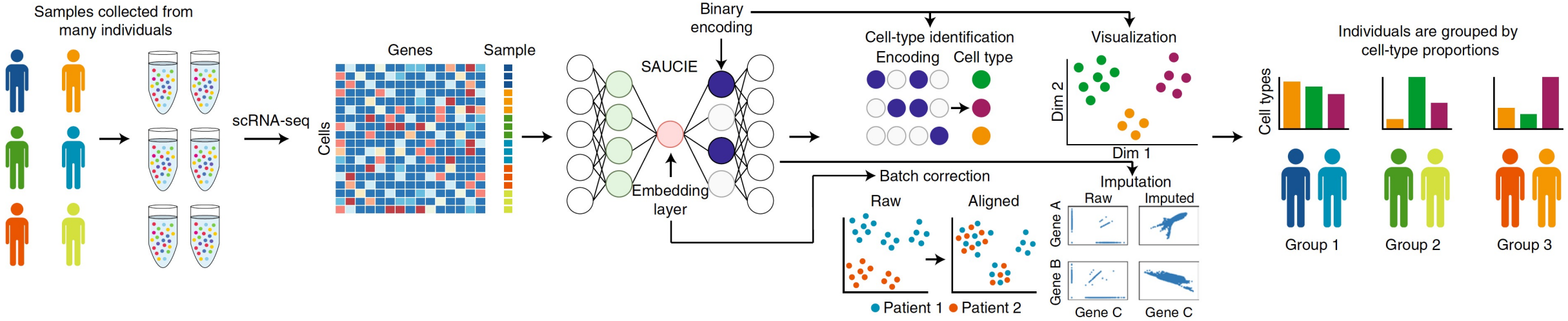
Using transfer learning from prior reference knowledge to improve the clustering of single-cell RNA-Seq data

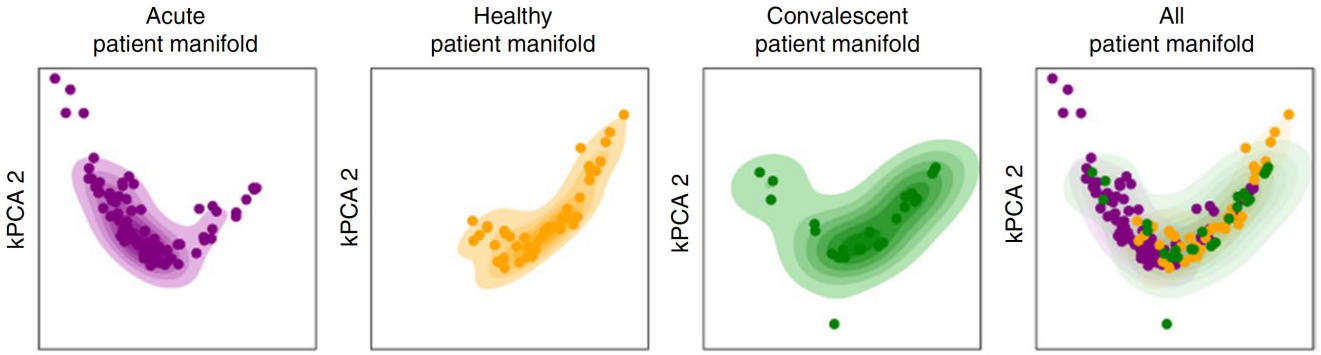


Mieth B et al, December 2019

Exploring single-cell data with deep multitasking neural networks

How to analyze 11 million T cells from 180 samples, 40 patients
(and control for batch effects, different sample preps)



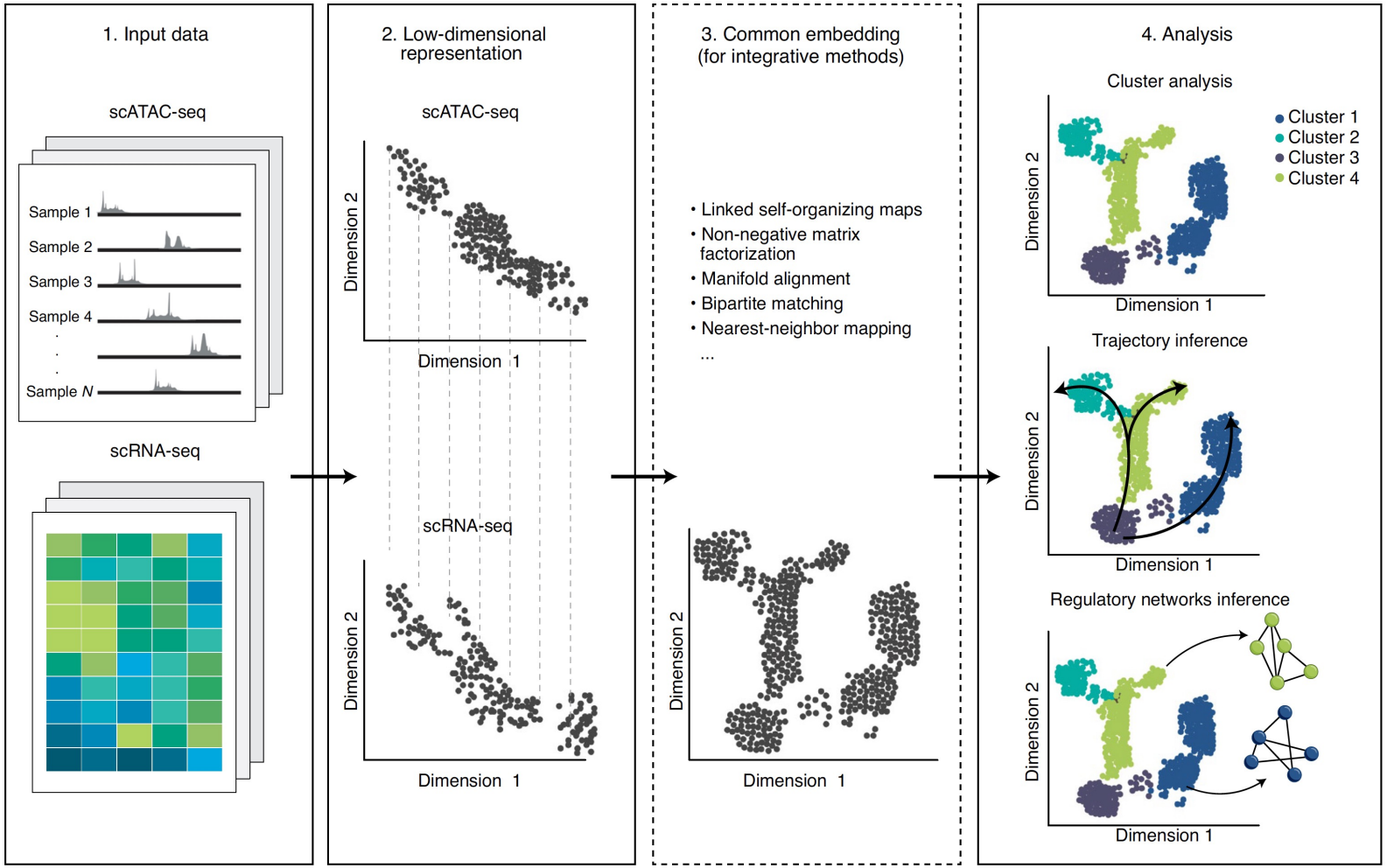


7 October 2019

nature methods

nature computational science

Machine learning for deciphering cell heterogeneity and gene regulation



"Opportunities to understand epigenetic regulation"

Scherer M, March 2021

An epic clash of cultures in ancient Mesoamerica p. 968

Music is another language pp. 974 & 1043

A primordial body in the Kuiper Belt pp. 980 & 998–1000

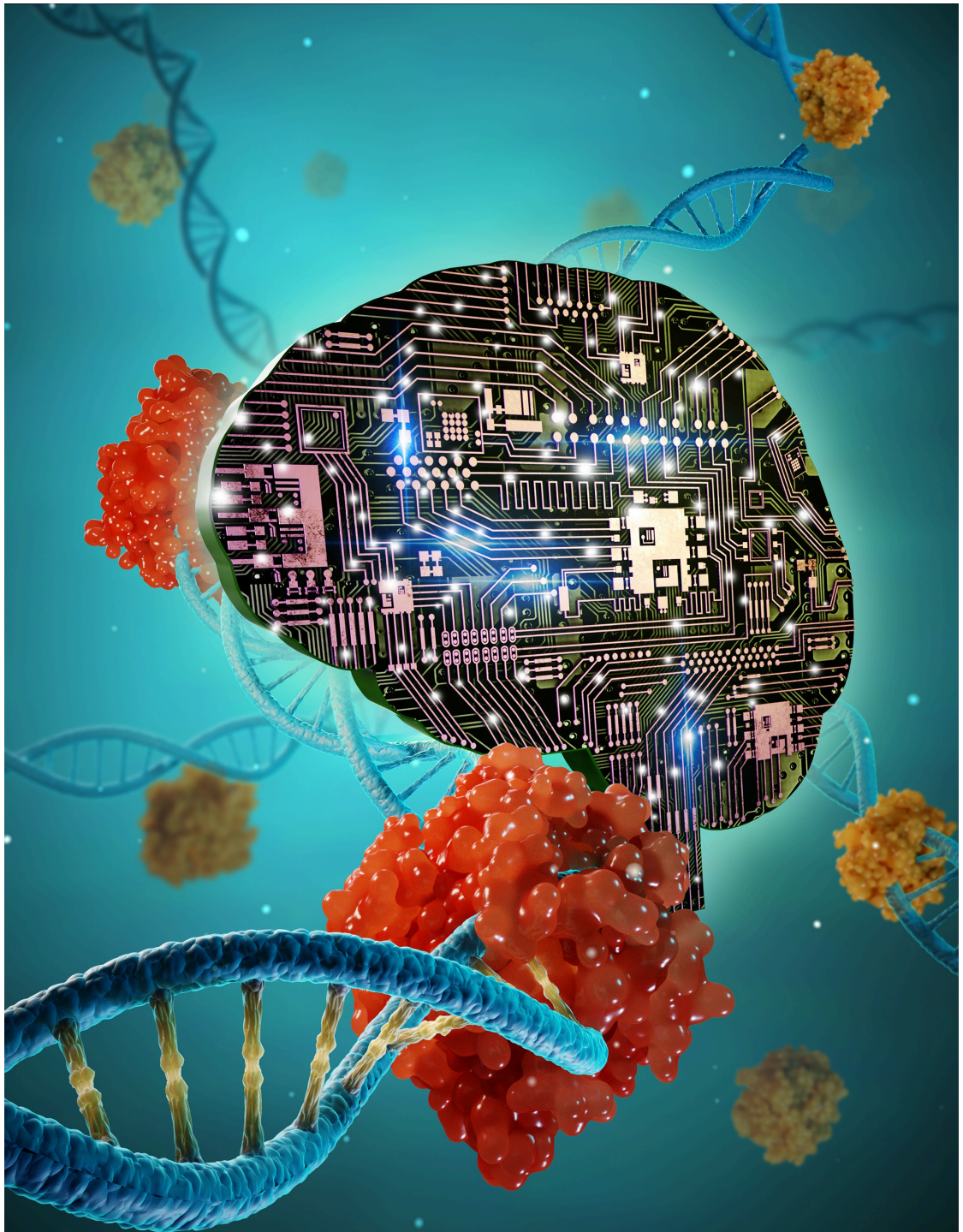
Science

$15
28 FEBRUARY 2020
sciencemag.org
AAAS

HUMAN CRISPR

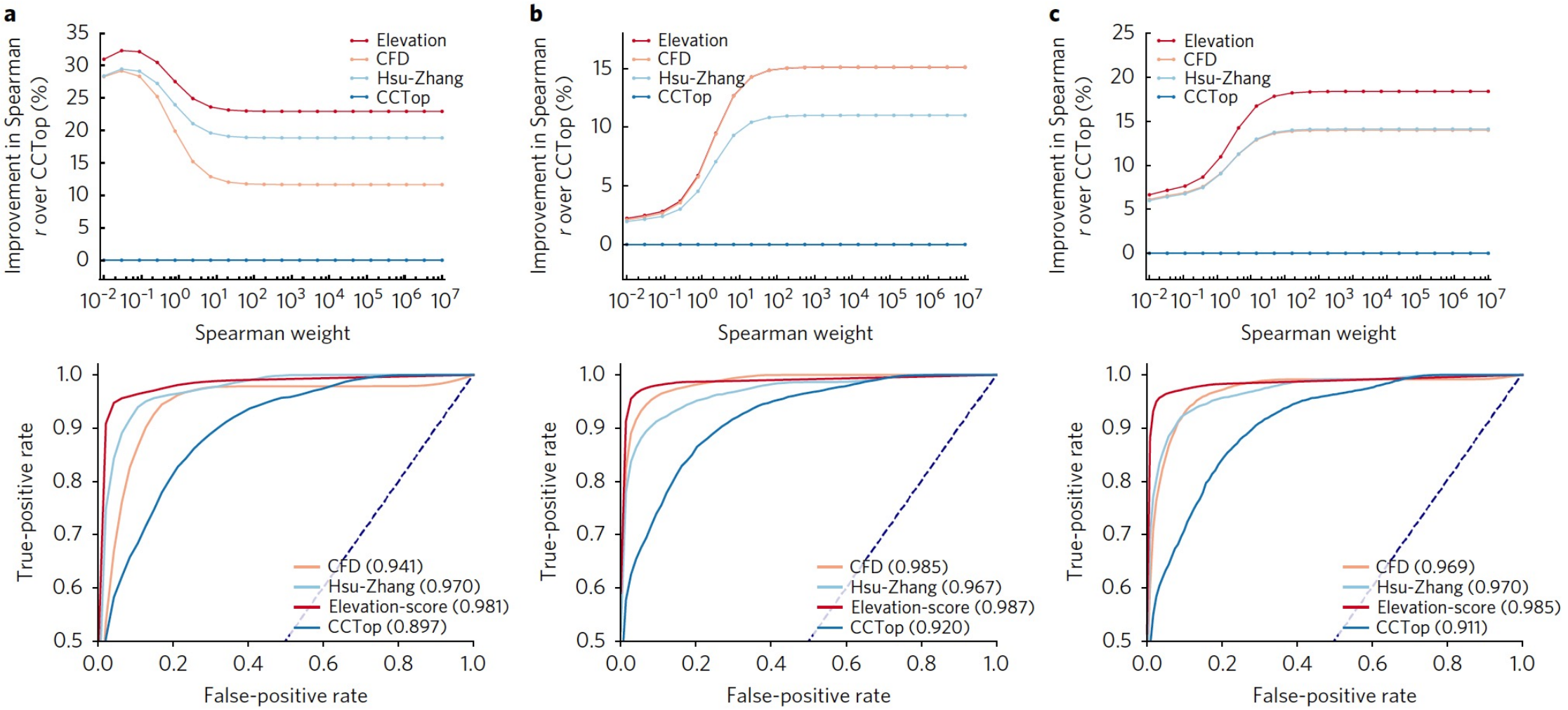
Gene editing meets cancer immunotherapy

pp. 976 & 1001



nature biomedical engineering

Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs



Listergarten J et al, Microsoft AI, 2018

Predictable an
CRISPR editing

Max W. Shen[1,2,12], Mandana Arbab[3,4,5,12], Jona
Christopher A. Cassa[8,10], David R. Liu[3,4,5]*, D

**Following Cas9 cleavage, DNA repair wit
impractical beyond gene disruption. Here
repair to a predicted genotype, enabling c
of 2,000 Cas9 guide RNAs paired with DI
genotypes and frequencies of 1- to 60-ba
five human and mouse cell lines. inDelp
'precise-50', yielding a single genotype
experimentally confirmed precise-50 inse
in primary patient-derived fibroblasts of
and Menkes disease. This study establishe**

Clustered regularly interspaced short palindromic
Cas9 has revolutionized genome editing, providing
tools and promising agents for the potential treatı
eases[1–3]. The DNA-targeting capabilities of Cas9 h
by the development of guide RNA (gRNA) desig
elling of factors leading to off-target DNA cleavag
Cas9 sequence fidelity by modifications to the n
and the evolution or engineering of Cas9 variar
PAM sequences[5]. Similarly, control over the proc
genome editing has been advanced by the develo
ing to achieve precise and efficient single-nucle
and the improvement of template-directed homol
(HDR) of double-stranded breaks[8]. Despite these
editing does not mediate insertions or deletions,
by low efficiency, particularly in non-dividing cell
by-products. As many human genetic variants asso
arise from insertions and deletions[9,10], methods to
insertions and deletions to alleviate pathogenic mu
able manner with a major single-genotype outcome
field of genome editing.

Non-homologous end joining (NHEJ) and
mediated end joining (MMEJ) processes are major
in the repair of Cas9-mediated double-stranded br
in highly heterogeneous repair outcomes comp
repair genotypes. Although end-joining repair of C
ble-stranded DNA breaks has been harnessed to fa
DNA templates[11,12] or deletion of intervening seq
cleavage sites[5], NHEJ and MMEJ are not generally c
precision genome editing applications. Previous v
the heterogeneous distribution of Cas9-mediated e

BRIEF COMMUNICATI
https://doi.org/10.1038/s41587-019-0203-2

Large dataset enabl
CRISPR–Cas9 editir

Ryan T. Leenay[1,17], Amirali Aghazadeh
Ryan Apathy[4], Eric Shifrut[4], Judd F. Hul
Hera Canaj[1], Manuel D. Leonetti[1], Ale
James Zou[1,2,15]*

**Understanding of repair outcomes after Cas9-i
cleavage is still limited, especially in primary hur
sequence repair outcomes at 1,656 on-target ger
primary human T cells and use these data to tra
learning model, which we have called CRISPR Rep
(SPROUT). SPROUT accurately predicts the leng
ity and sequence of nucleotide insertions and d
will facilitate design of SpCas9 guide RNAs in th
important primary human cells.**

Primary T cells are a promising cell type for therap
editing, as they can be engineered efficiently ex vivo a
transferred to patients[1]. However, detailed informat
genomic outcomes of Cas9-dependent editing in pr
cells is lacking. Here we systematically characterize
pyogenes Cas9 (SpCas9) repair outcomes in primary T
healthy blood donors (Supplementary Fig. 1).

Targeted sequencing was applied to 1,656 un
locations within 559 genes in primary CD4+ T cells
(gRNAs) were combined with SpCas9 to assembl
protein complexes (RNPs) and electroporated into T
was isolated from cells after 6 d of recovery and exp
180- to 260-base-pair region around each site was I
and sequenced (Fig. 1). We quantified the distribu
outcomes at each target site from the generated an
using CrispRVariants[4] (Fig. 1). In total, 31% of reads c
tions centered around the cut site with an average del
13 base pairs. We also found that 20% of the reads ha
the cut site, and 95% of these insertions were exactl
(Supplementary Fig. 2). Only 0.008% of reads cont
insertion and a deletion.

There were an average of 98 discrete repair outc
get site that were observed at a frequency greater th
reads, and different sites were highly variable in the p
length distribution of insertions and deletions. The re
from each target site were similar between donors, but

Predicti
specific
attentio
based g

Qiao Liu[1], Di H
1 Department of C
United States of A
of New York, New
The Graduate Cer
4 Helen and Robe
Institute, Weill Cor
* lei.xie@hunter.c

Abstract

CRISPR-Cas is
gene therapy. S
many safety co
Cas treatment,
these issues, it
Existing single-
the local inform
ence in the cell
specific informa
ing models, wh
network with ce
sion profile, for
ity. In benchma
Furthermore, w
cell-specific pos
and safe CRISI
may bolster dev
spectrum of bic
safe CRISPR-b

Deep learning improves prediction of CRISPR–Cpf1 guide RNA activity

Hui Kwon Kim[1,2,9], Seonwoo Min[3,9], Myungjae Song[1,4], Soobin Jung[1,2], Jae Woo Choi[1,5], Younggwang Kim[1,2], Sangeun Lee[1,2], Sungroh Yoon[3,6] & Hyongbum (Henry) Kim[1,2,5,7,8]

We present two algorithms to predict the activity of AsCpf1 guide RNAs. Indel frequencies for 15,000 target sequences were used in a deep-learning framework based on a convolutional neural network to train Seq-deepCpf1. We then incorporated chromatin accessibility information to create the better-performing DeepCpf1 algorithm for cell lines for which such information is available and show that both algorithms outperform previous machine learning algorithms on our own and published data sets.

Cpf1 (from *Prevotella* and *Francisella* 1) is an effector endonuclease protein of the class 2 CRISPR–Cas system and allows genome editing in various species and cell types, including human cells[1–7]. We previously reported a high-throughput method for evaluating Cpf1 activity in human cells, which allowed the development of a prototype computational algorithm for predicting the activity of AsCpf1 (Cpf1 from *Acidaminococcus sp. BV3L6*) based only on target sequence composition[8]. However, our initial program was based on conventional (non-neural network) machine learning trained on a medium-scale (1,251 target sequences) data set of Cpf1 activities. Here we developed programs with significantly improved accuracy for predicting AsCpf1 activity at endogenous target sites.

We first obtained large-scale data sets of AsCpf1 activity at 16,292 (experiment A) and 2,963 (experiment B) lentivirally integrated target sequences using our high-throughput method and 20-nt guide sequences[8] in HEK293T cells. The high-throughput experiments A and B led to the generation of data sets HT 1 and HT 2, respectively, which consist of target sequence compositions and corresponding indel frequencies (**Supplementary Tables 1** and **2**). Data set HT 1 was split into data sets HT 1-1 (n = 15,000) and HT 1-2 (n = 1,292) by random sampling.

To build Seq-deepCpf1, a deep-learning-based regression model that predicts AsCpf1 activity based on target sequence composition, we used an end-to-end deep-learning framework based on a convolutional neural network (**Fig. 1a** and **Supplementary Fig. 1**). We performed nested cross-validation with data set HT 1-1 to evaluate the generalization performance of model selection and training of Seq-deepCpf1 (**Supplementary Figs. 2** and **3** and **Supplementary Table 3**). We found that 34 bp was adequate as the input target sequence (**Supplementary Fig. 4**).

nature biotechnology

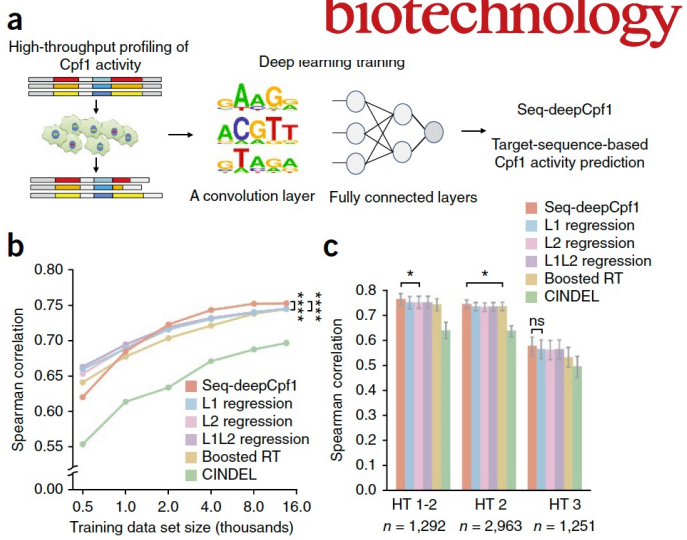
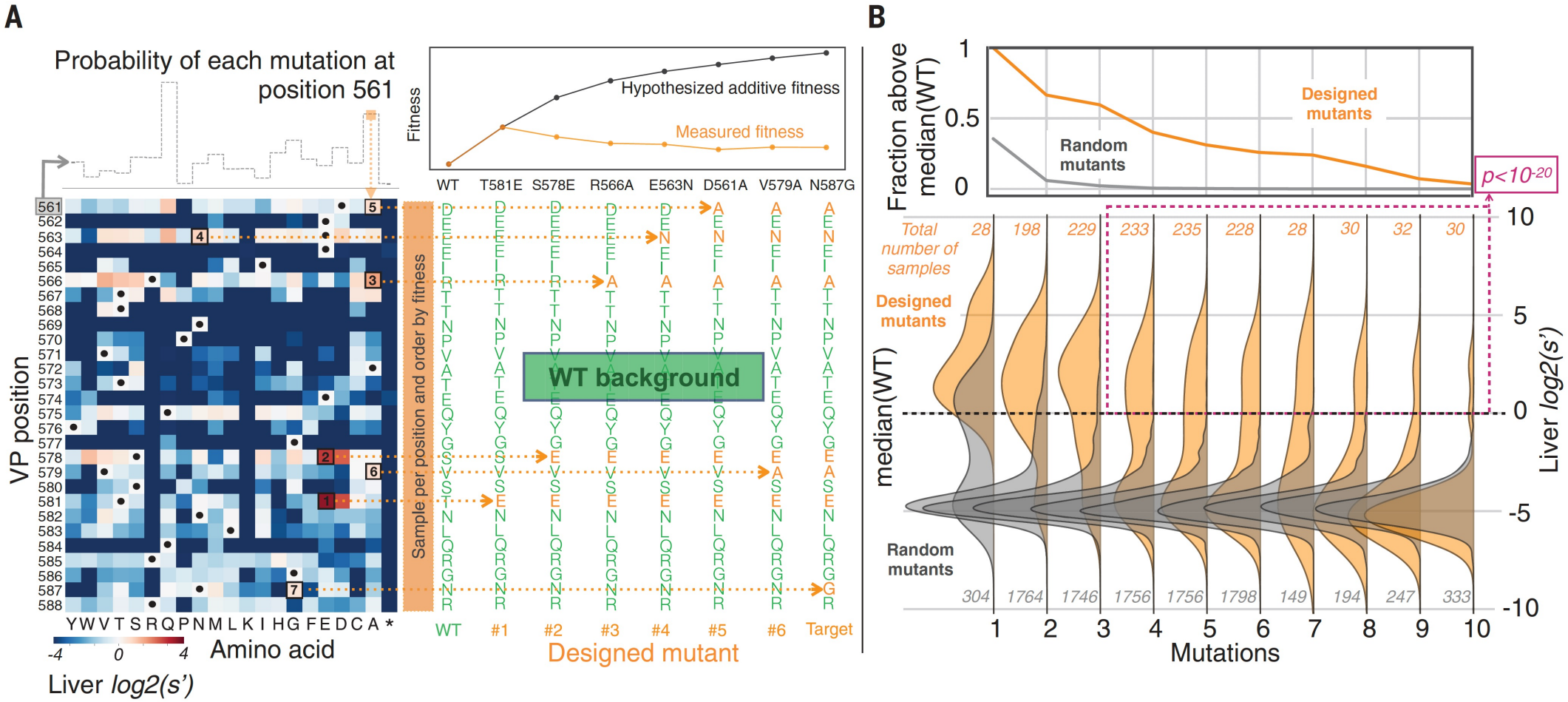


Figure 1 Deep learning outperforms conventional machine learning for the task of predicting Cpf1 activity based on the target sequence composition. (**a**) Schematic representation of deep learning for the target-sequence-dependent Cpf1 activity prediction. (**b**) Nested cross-validation of Cpf1 activity prediction models trained on different size data sets. Each point represents the average result of ten outer folds. The Spearman correlation coefficients between experimentally obtained indel frequencies and predicted scores from Seq-deepCpf1 and other conventional machine learning approaches are plotted for six different training data set sizes. For the sake of clarity, results from statistical significance testing are shown only between the best model and the two next-best models (Seq-deepCpf1 vs. L1L2 regression, $^{****}P = 6.5 \times 10^{-6}$; Seq-deepCpf1 vs. L2 regression, $^{****}P = 5.5 \times 10^{-6}$; Steiger's test). The confidence intervals are described in **Supplementary Table 7**. (**c**) Performance comparison of prediction models. For three independent test data sets (HT 1-2, HT 2, HT 3), the Spearman correlation coefficients between measured indel frequencies and predicted Cpf1 activity scores are shown. For the sake of clarity, results from statistical significance testing are shown only for the pair of the best and the next-best models (left to right; $^{*}P = 0.015$, $^{*}P = 0.026$, and ns = not significant; Steiger's test). Error bars represent 95% confidence intervals, which are also described in **Supplementary Table 7**. Boosted RT, gradient-boosted regression trees.

Science

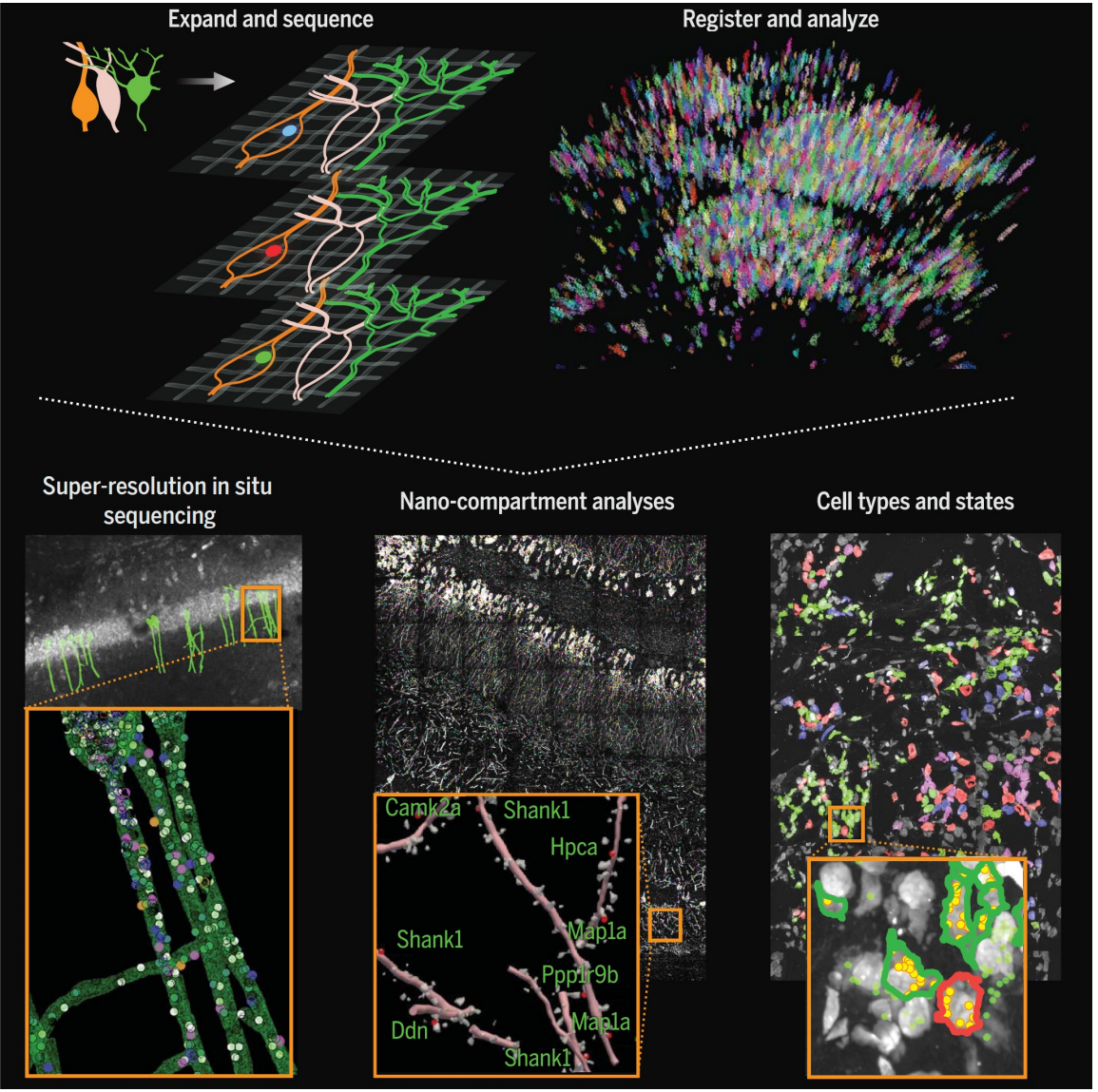
Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design

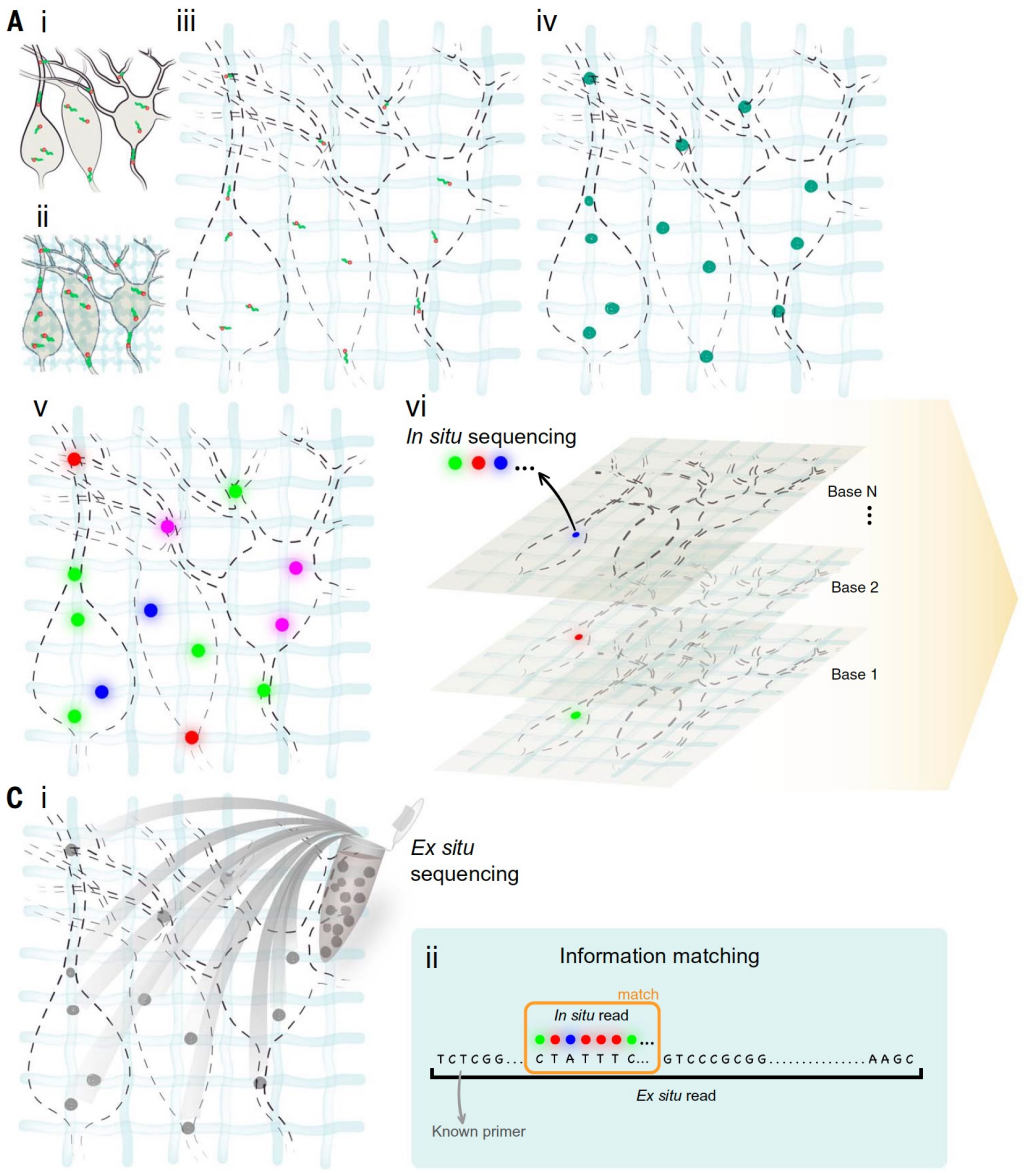


Ogden P et al, November 2019

IN SITU SEQUENCING

Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems

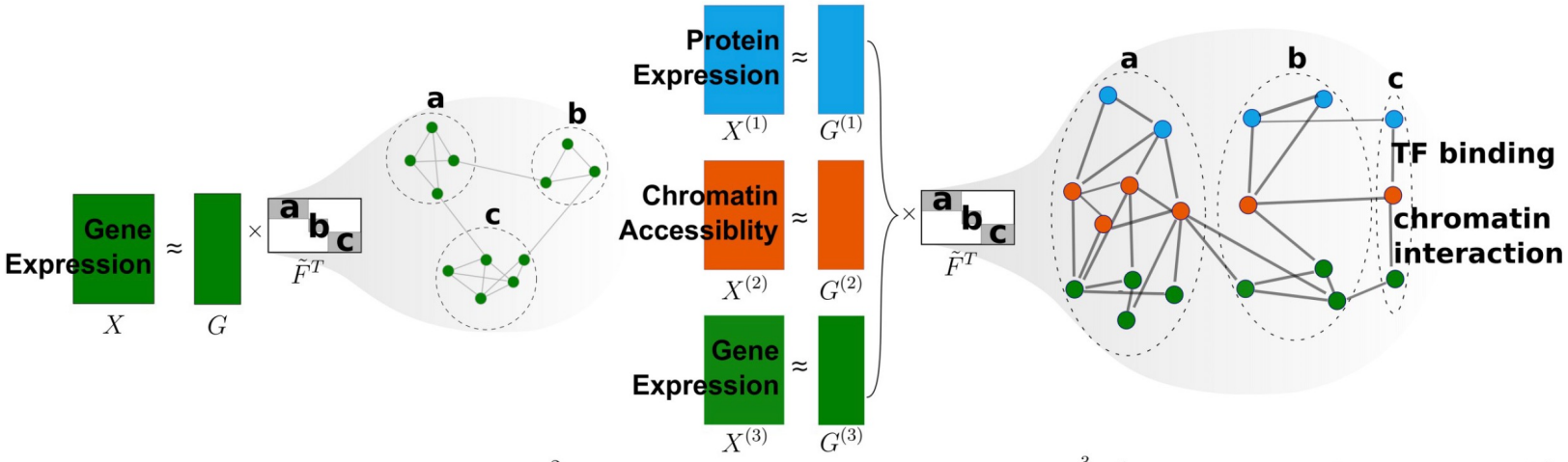




In situ sequencing of physically expanded specimens enables multiplexed mapping of RNAs at nanoscale, subcellular resolution throughout intact tissues.

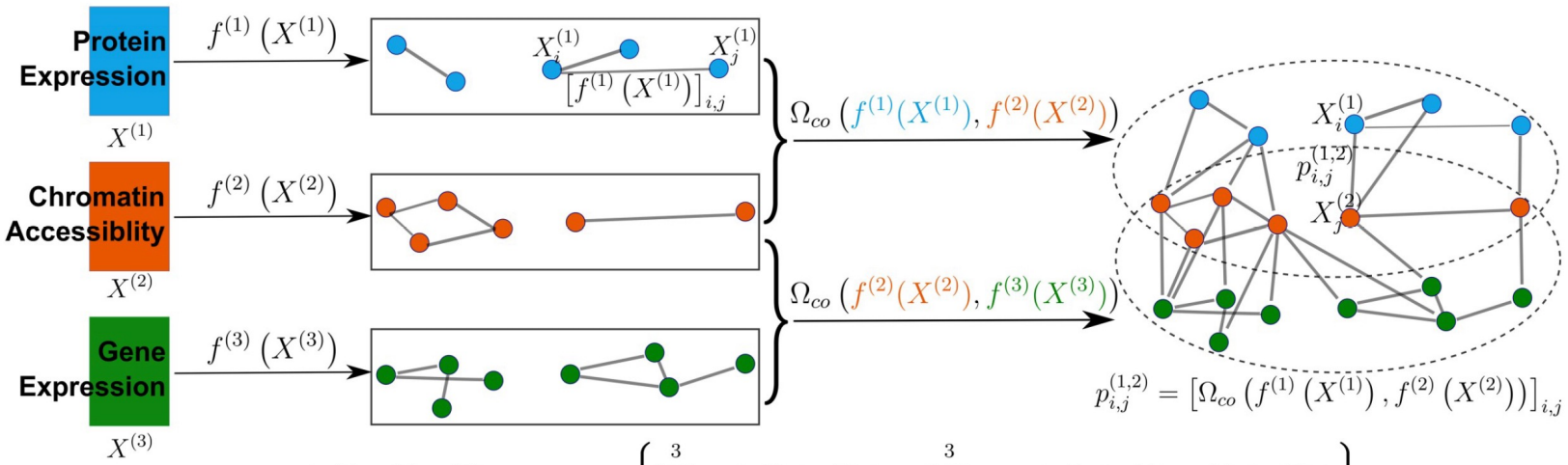
Science

Alon S et al, January 2021



$$\{G^*, \tilde{F}^*\} \in \arg\min_{G,F\geq 0} \left\| X - G\tilde{F}^T \right\|_F^2$$

$$\left(G^{(1)*}, G^{(2)*}, G^{(3)*}, \tilde{F}^*\right) \in \arg\min_{G^{(i)},F^{(i)},\tilde{F}\geq 0} \sum_{i=1}^{3} \left\{ \left\| X^{(i)} - G^{(i)} F^{(i)T} \right\|_F^2 + \lambda \left\| F^{(i)} - \tilde{F} \right\|_F^2 \right\}$$

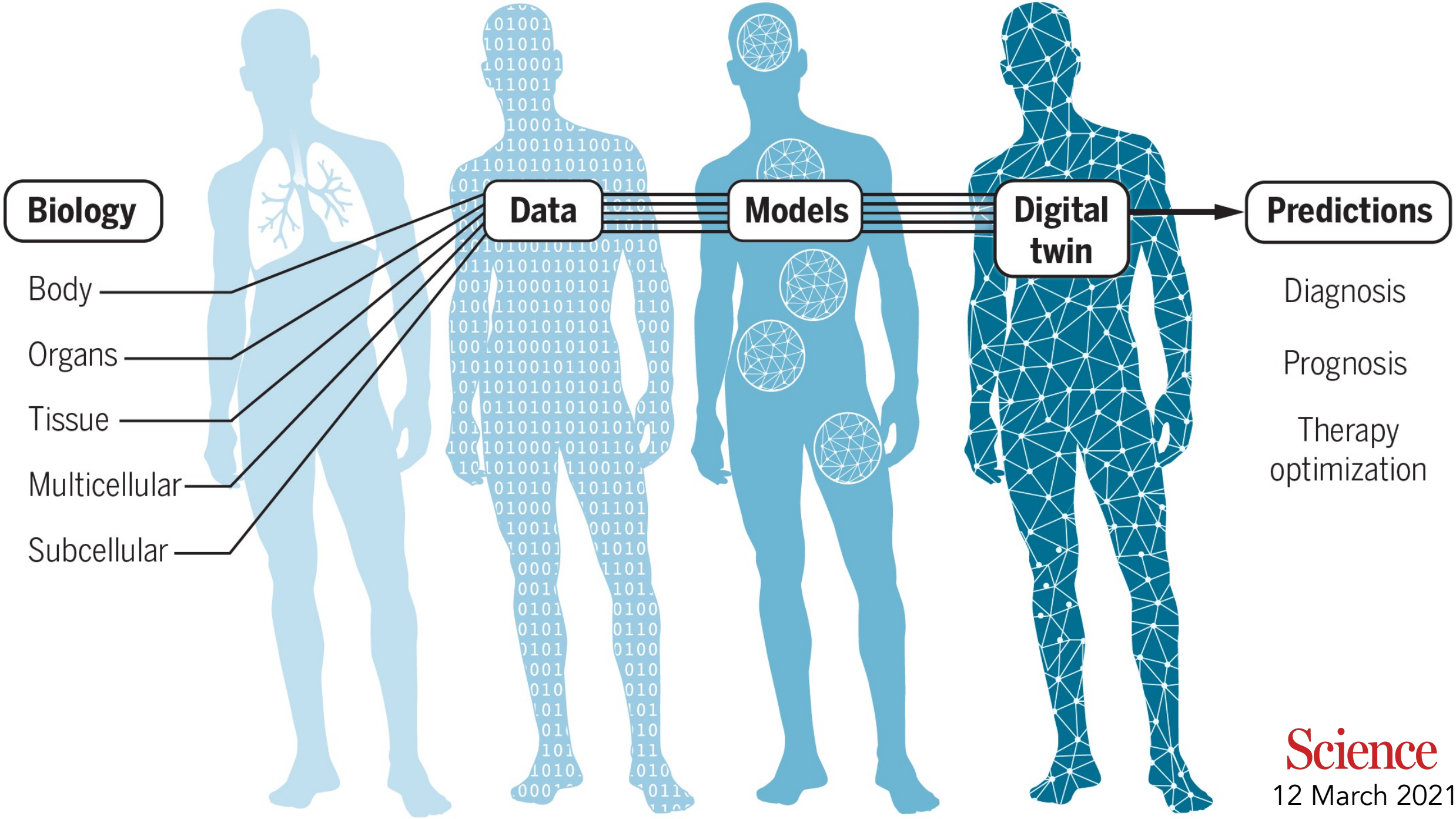


$$\left(f^{(1)}, f^{(2)}, f^{(3)}\right) \in \arg\min \left\{ \sum_{i=1}^{3} \Omega\left(f^{(i)}\left(X^{(i)}\right)\right) + \sum_{i>j=1}^{3} \Omega_{co}\left(f^{(i)}\left(X^{(i)}\right), f^{(j)}\left(X^{(j)}\right)\right) \right\}$$

PLOS COMPUTATIONAL BIOLOGY

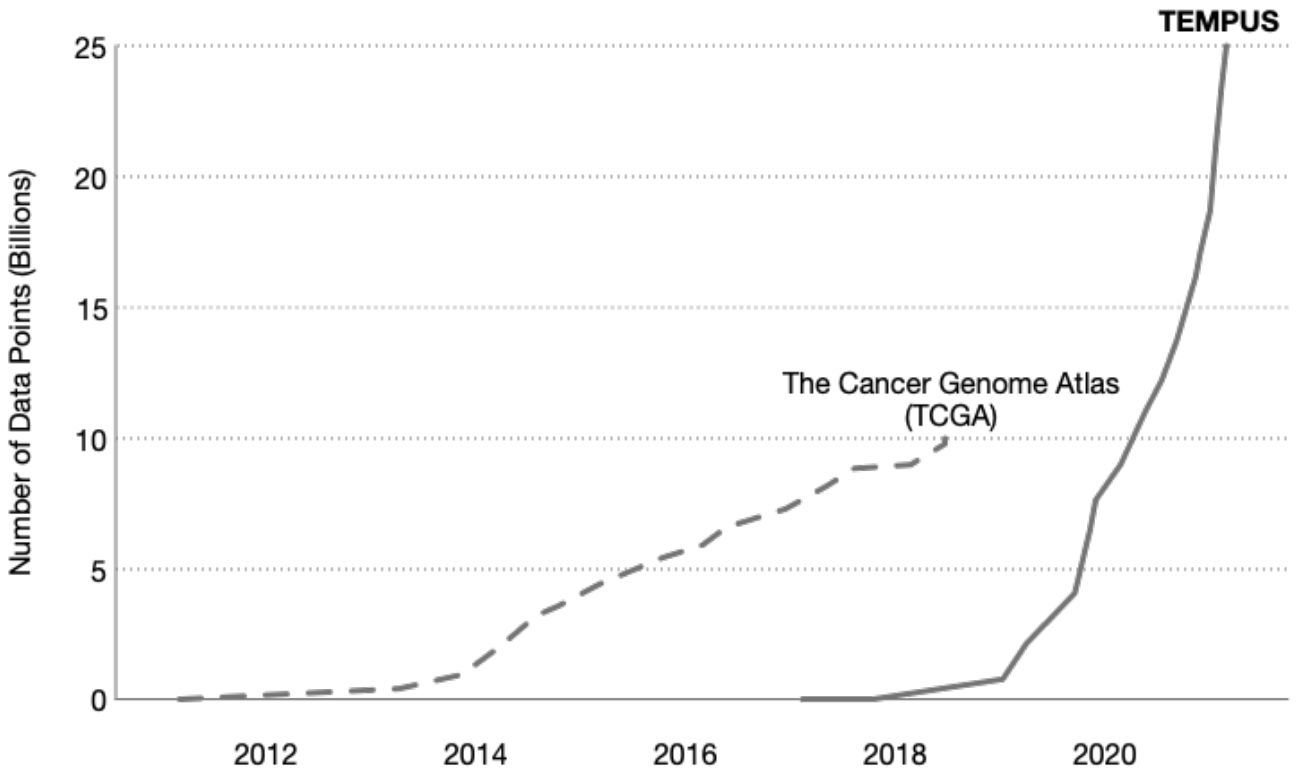
Nguyen N and Wang D, April 2020

Multiview learning for understanding functional multiomics



Science
12 March 2021

A Digital Twin Infrastructure for Cancer



~200,000 patients sequenced, multi-omics, EHR, treatment, outcomes, path and scans digitized

~400,000 patients EHR and molecular diagnostics

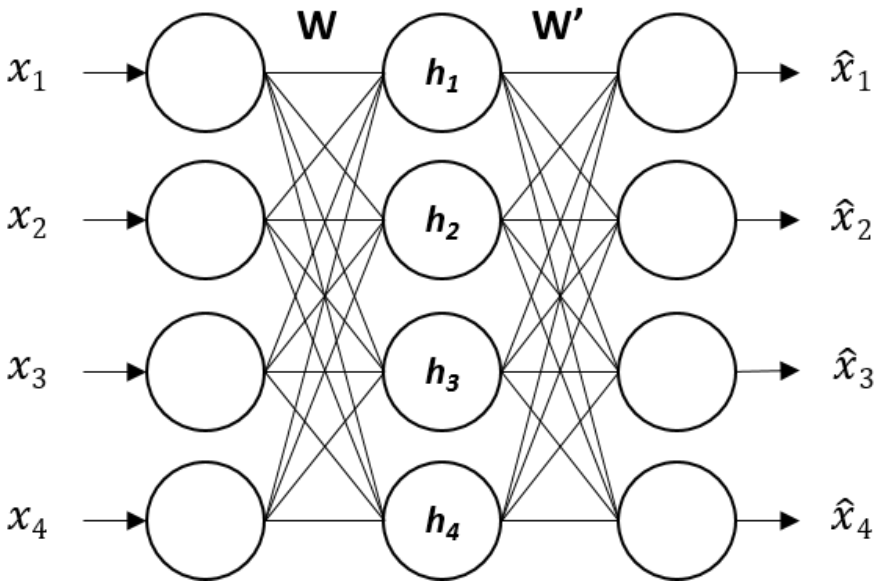
~800,000 imaging digitized

Scripps Research Imputation Project

Sparse Data
~4 million genetic

```
0 0 ? ? ? 0 0 1 1 ? ? 0 ? ? ? ?
0 0 ? ? ? 0 0 1 1 ? ? 0 ? ? ? ?
0 1 ? ? ? 1 0 1 1 ? ? 0 ? ? ? ?
0 0 ? ? ? 0 0 1 1 ? ? 1 ? ? ? ?
0 1 ? ? ? 1 0 1 1 ? ? 0 ? ? ? ?
0 1 ? ? ? 1 0 1 1 ? ? 0 ? ? ? ?
```

Affymetrix
and Axiom
arrays



Full Data
~80M genetic variants

```
0 0 1 0 0 0 0 1 1 1 1 0 1 1 0 1
0 0 1 0 0 0 0 1 1 1 1 0 1 1 0 1
0 1 0 1 0 1 0 1 1 1 1 0 1 1 0 1
0 0 1 0 0 0 0 1 1 0 0 1 0 1 1 0
0 1 0 1 0 1 0 1 1 1 1 0 1 1 0 1
0 1 0 1 1 1 0 1 1 1 1 0 1 1 0 1
```

Scripps Research
Science Changing Life

Scripps Research
Translational Institute

Scripps Research Imputation Project

9p21

Chromosome 22*

Whole genome

*~1,000 very complex regions

Raquel Dias PhD
KL2 Scholar

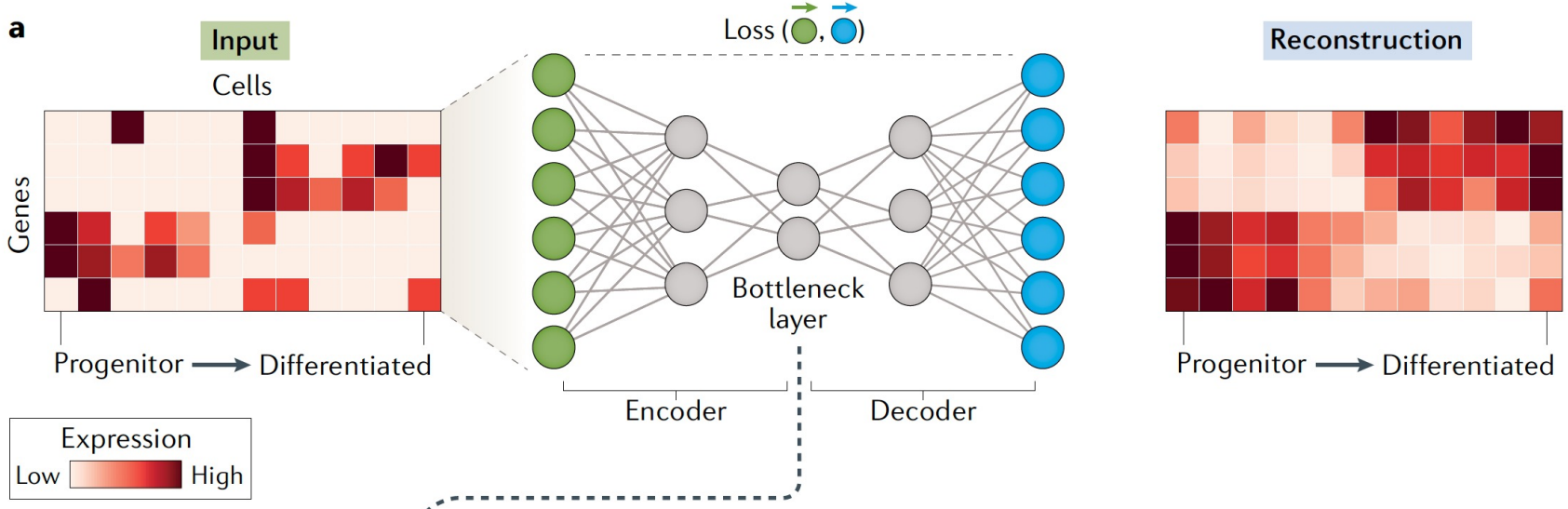
NVIDIA®

R01 Funded

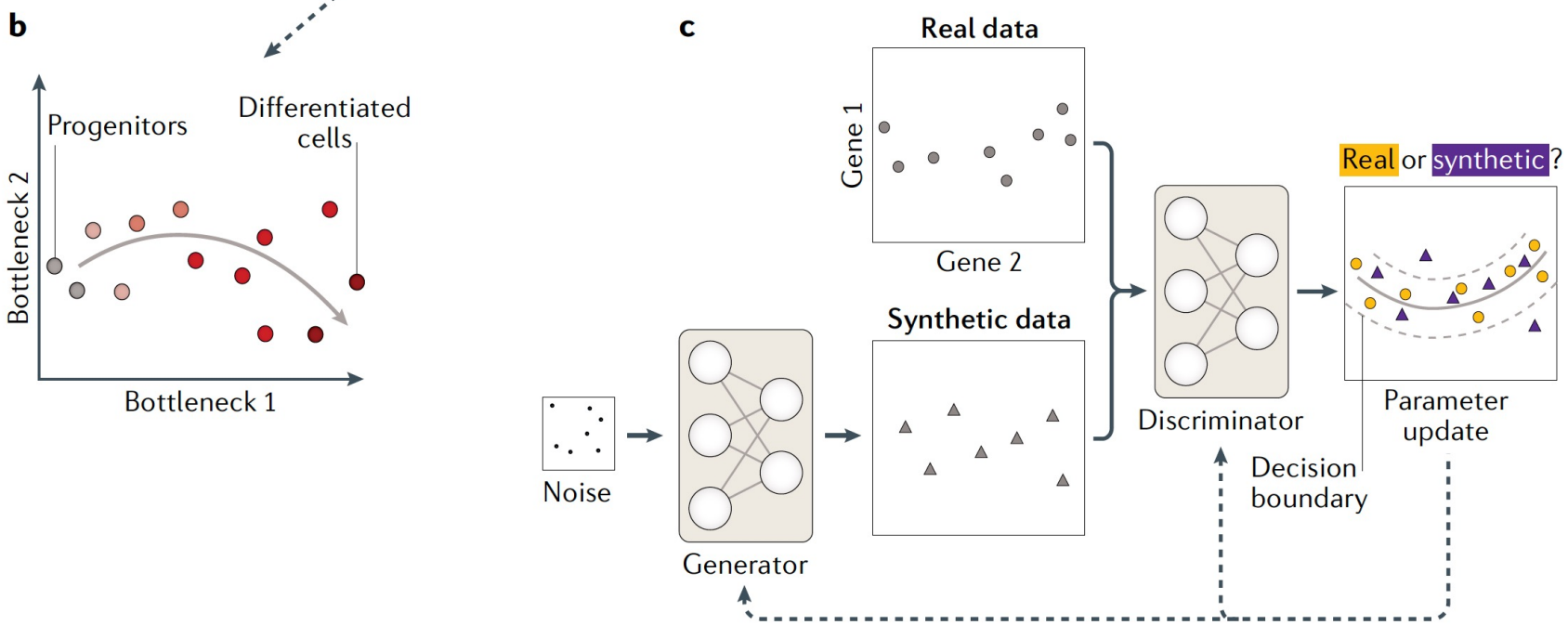
NIH National Human Genome Research Institute

Unsupervised learning

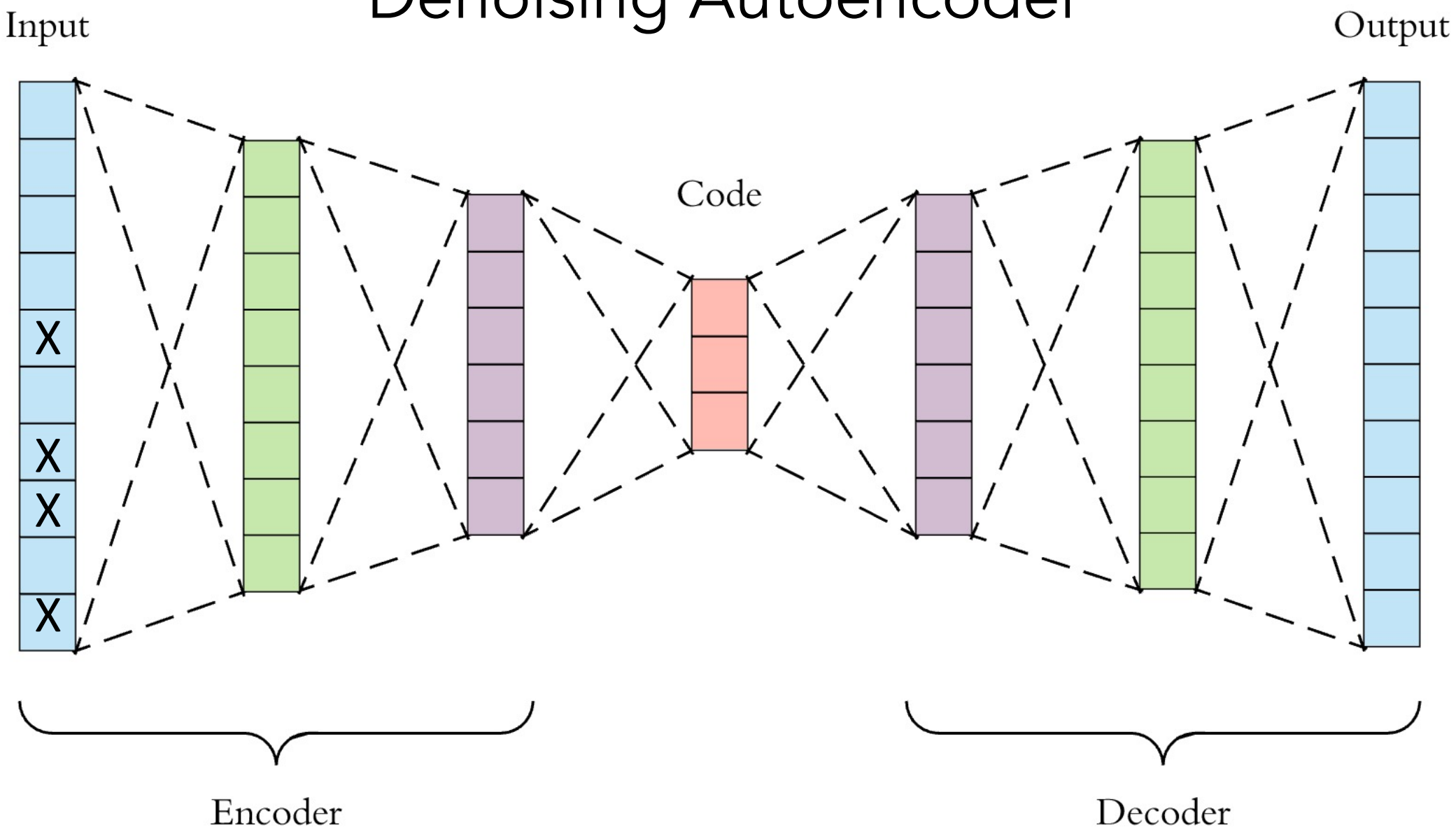
Autoencoder



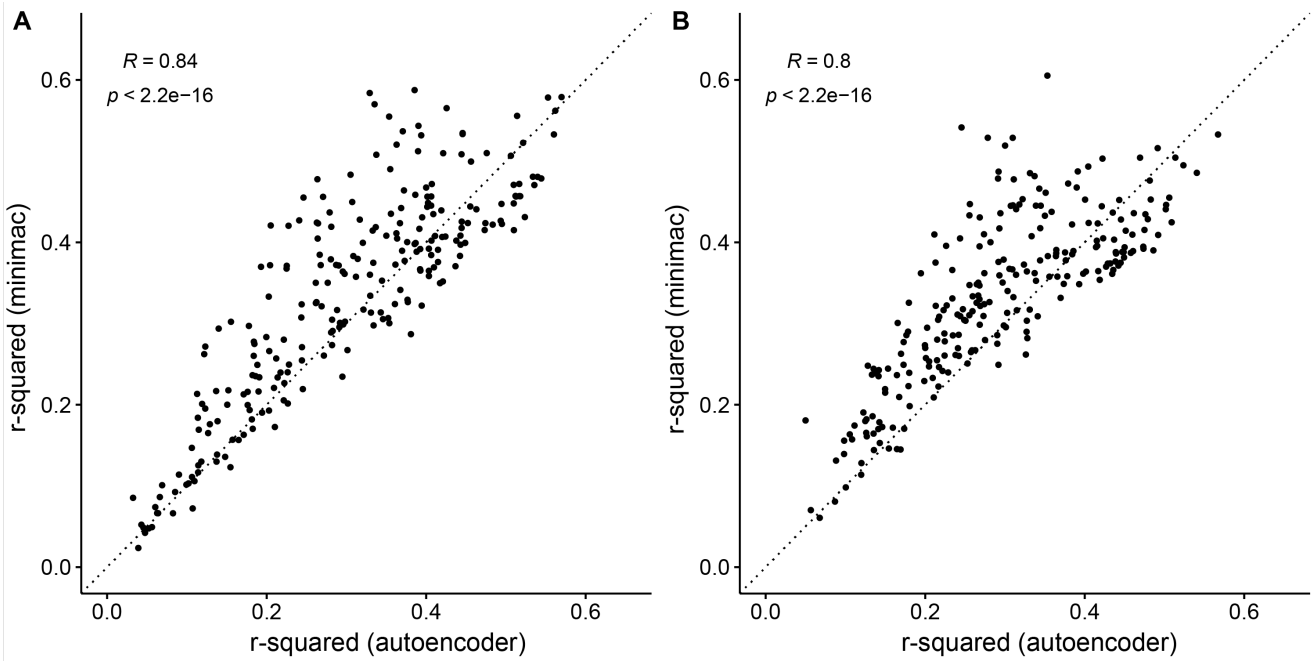
Generative Adversarial Networks (GANs)



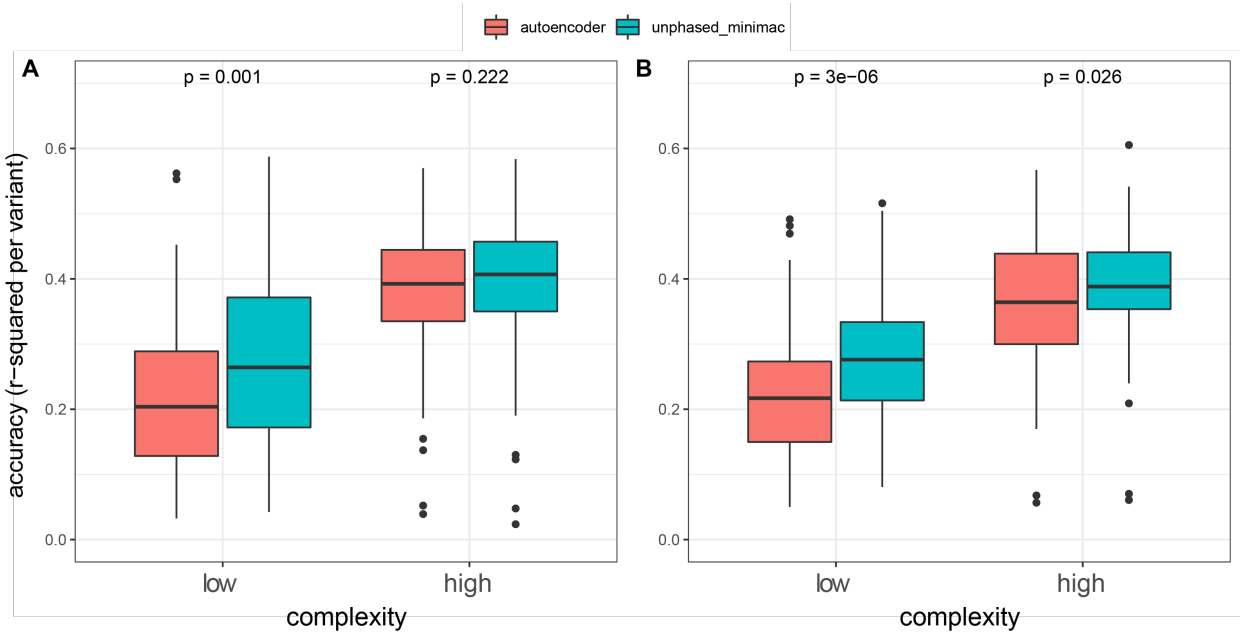
Denoising Autoencoder



Minimac (HMM) compared with an Autoencoder



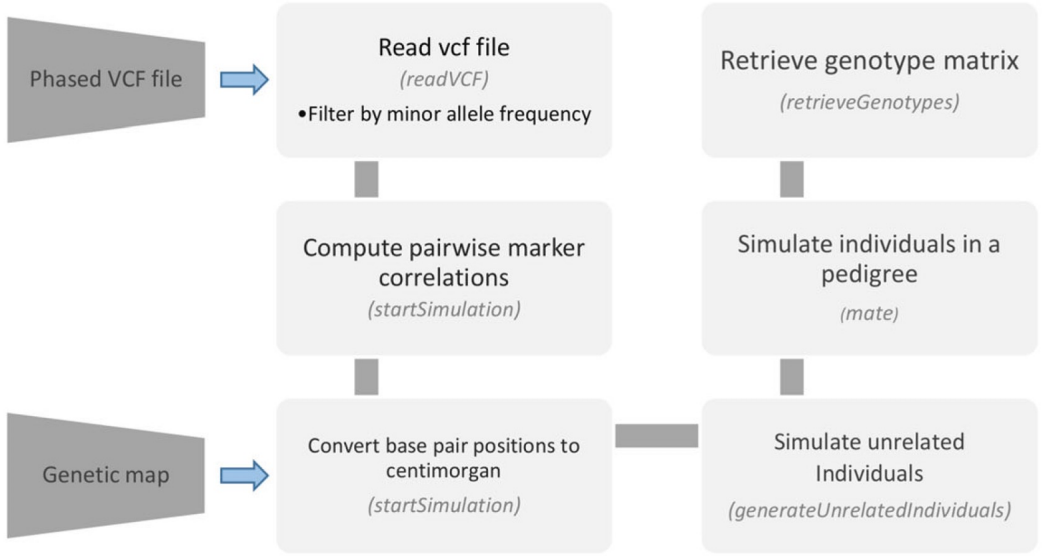
Chromosome 22
Each dot=LD block
250 regions



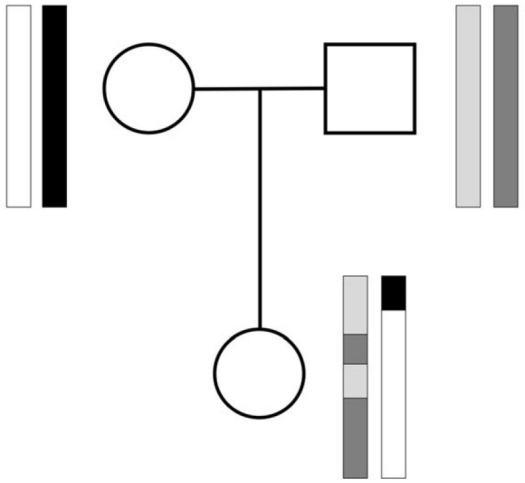
The autoencoder underperformed

A Lesson on Data Inputs

Synthetic Data: 30,000 "Virtual Babies"



(a) Overview of simulation workflow



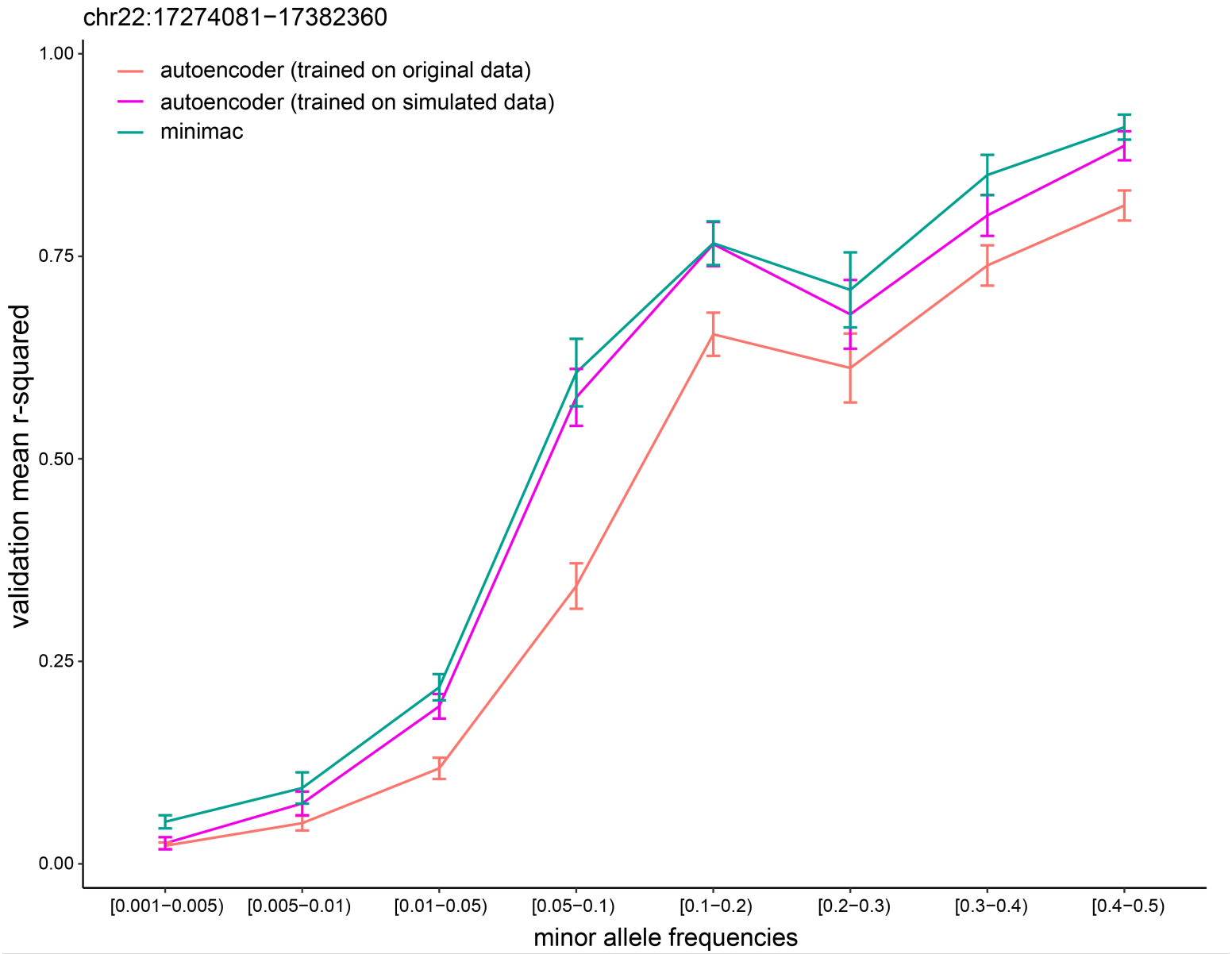
(b) Recombination process for the simulation of related individuals

Genomic data Augmentation
More admixed, doubled size of training data

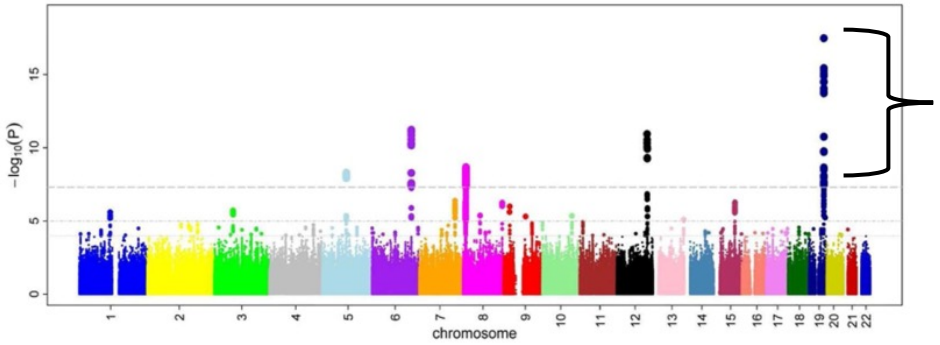
sim1000G: a user-friendly genetic variant simulator in *R* for unrelated individuals and family-based designs

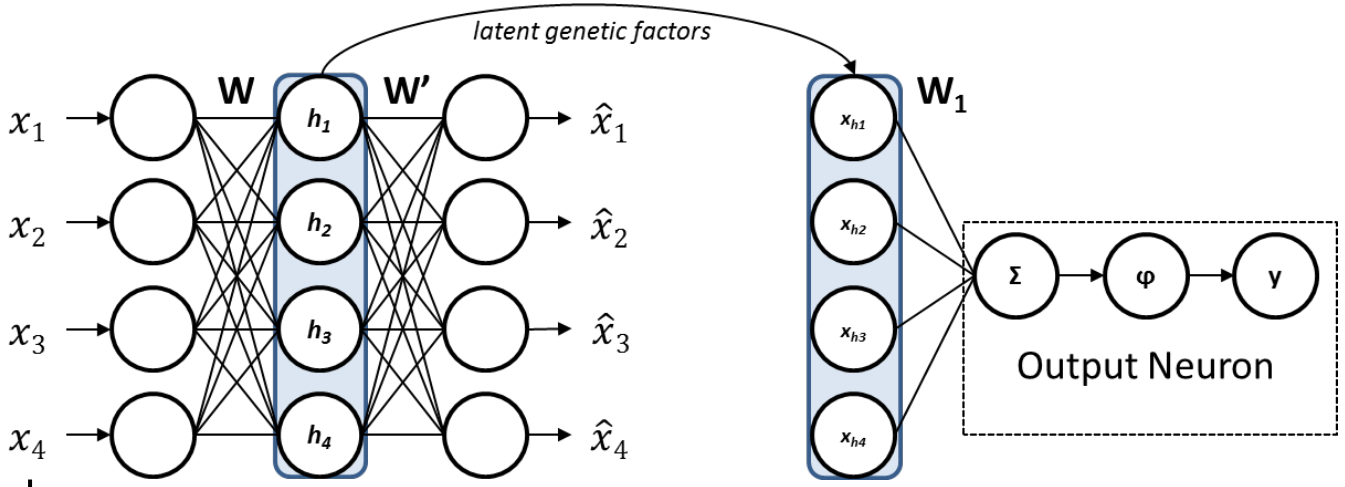
Dimitromanolakis *et al. BMC Bioinformatics* (2019) 20:26

Improving Performance with Genomic Data Augmentation



Deep Learning-Based Polygenic Risk Prediction



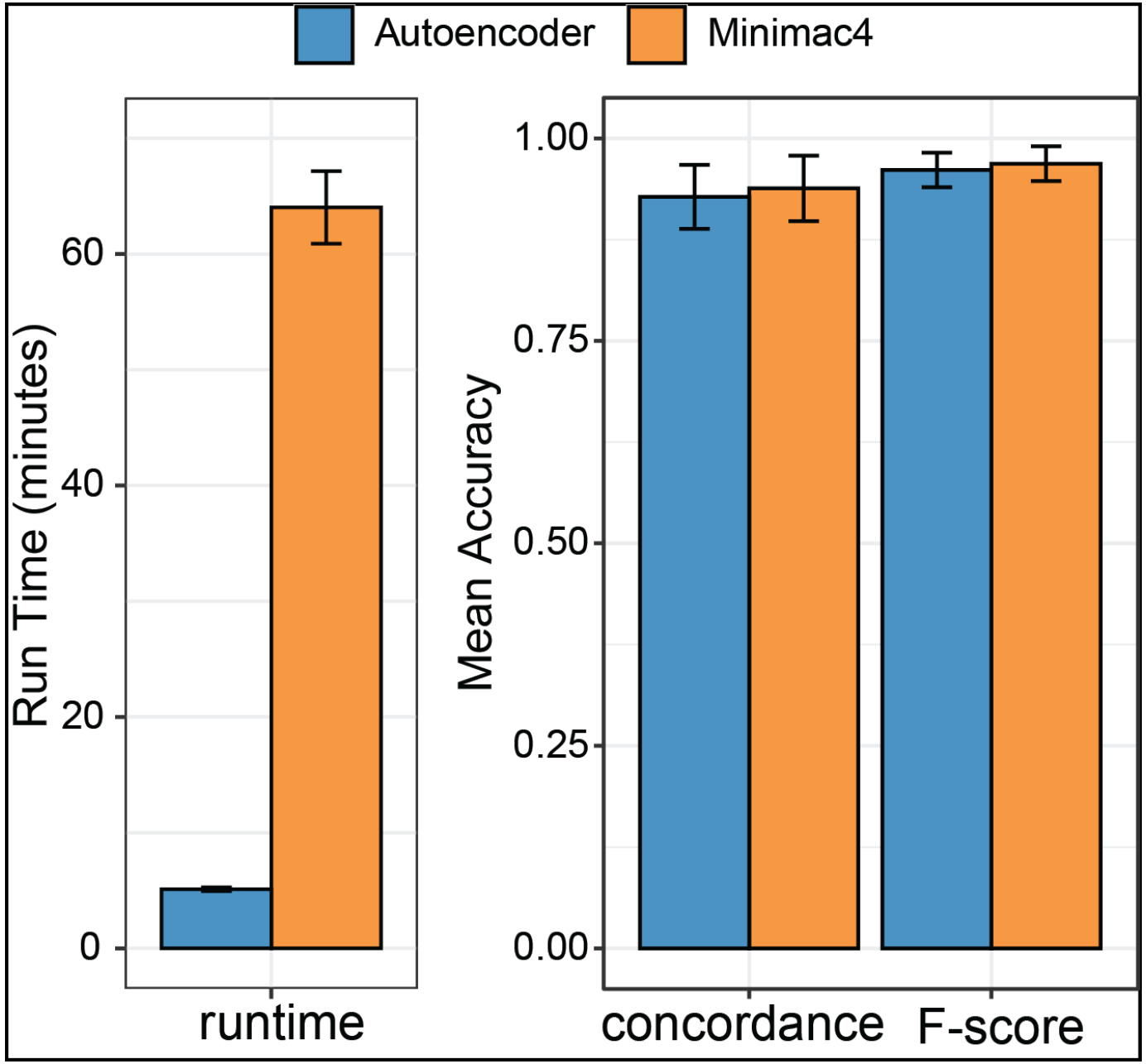


Genotypes

Hidden Encoding

Extend to:

- Deeper networks
- Inclusion of clinical risk factors
- Multi-task deep learning


Autoencoder
Minimac4
Run Time (minutes)
0
20
40
60
runtime
Mean Accuracy
0.00
0.25
0.50
0.75
1.00
concordance
F-score

A GPU Guzzler

4 models per GPU

Project for chromosome 22 with 18 GPUs
running in parallel

Each model takes 3-5 days to train

Extrapolation for Whole genome
~500,000 GPUs

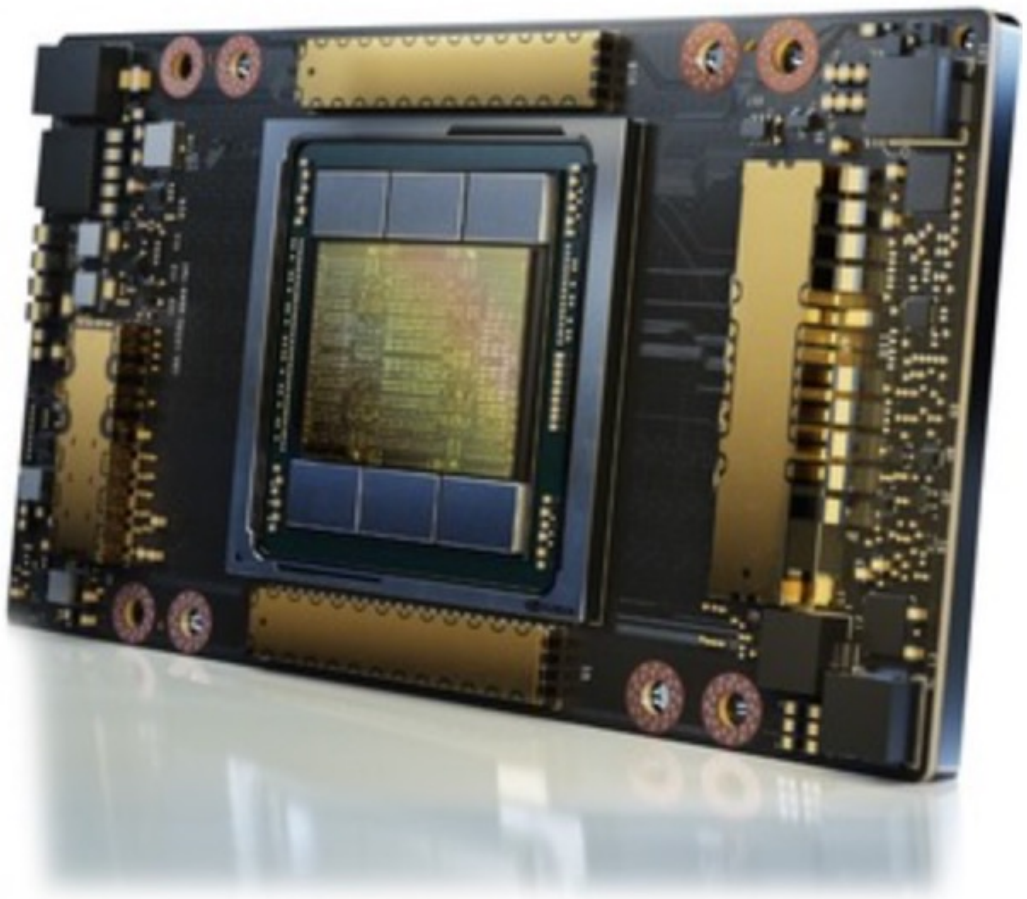

NVIDIA
A100

NVIDIA CORP

Nvidia VCX 900-21001-0000-000 A100 40GB CoWoS HBM2 PCIe 4.0 Passive Cooling

Mfg Part Number: **900-21001-0000-000** , Item #: 4162417

	NVIDIA A100 for NVLink	
Peak FP64	9.7 TF	
Peak FP64 Tensor Core	19.5 TF	
Peak FP32	19.5 TF	
Tensor Float 32 (TF32)	156 TF | 312 TF*	
Peak BFLOAT16 Tensor Core	312 TF | 624 TF*	
Peak FP16 Tensor Core	312 TF | 624 TF*	
Peak INT8 Tensor Core	624 TOPS | 1,248 TOPS*	
Peak INT4 Tensor Core	1,248 TOPS | 2,496 TOPS*	
GPU Memory	40GB	80GB
GPU Memory Bandwidth	1,555 GB/s	2,039 GB/s
Interconnect	NVIDIA NVLink 600 GB/s** PCIe Gen4 64 GB/s	
Multi-Instance GPU	Various instance sizes with up to 7 MIGs @ 10 GB	
Form Factor	4/8 SXM on NVIDIA HGX™ A100	
Max TDP Power	400 W	400 W



Price: **$13,209.07** + Free Shipping

Quantity: 1 **Add to Cart**

More to Come

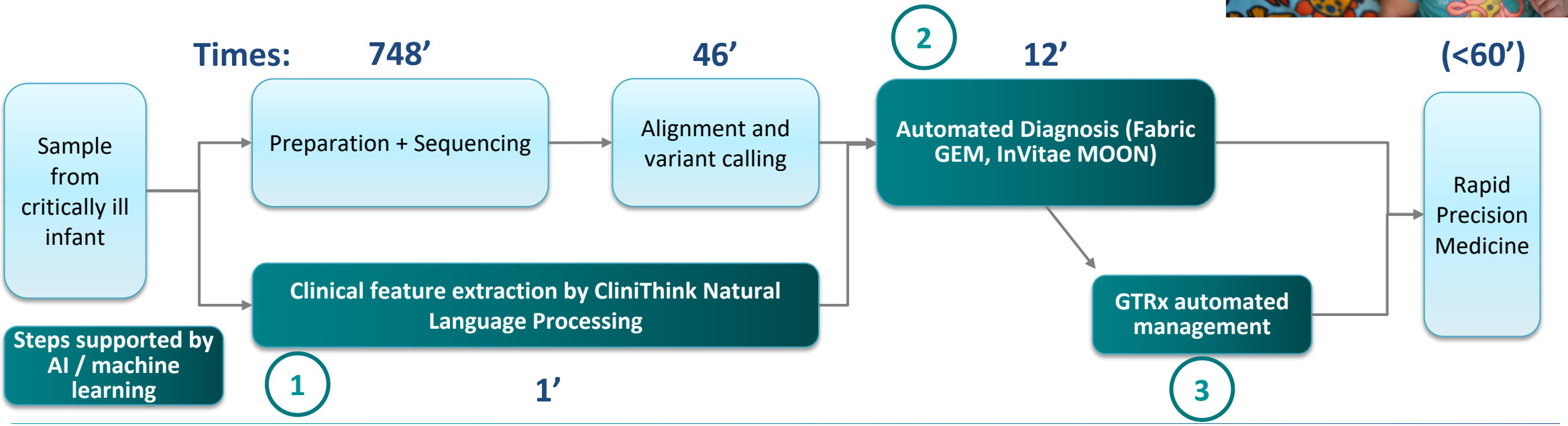
Last Updated: Mar 26, 2021

Rapid AI Medicine for Critically Ill Infants

Rady Children's Institute
Genomic Medicine

Every minute without a diagnosis counts

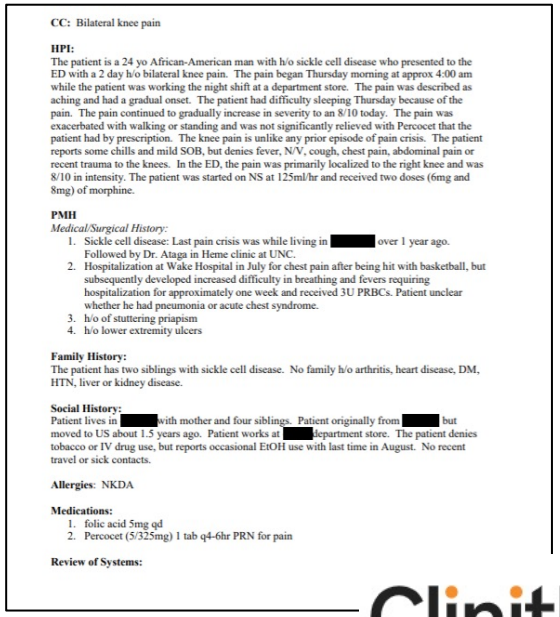
The fastest possible diagnosis in 13.5 hours for these babies, using AI tools to optimize for speed and performance



Clinical Natural Language Processing (CNLP)

CliniThink or CLAMP, **transforms unstructured EHR data** into a **structured list of Human Phenotype Ontology Terms**

CliniThink has been **iteratively trained on Rady Children's electronic health records** for optimum performance in extraction of terms relevant to rare genetic diseases



CC: Bilateral knee pain

HPI:
The patient is a 24 yo African-American man with h/o sickle cell disease who presented to the ED with a 2 day h/o bilateral knee pain. The pain began Thursday morning at approx 4:00 am while the patient was working the night shift at a department store. The pain was described as aching and had a gradual onset. The patient had difficulty sleeping Thursday because of the pain. The pain continued to gradually increase in severity to an 8/10 today. The pain was exacerbated with walking or standing and was not significantly relieved with Percocet that the patient had by prescription. The knee pain is unlike any prior episode of pain crisis. The patient reports some chills and mild SOB, but denies fever, N/V, cough, chest pain, abdominal pain or recent trauma to the knees. In the ED, the pain was primarily localized to the right knee and was 8/10 in intensity. The patient was started on NS at 125ml/hr and received two doses (6mg and 8mg) of morphine.

PMH
Medical/Surgical History:
1. Sickle cell disease: Last pain crisis was while living in [redacted] over 1 year ago. Followed by Dr. Ataga in Heme clinic at UNC.
2. Hospitalization at Wake Hospital in July for chest pain after being hit with basketball, but subsequently developed increased difficulty in breathing and fevers requiring hospitalization for approximately one week and received 3U PRBCs. Patient unclear whether he had pneumonia or acute chest syndrome.
3. h/o of stuttering priapism
4. h/o lower extremity ulcers

Family History:
The patient has two siblings with sickle cell disease. No family h/o arthritis, heart disease, DM, HTN, liver or kidney disease.

Social History:
Patient lives in [redacted] with mother and four siblings. Patient originally from [redacted] but moved to US about 1.5 years ago. Patient works at [redacted] department store. The patient denies tobacco or IV drug use, but reports occasional EtOH use with last time in August. No recent travel or sick contacts.

Allergies: NKDA

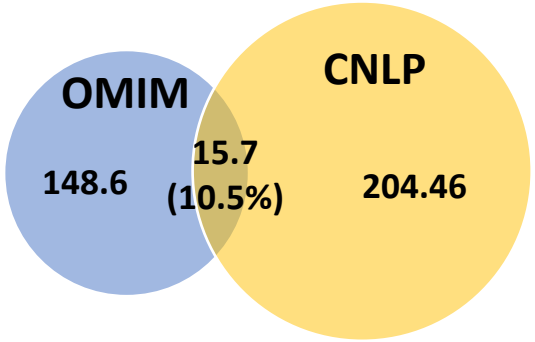
Medications:
1. folic acid 5mg qd
2. Percocet (5/325mg) 1 tab q4-6hr PRN for pain

Review of Systems:

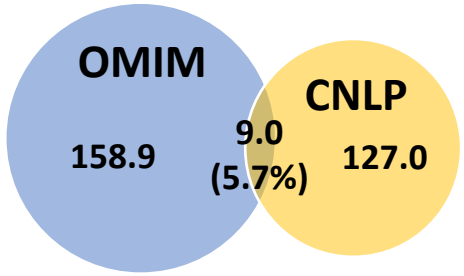
Clinithink

HPO Code	HPO Term	Frequency
hp0000152	Abnormality_of_head_or_neck	10
hp0000234	Abnormality_of_the_head	10
hp0002086	Abnormality_of_the_respiratory_system	9
hp0000765	Abnormality_of_the_thorax	8
hp0001438	Abnormality_of_abdomen_morphology	8
hp0002170	Intracranial_hemorrhage	8
hp0000707	Abnormality_of_the_nervous_system	7
hp0001892	Abnormal_bleeding	7
hp0001928	Abnormality_of_coagulation	7
hp0002088	Abnormality_of_lung_morphology	7
hp0002107	Pneumothorax	7
hp0003256	Abnormality_of_the_coagulation_cascade	7
hp0003270	Abdominal_distention	7
hp0100750	Atelectasis	7
hp0001298	Encephalopathy	6
hp0001342	Cerebral_hemorrhage	6
hp0012443	Abnormality_of_brain_morphology	6
hp0000969	Edema	5
hp0001626	Abnormality_of_the_cardiovascular_system	5
hp0001941	Acidosis	5
hp0002118	Abnormality_of_the_cerebral_ventricles	5
hp0002615	Hypotension	5
hp0011024	Abnormality_of_the_gastrointestinal_tract	5
hp0025031	Abnormality_of_the_digestive_system	5
hp0030746	Intraventricular_hemorrhage	5

Infant Deaths (n=30)



Infants who did not die (n=122)

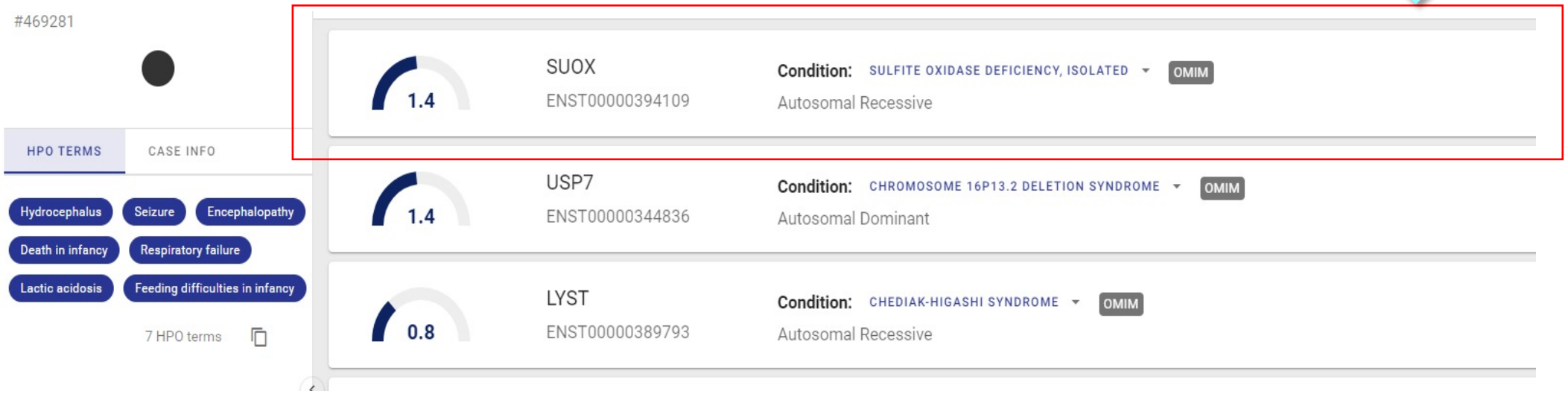


Automated Interpretation Reduces Analytic Time

GEM (Fabric Genomics) and MOON (InVitae)

Prioritizes and ranks variants

GEM accurately made this diagnosis of **isolated sulfite oxidase deficiency** due to a mutation in the ***SUOX* gene**



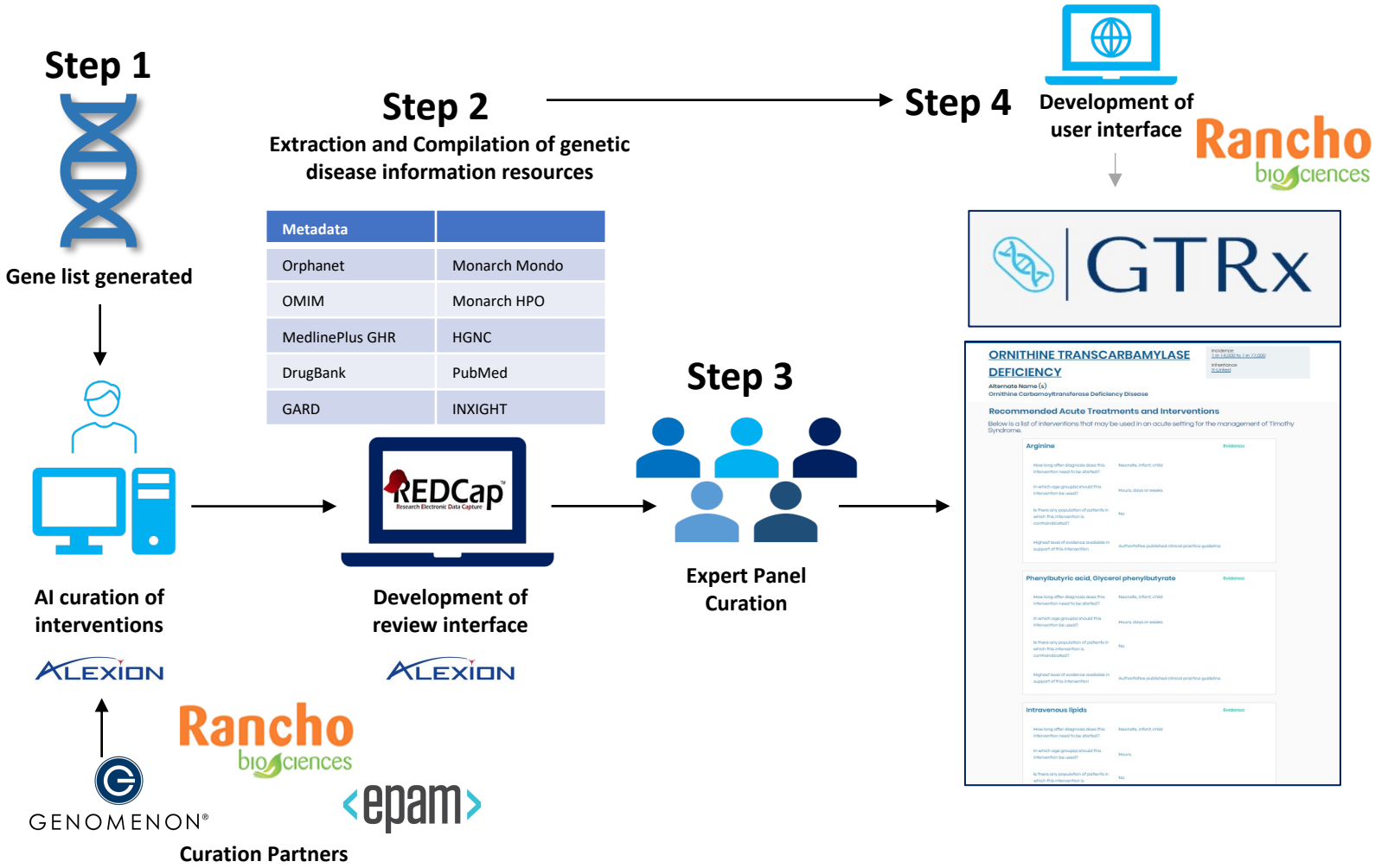
Rapid, Precision Management for Rare Genetic Conditions

Genome-To-Treatment (GTRx)

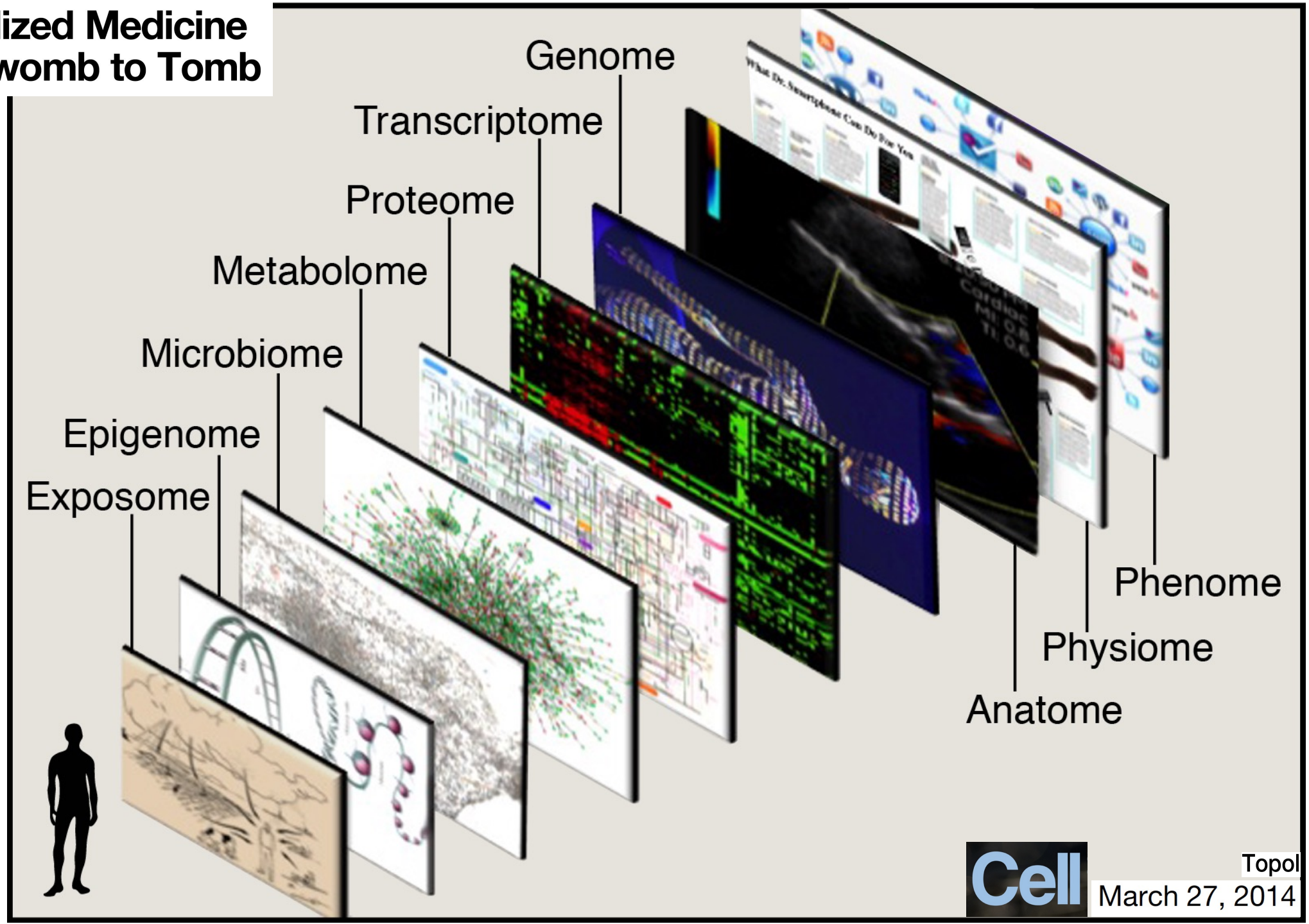
an automated system for immediate, 24-hour management of newly diagnosed genetic conditions

AI was used to pull references for a list of 358 severe, treatable genetic conditions and extract relevant interventions

Available in a web resource for frontline clinicians

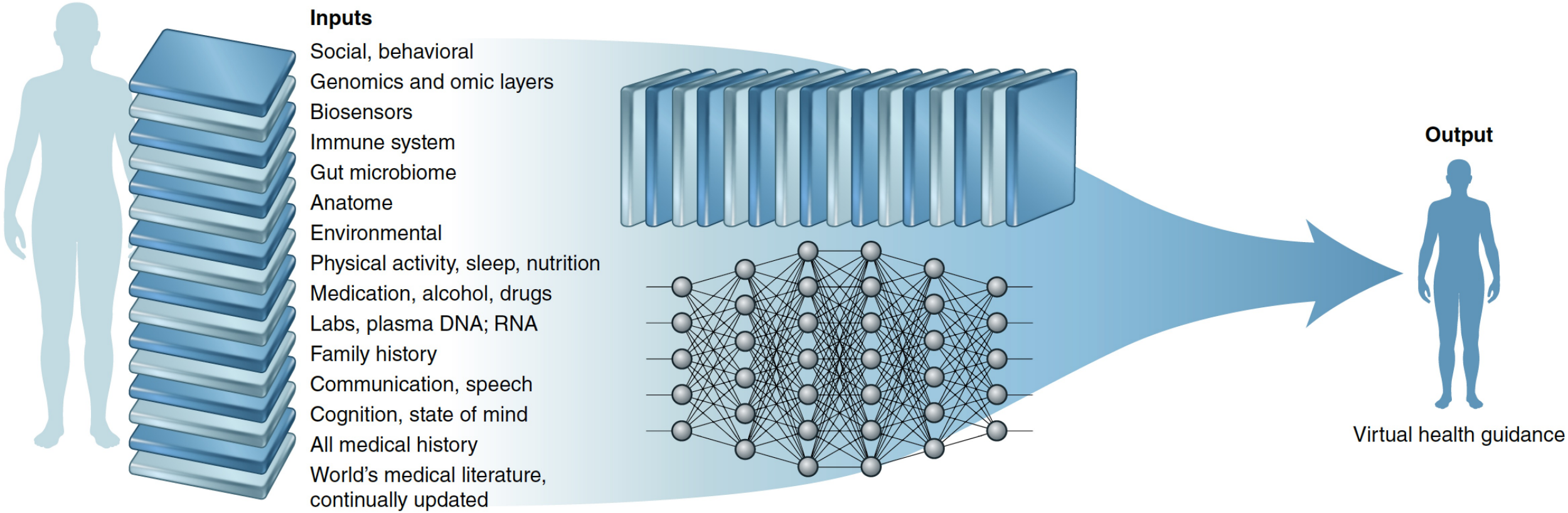


Individualized Medicine from Prewomb to Tomb



Topol
March 27, 2014

The Virtual Medical Assistant



7 Jan 2019

nature medicine

Acknowledgements

Ali Torkamani
Steven Steinhubl
Kristian Andersen
Andrew Su
Emily Spencer
Evan Muse
Katie Baca-Motes
Pejman Mohammadi
Laura Nicholson
Athena Philis-Tsimikas
Nathan Wineinger
Michelle Miller
Kristina Haro
Chunlei Wu
Gail Ebner
Julia Moore Vogel
Daniel Oran
Lorraine Evangelista
Luc Teyton
Jill Waalen
Lase Ajayi
Jennifer Radin
Giorgio Quer
Lauren Ariniello
Matteo Gadaleta
Theresa Hill
Kendall Laycock
Refugio Robles-Sikisaka
Julia Menard
Anna Andersen
Doug Evans
Elias Salfati
Raquel Dias
Sarah Topol
Will Liu
Tanya Hearne
Erin Coughlin
Amanda Schneider
Torrey Leavy
Gayle Simon
Dina Hamideh
Paula King
Shaquille Peters
Elise Felicione
Colleen McShane
Jane Samaniego
Maribel Perez-Medina
Amitabh Pandey
Stuti Jaiswal
Addie Fortmann
Courtney Prato
Meghan Grubel
Katie Quartuccio
Ed Ramos
Kalyani Kottilil
Sasri Dedigama
Nicole Phoenix
Matthew Thombs
Danielle Jones
Vik Kheterpal
Royan Kamyar
Paul Montgomery
Whitney Baldridge
Siddhartha Sharma
Don Clarke
Sophie Shevick
Dillon Flood
Marcela Mendoza Martinez
Tridu Huynh
Wesleigh Edwards

NIH National Institutes of Health
NIH National Center for Advancing Translational Sciences
CTSA Clinical & Translational Science Awards Program
CD2H National Center for Data to Health
NIH National Human Genome Research Institute
All of Us Research Program