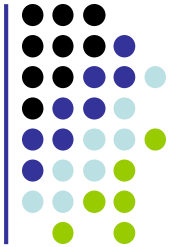
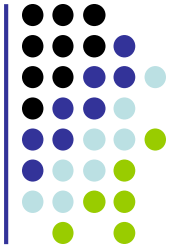


Why Deep Learning in Biomedicine

Serafim Batzoglou
Stanford University

Why deep learning in biomedicine

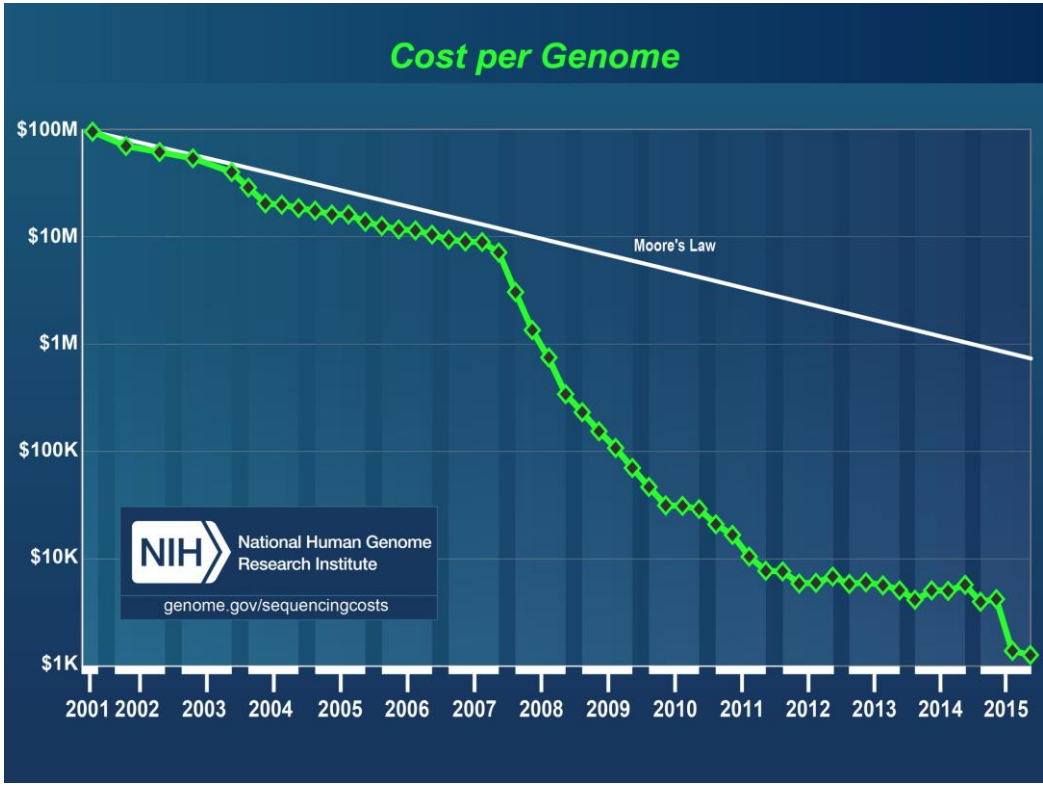


- Transformation of biomedicine into data science
- The future role of deep learning
- Why ***deep*** learning?
- The biggest obstacle and suggestions to overcome it

Sequencing Growth

Cost of one human genome

- 2000: $3,000,000,000
- 2004: $30,000,000
- 2008: $100,000
- 2010: $10,000
- **2015: $1000**
- ???: $300



Roche 454 SEQUENCING

SOLID

Helicos BioSciences Corporation

2nd Base A C G T
1st Base A C G T

Ion Torrent™ Next-Generation Sequencing
Find out more ›

PACBIO

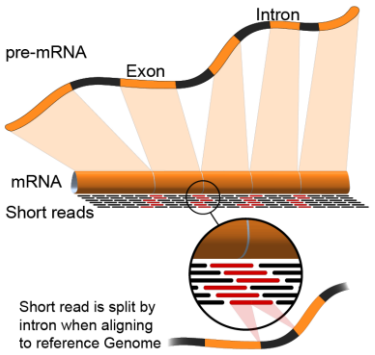
Oxford NANOPORE Technologies

Complete genomics

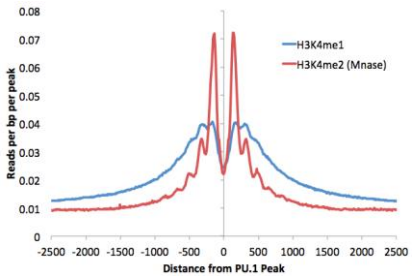
illumina

A Data-Acquisition Technology Explosion Enabled by Inexpensive Sequencing

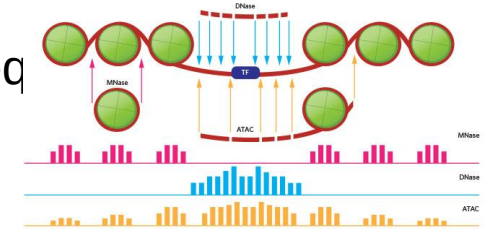
- RNAseq



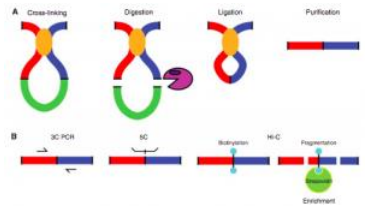
- DNase-seq



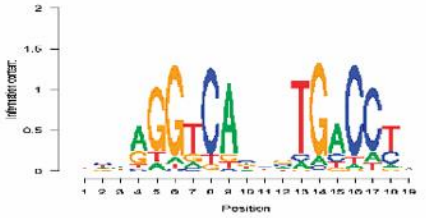
- ATAC-seq

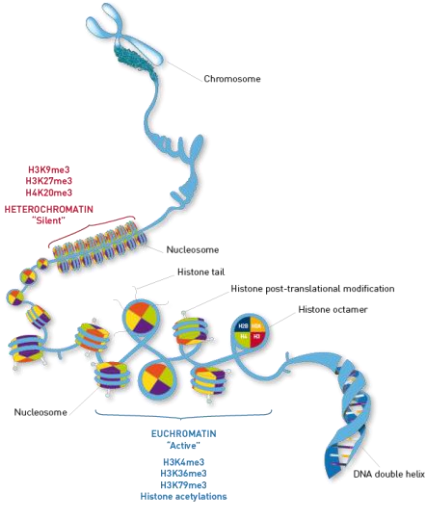


- Hi-C

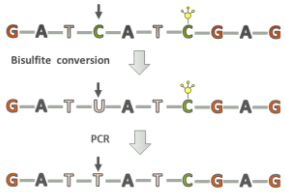


- ChIP-seq for
  - Transcription factor binding
  - Nucleosome positioning
  - Histone Modifications

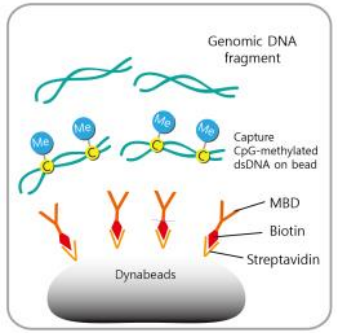




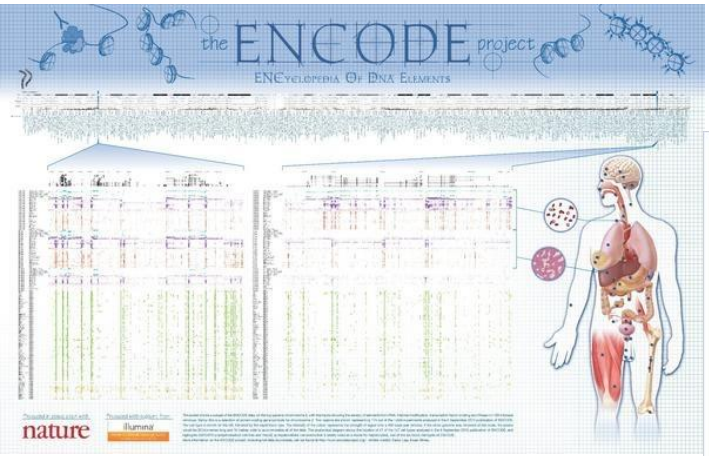
- Bisulfite treatment for methylation



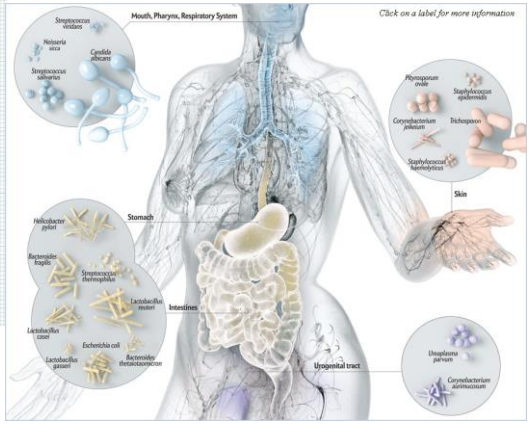
- MeDIP for methylation

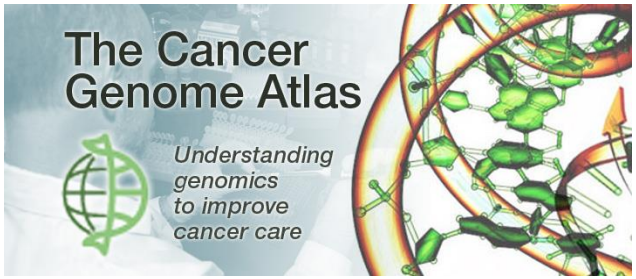


Major Data Acquisition Efforts

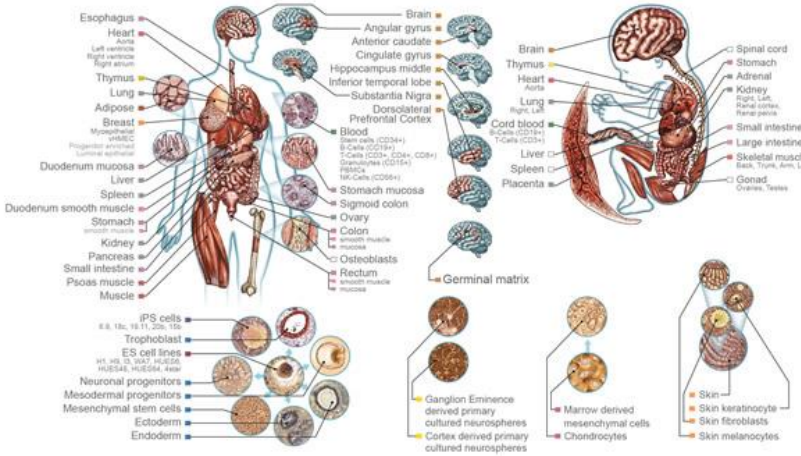


Human Microbiome



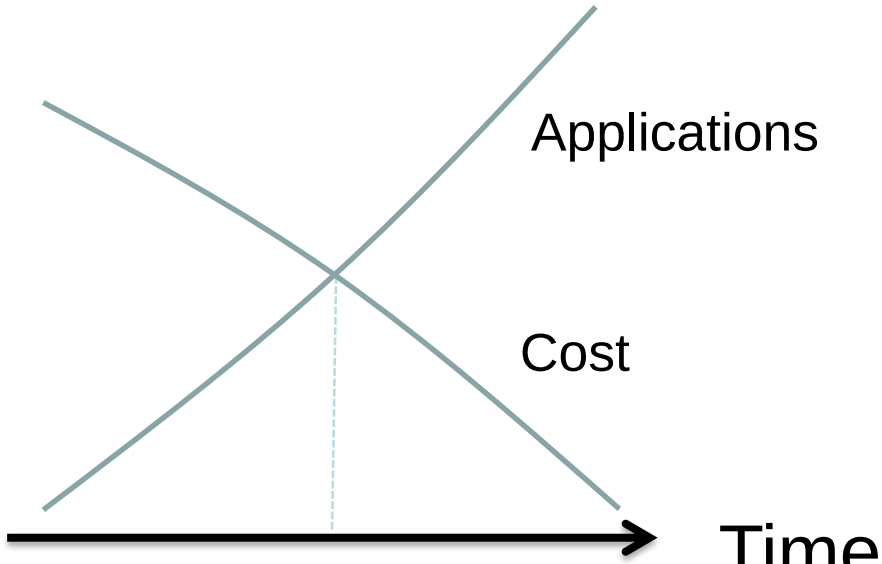


Roadmap Epigenomics

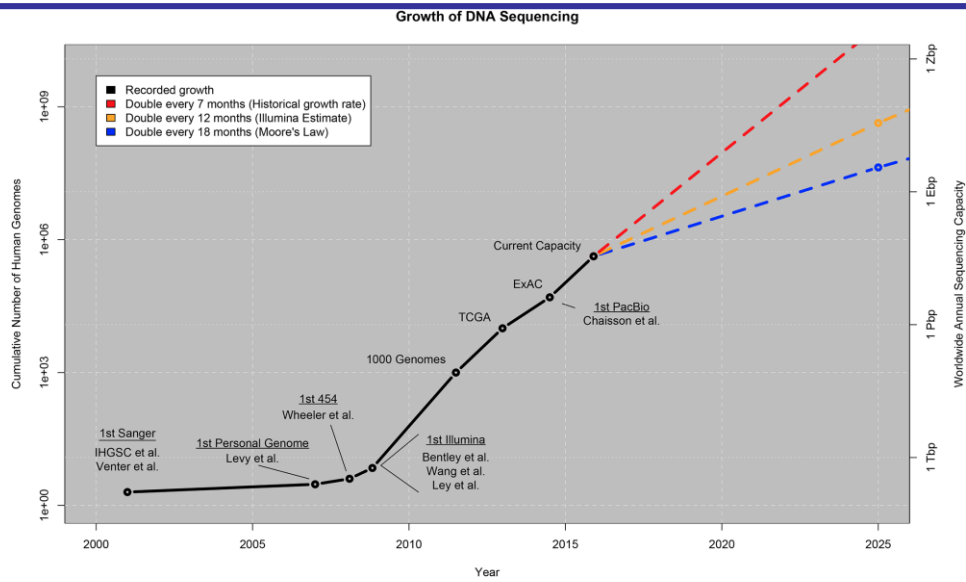


1000 Genomes

How soon will we all be sequenced?

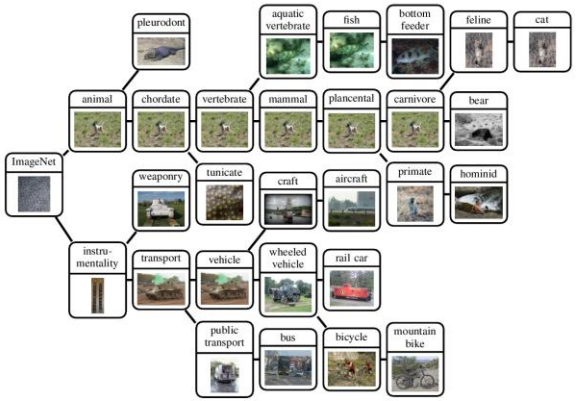


2016?
2022?

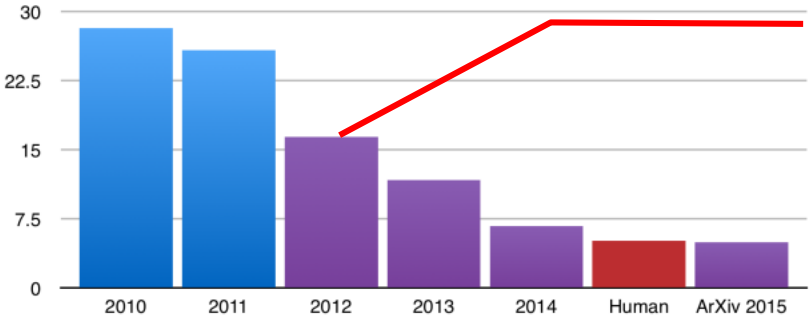


- 1Bn individuals by 2022-2026

Deep Learning – major AI breakthrough



ILSVRC top-5 error on ImageNet



Introduction of deep NNs


"man in black shirt is playing guitar."


"construction worker in orange safety vest is working on road."


"two young girls are playing with legos toy."


"boy is doing backflip on wakeboard."


"girl in pink dress is jumping in air."

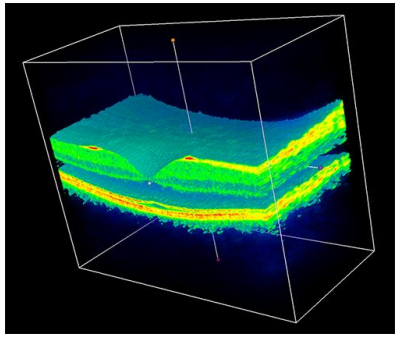

"black and white dog jumps over bar."

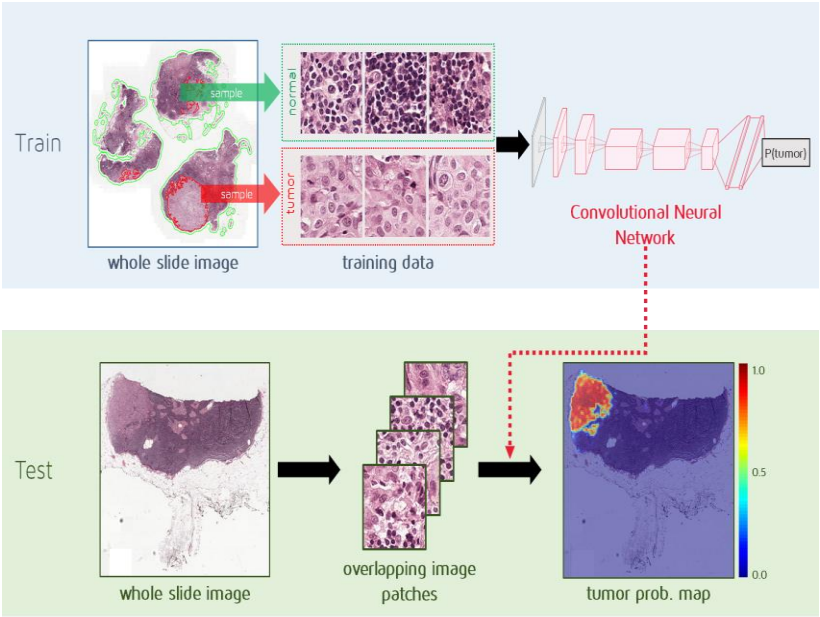

"young girl in pink shirt is swinging on swing."


"man in blue wetsuit is surfing on wave."

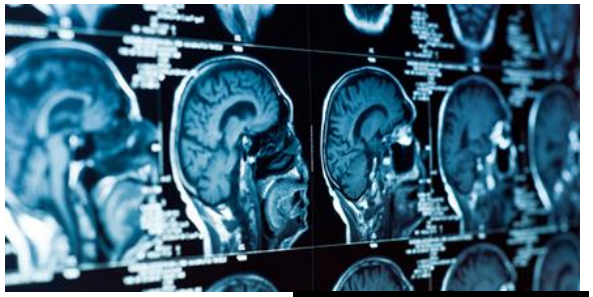
The beginning of deep learning in medicine

DeepMind Health





enlitic



MGH MASSACHUSETTS GENERAL HOSPITAL

NVIDIA

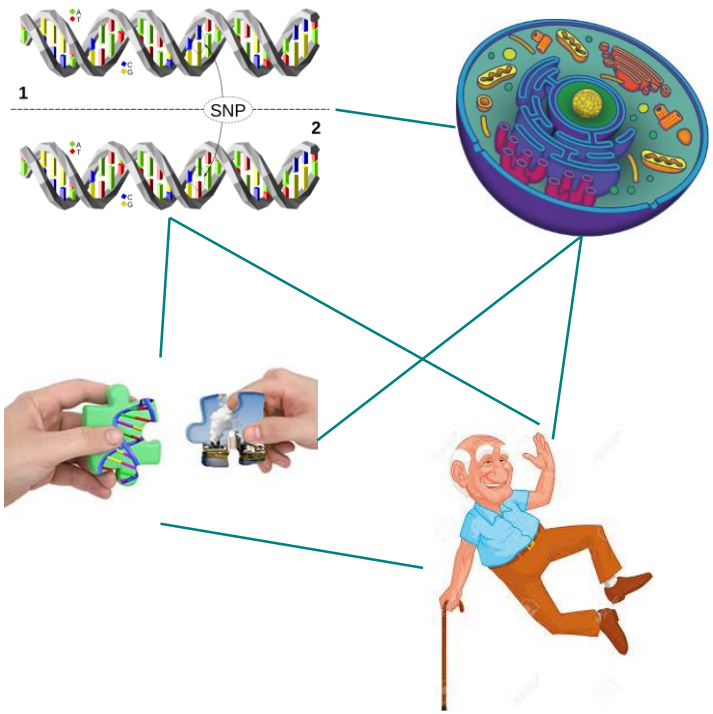
Bet: Biomedicine amenable to deep learning

Current medicine:

- With nurse help
- Oncology, heart, infectious diseases, dermatology, pharmacology, …, diagnosis, treatment

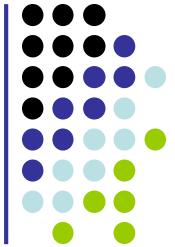
Benefits:

- Improved outcomes
- Prediction, prevention
- Low cost



- Personalized health & prevention
- Powerful AI, learning from billions of examples
- New drugs & treatments, gene therapy, CRISPR, …

OK, but why "*deep* learning"?



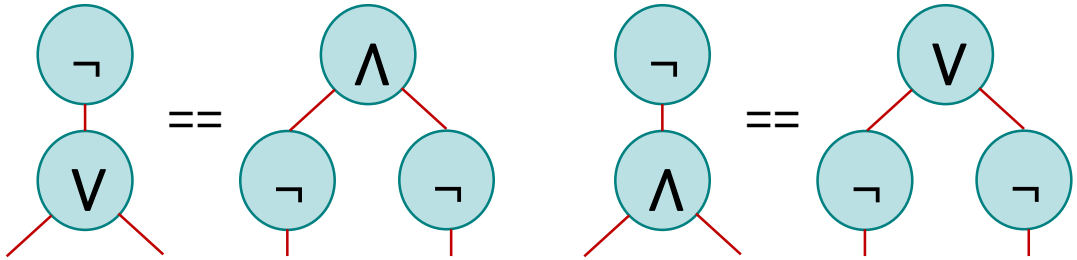
Let's digress and talk about

Boolean circuits

OK, but why "*deep* learning"?

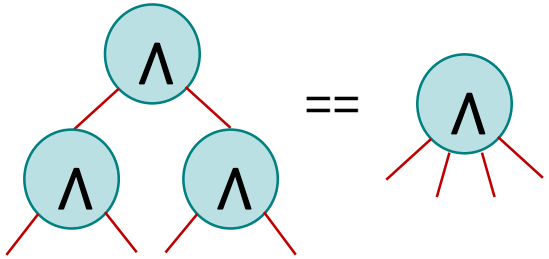
Consider circuits built of: ∨, ∧, ¬

Can always push ¬ to bottom layer

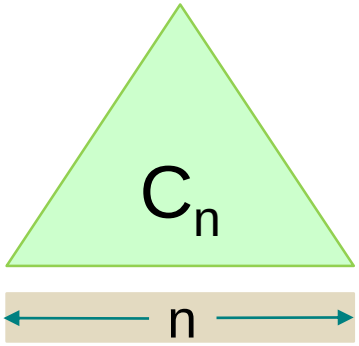


Assume alternating ∨, ∧

Unlimited fan-in



Assume a circuit C_n that handles all inputs of length n

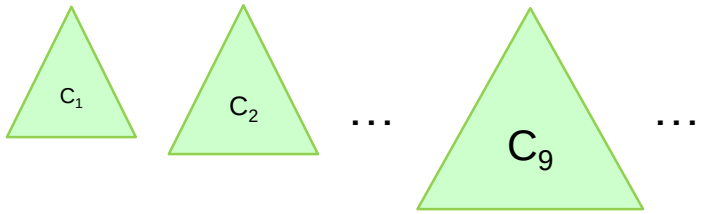


OK, but why "*deep* learning"?

A collection of circuits $C_1, \ldots, C_n, \ldots$ is

P-uniform: There is a poly-time Turing Machine that on input n, outputs C_n

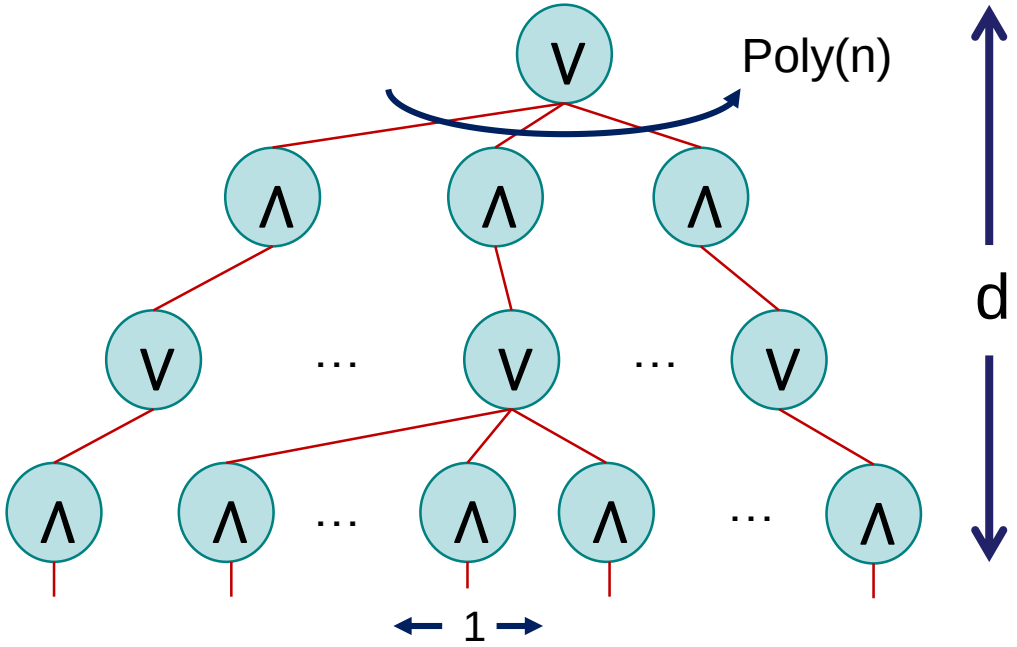
A language L is computable by a P-uniform circuit family, if and only if $L \in P$



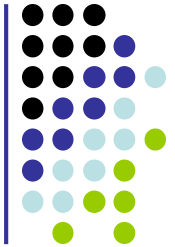
Class AC^0:

Unlimited fan-in circuits of constant depth d

Assume bottom layer fan-in 1



What can we compute with a constant-depth circuit?

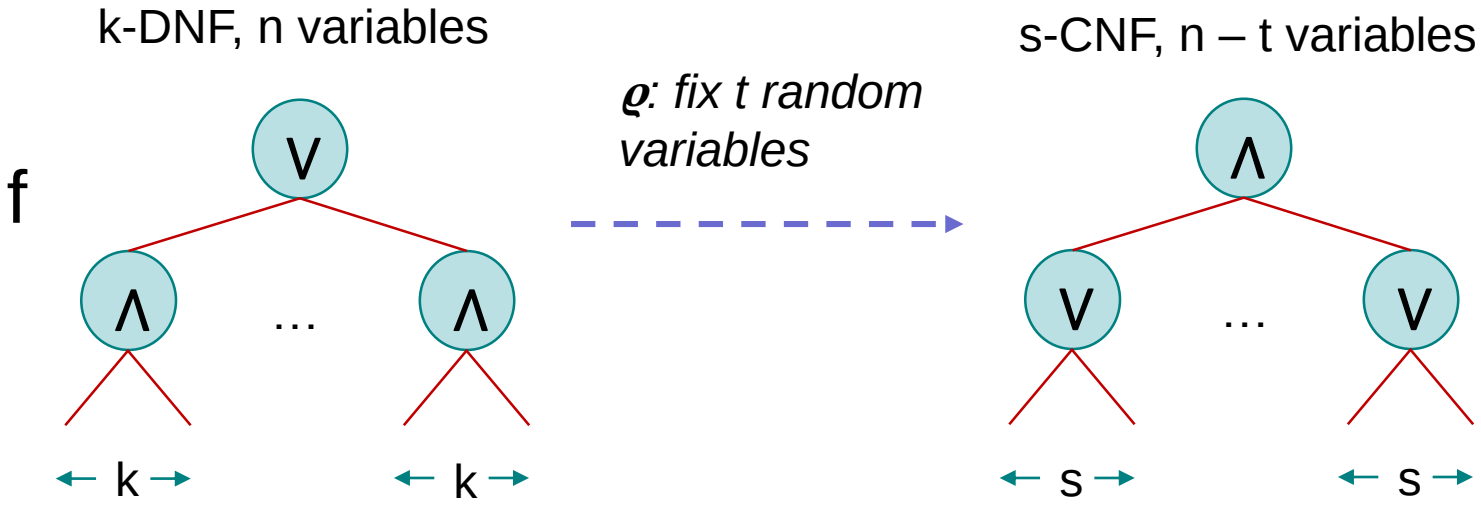


- Start with **depth-d** circuit
- **Collapse 1st & 2nd layers** – Switching Lemma

…

- Flatten to **2 layers** without making it “too fat”
- What can we now compute?

Switching Lemma



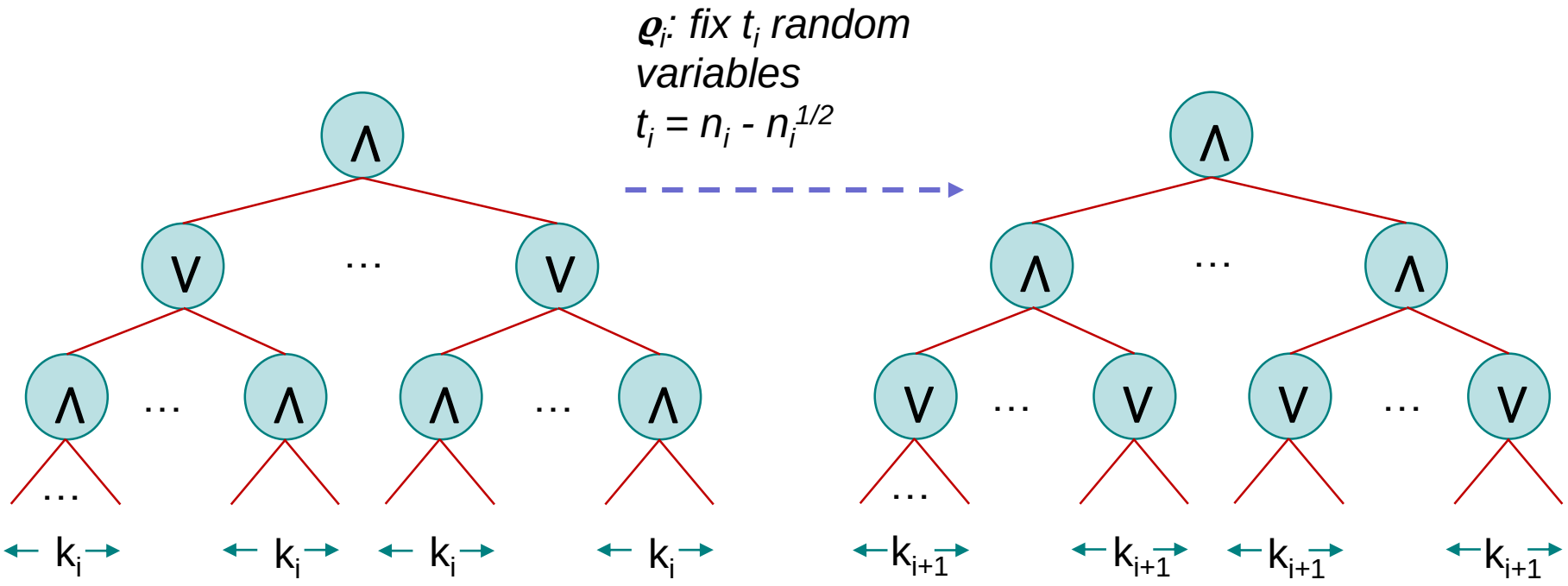
Switching Lemma (CNF ←→ DNF)

Prob [$f|_{\varrho}$ not expressed as a s-CNF] ≤ q]

Where q is equal to the ugly:

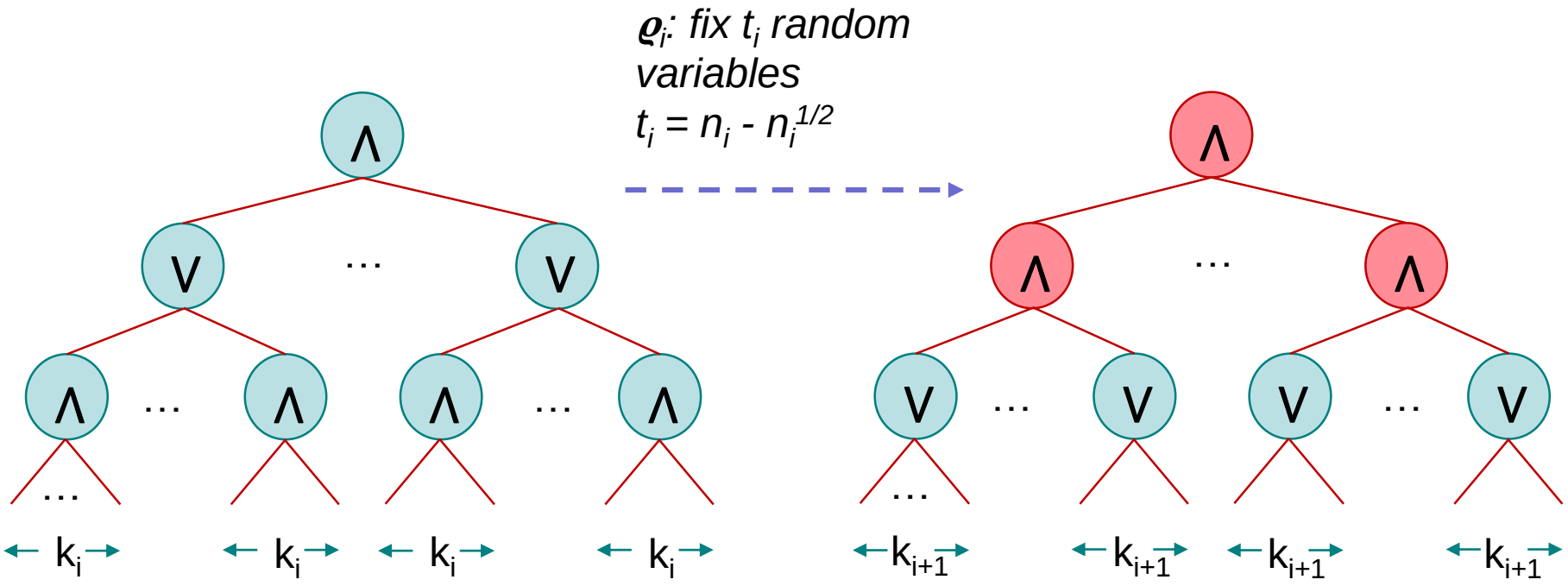
$$q = q(k, s, t, n) = ((n-t)k^{10}/n)^{s/2}$$

Reducing depth by 1



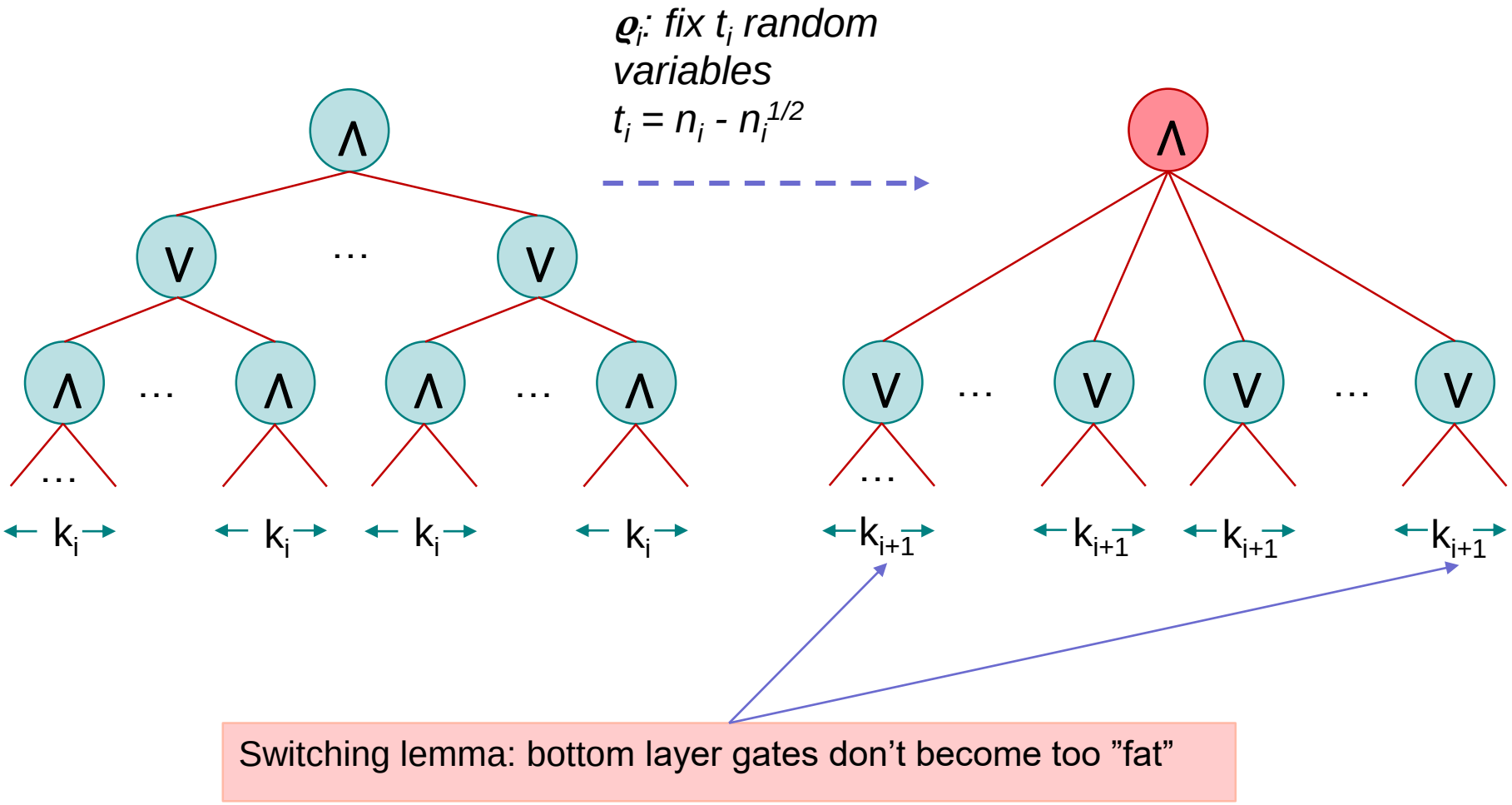
Always reduce the bottom layer, initially of fan-in 1

Reducing depth by 1



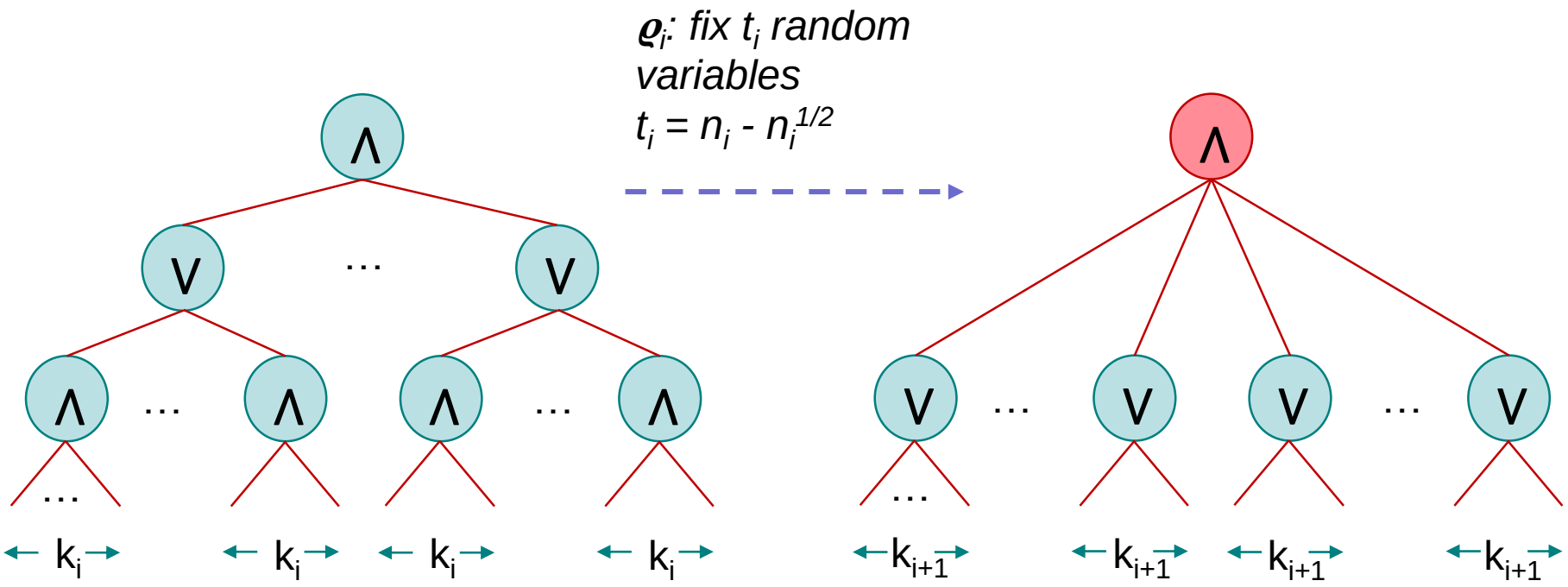
Always reduce the bottom layer, initially of fan-in 1

Reducing depth by 1



Switching lemma: bottom layer gates don't become too "fat"

Reducing depth by 1

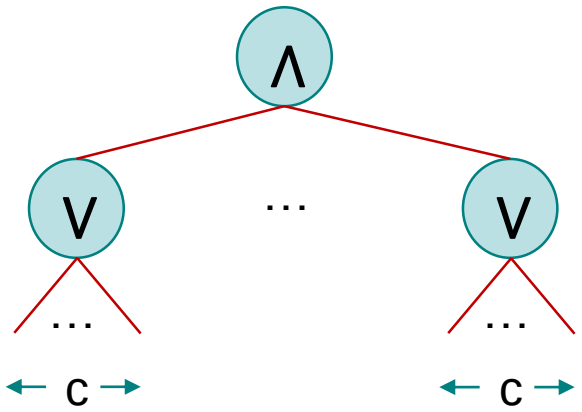


After some hairy calculations,

After d-2 steps, each applied to all layer-2 gates:

# variables left:	$n^{1/(2^{d-2})}$
depth:	2
fan-in:	10 b 2^{d-2}, *where circuit size = n^b*

Final Circuit



# variables left:	$n^{1/(2^{d-2})}$
depth:	2
fan-in:	c = 10b2^{d-2}

Turn any circuit family in AC^0 into 2-layer

If n large, #vars >> fan-in
Fix c vars, make output constant = 0

Now take **f** = parity(x_1, …, x_n)
≠ constant by fixing any s < n variables

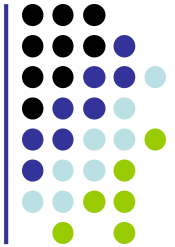
Therefore, **f** ∉ AC^0

A "shallow" ML method cannot compute parity unless exponential # features

How about HMM, K states? N layers, each layer receives log K bits from previous layer. Can do parity, but limited to simple functions

Shallow ML seems unable to replace general algorithms

Biggest Obstacle: wide, free data availability



Ideal:

- Millions of publicly available genomes, m… …onment
- Far-reaching anti data-based discrim…

Not ideal:

- Data silos in hospitals, health care pr… …nies
- Specialized large-scale projects (1000 ge… …ical records)

"I don't want to live in a world where someone else makes the world a better place better than I do."

What about privacy?

- No cancer patient has died yet because of data privacy breach
- My Facebook page reveals much more private stuff about me than my genome or medical record

Let's make our data public – if 5% of us do, probably enough to solve medicine

- Let's do so without waiting for legislation

Some examples to follow

You think recycling saves the world?

Try data sharing!

Anti-discrimination law and public awareness extends to all genetic and medical information

Data hoarders (academia, hospitals, industry) akin to polluters