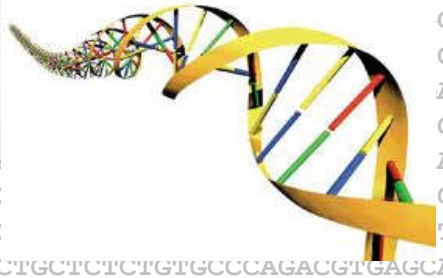
Build Systems to Reverse Engineer The Most Amazing Operating System on the Planet



One GENOME to Rule Them All

Gill Bejerano

Today's Agenda

You should rotate.

You should rotate in Genomics.

You should rotate in Genomics with the Bejerano lab.

Why Rotate?

Welcome to (one of) the best CS depts in the world!

Our CS dept. encourages you to undertake three rotations, over three quarters before aligning for your PhD.

So you try new things before picking your field.

So you are happier by graduation with your choice of PhD, often with the choice of a career for life.

Don’t rush.

Why Rotate *Away* from your Undergrad Field?

To get into Stanford you must have done well as undergrad in some field of CS.

However, researchers, at all levels, often do their best work when they cross fields, taking ideas from one field to another. You should cross now.

Moreover, regardless of your lab choice, in 3-4 years, before graduating, you will be happier if you've rotated. (We know because we polled our students).

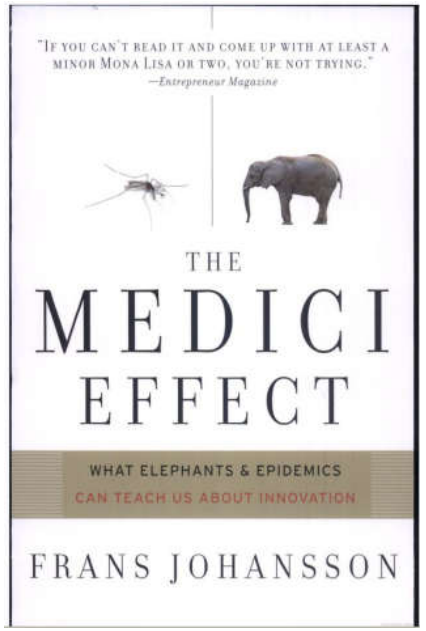
Don't rush.

Why Rotate in a *Young* Field?

The most exciting science,
the most opportunities to change the world,
the most faculty positions,
the most industry / start-up opportunities,
lie in young fields.

In particular, most opportunities lie
at new intersections of established fields.

Don't rush.



Why Rotate in Genomics?

Genomics is at the intersection of CS & BioMedicine.

Genomics is fundamentally a computational science.
(Each genome is a program instance).

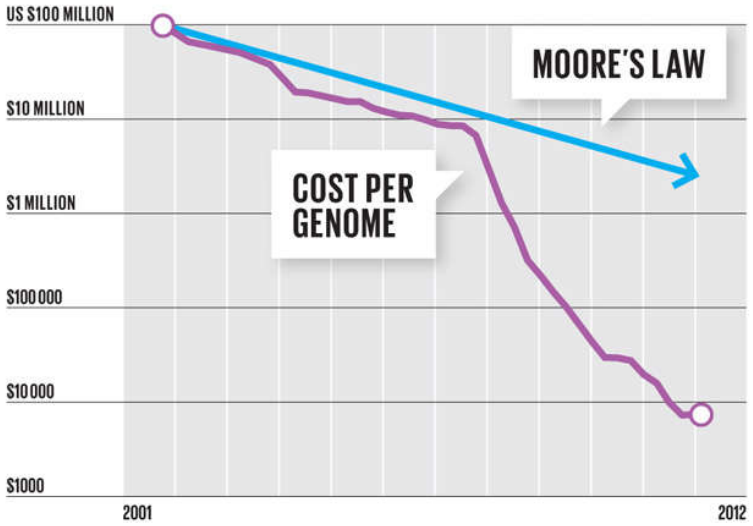
This century is owned by Genomics.

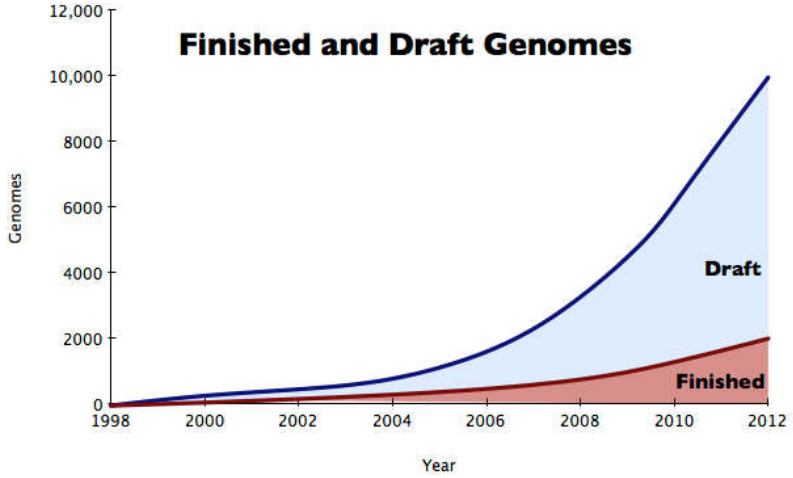
"There is gold in them thar hills" – lots to build & discover!

"I don't know any biology.." "Not a problem!"

Neither did many of us when we started.
(I hadn't taken a biology class until grad school;
The human genome was first assembled by a graphics programmer).

Genomics is affecting multiple CS fields





AI
Crypto
Systems
Databases
HCI
etc.

Even if you do not want to be a genomicist, some of the most exciting challenges in <u>your</u> field will be at the interface with Genomics. Understand it.

Everybody wants a piece of the action.

Wow… I should learn some Genomics!

CS 273B: Deep Learning in Genomics and Biomedicine

Recent breakthroughs in high-throughput genomic and biomedical dat
data" disciplines. In parallel, progress in deep neural networks are rev
natural language processing and, more broadly, AI. This course explore
advances. The course will start with an introduction to deep learning
and high-throughput biotechnology, focusing on the available data and
developments in deep learning (supervised, unsupervised and generati
these methods to biomedical data, which are beginning to produced
modeling, the course emphasizes how to visualize and extract interpre
papers from the literature will be presented and discussed. Stude mo

Terms: Aut | Units: 3 | Grading: Medical Option (Med-Ltr-CR/NC)

Instructors: Kundaje, A. (PI) ; Zou, J. (PI)

Schedule for CS 273B

Fall Quarter

`http://cs273a.stanford.edu`

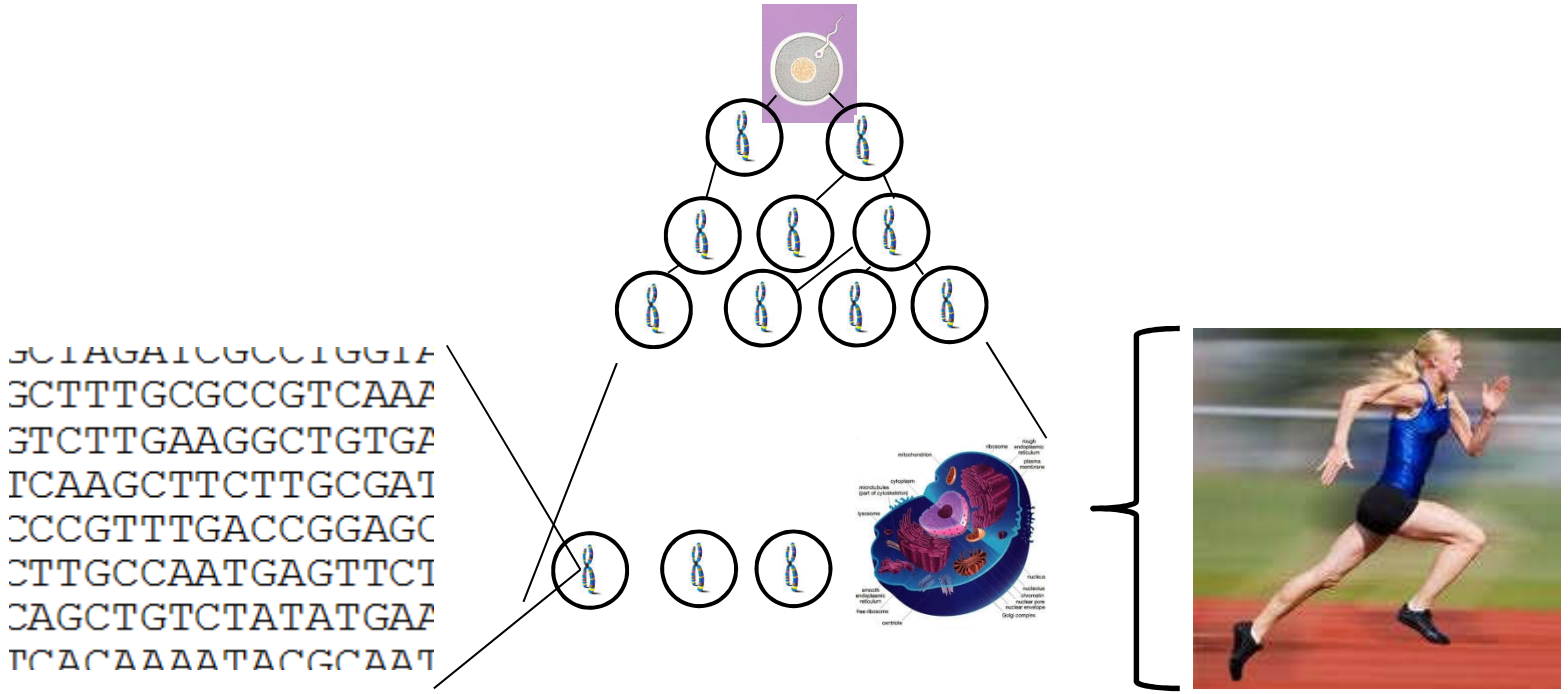
Winter Quarter

Okay, Fine, What is Genomics?

The Most Amazing Operating System on the Planet: The Human Genome

hOS

The Human Genome



Genome 10^{13} cells ≈ Human

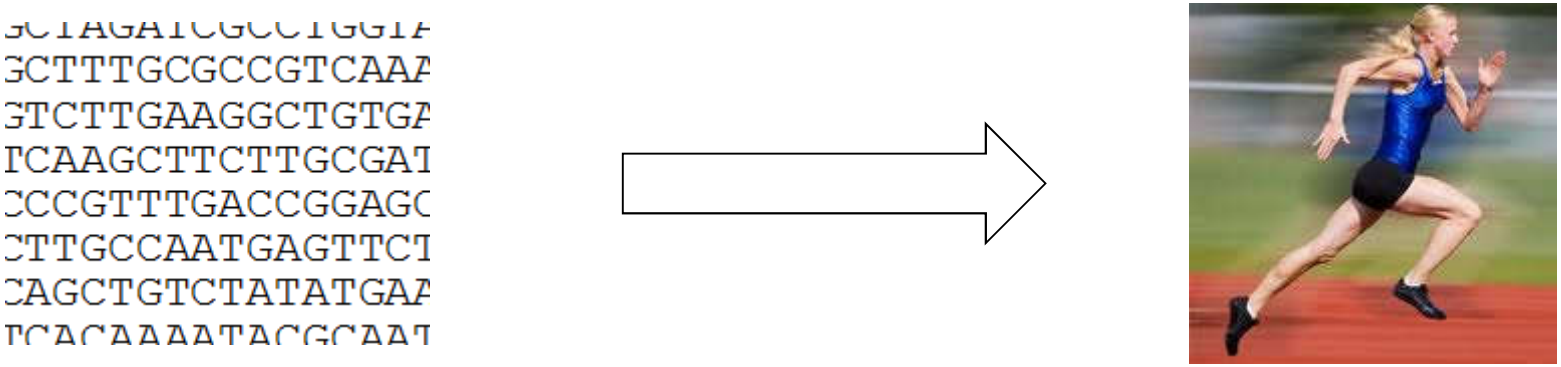
Genome = The Operating System that runs <u>every</u> cell in our body

$3*10^9$ letters long, over the DNA alphabet = {A,C,G,T}

The Biggest Challenge in Genomics…

… is computational:

How does this



Program

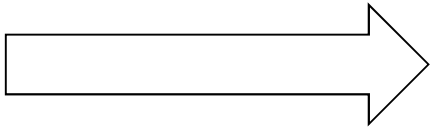
Output

This “coding” question has profound implications for our lives

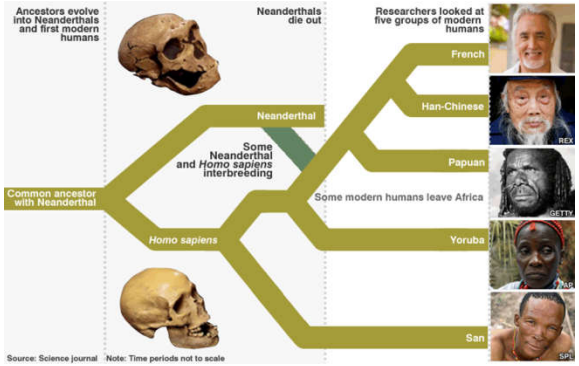
The Biggest Challenge in Genomics…

… is computational:

How does this



encode *this*



Program

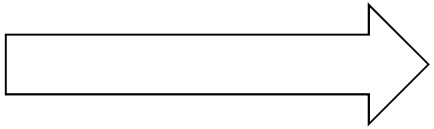
Forks & re-merges

Where did we come from? How are we different from each other?

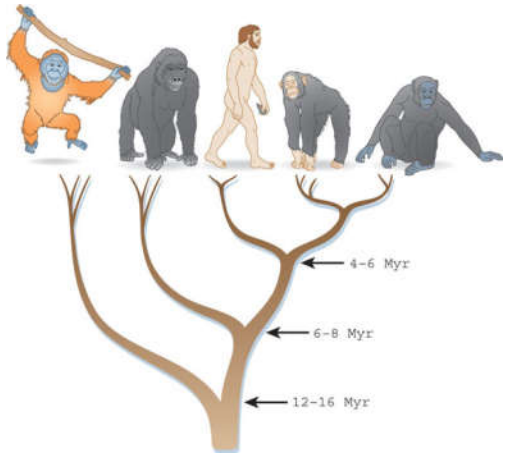
The Biggest Challenge in Genomics…

… is <u>computational</u>:

How does this



encode *this*



Program

Suite of related products

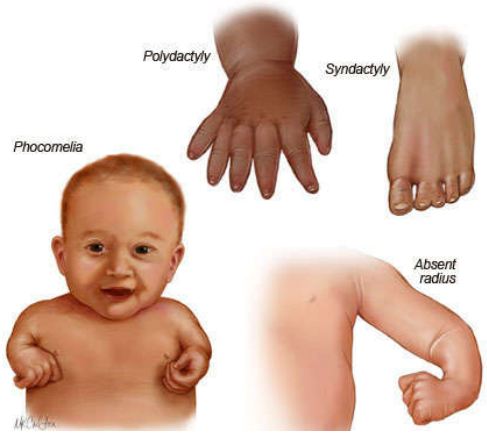
What in our genomes make us different from other species?

The Biggest Challenge in Genomics…

… is computational:

How does this

encode *this*



Program

Bugs

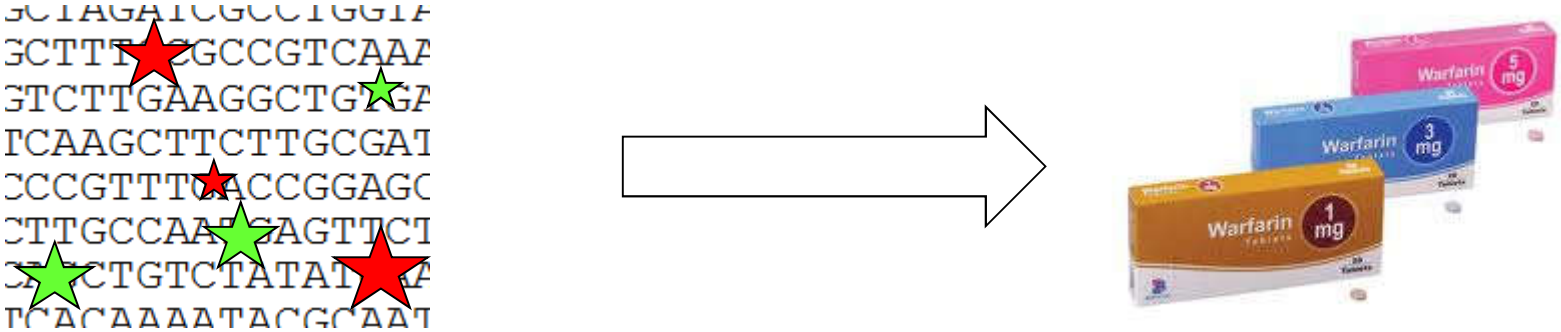
Output

What genomic mutations predispose us to disease?

The Biggest Challenge in Genomics…

… is computational:

How does this encode *this*



Program Bugs Patching

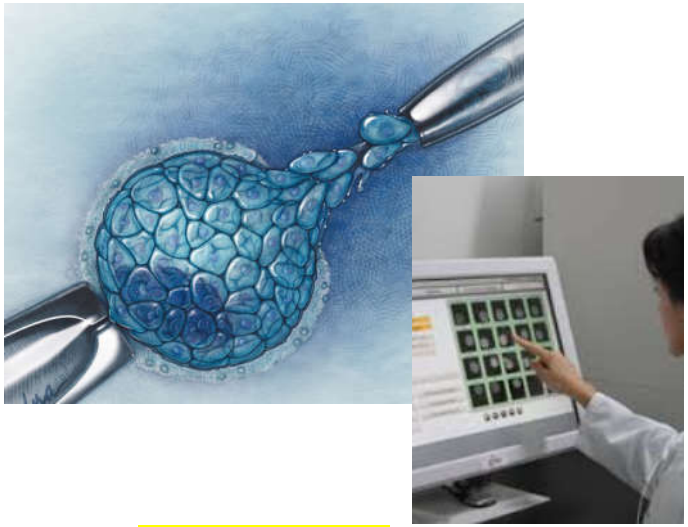
What genomic mutations determine our drug response?

The Biggest Challenge in Genomics…

… is computational:

How does this

encode *this*



Program

Bugs

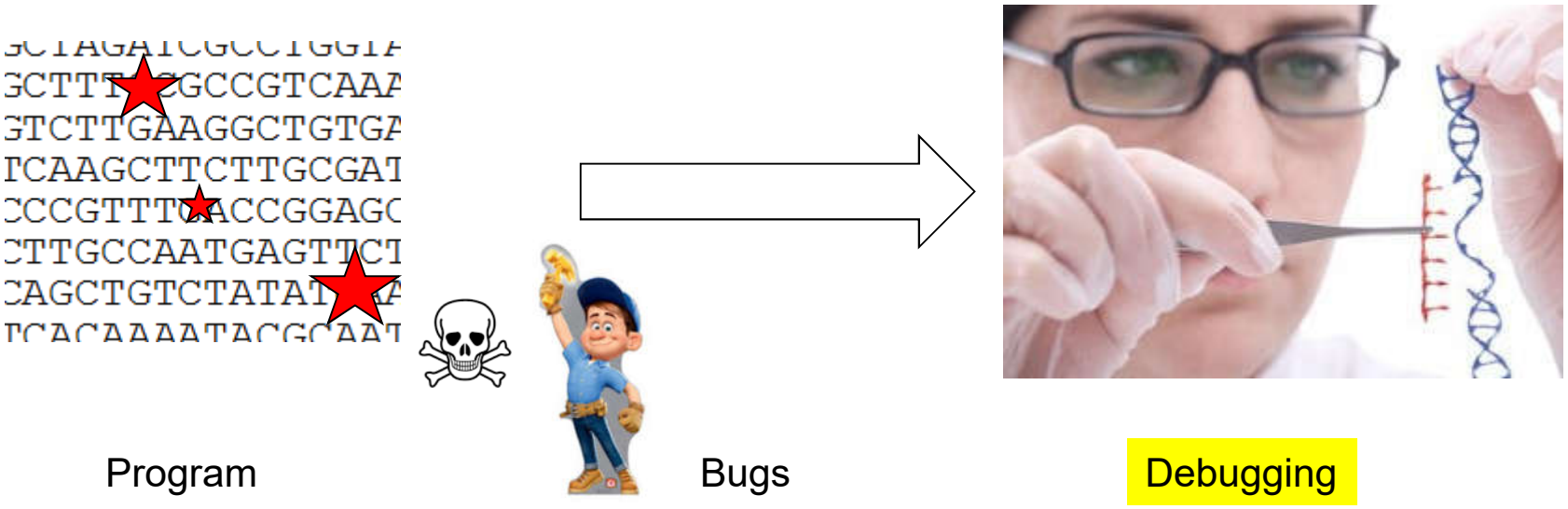
Verification

We can eliminate suffering by not “booting” “buggy” embryos

The Biggest Challenge in Genomics…

… is computational:

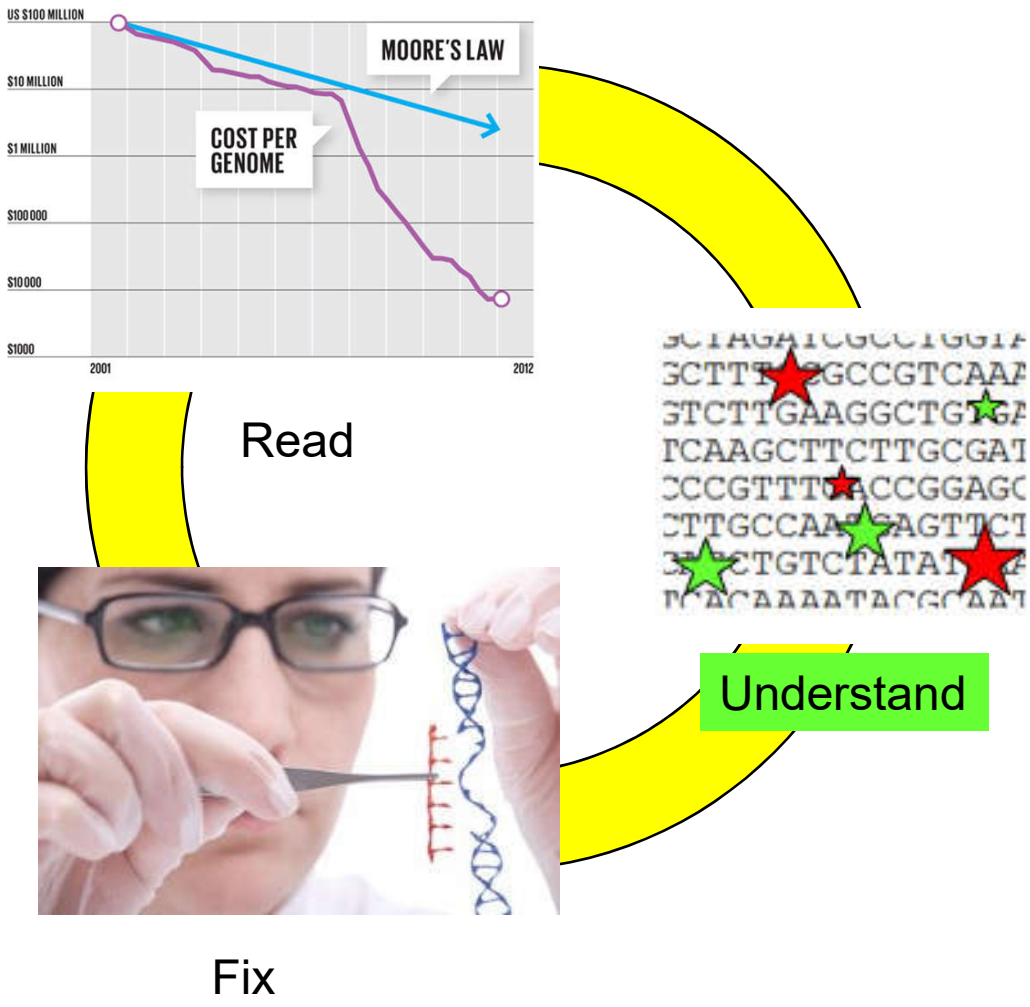
How does this encode *this*



Program Bugs Debugging

We can eliminate suffering by fixing people's "buggy" genomes

Literally Save Lives From Your Keyboard!



Gene Therapy: Find the Cause, Bring the Cure

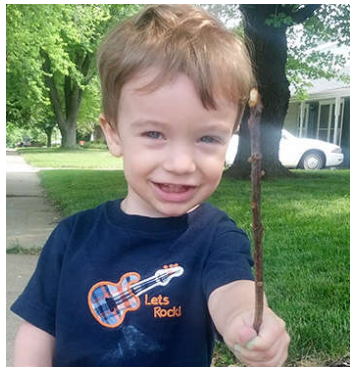
The Atlantic SUBSCRIBE SEARCH MENU≡

HEALTH

How to Cure a Bubble Boy

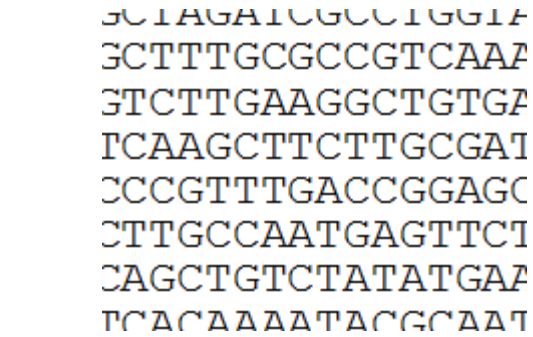
Thanks to gene therapy, a boy born without an immune system can now play in the yard.

Life saving
“code injection”

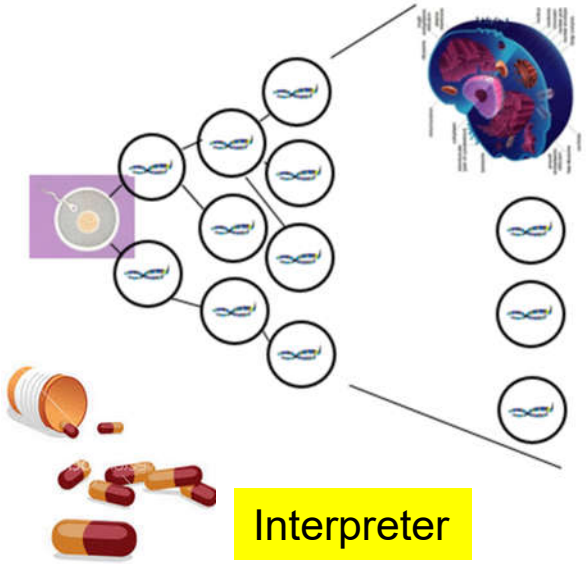


Biomedicine is facing a phase transition

From an obsession with the interpreter,



Code



Interpreter

Output

To an obsession with the code.

(When your code has a bug what do you fix?..)

Okay.. what do *we* do in Genomics?

We build systems to
Reverse Engineer the Most Amazing
Operating System on the Planet:
The Human Genome

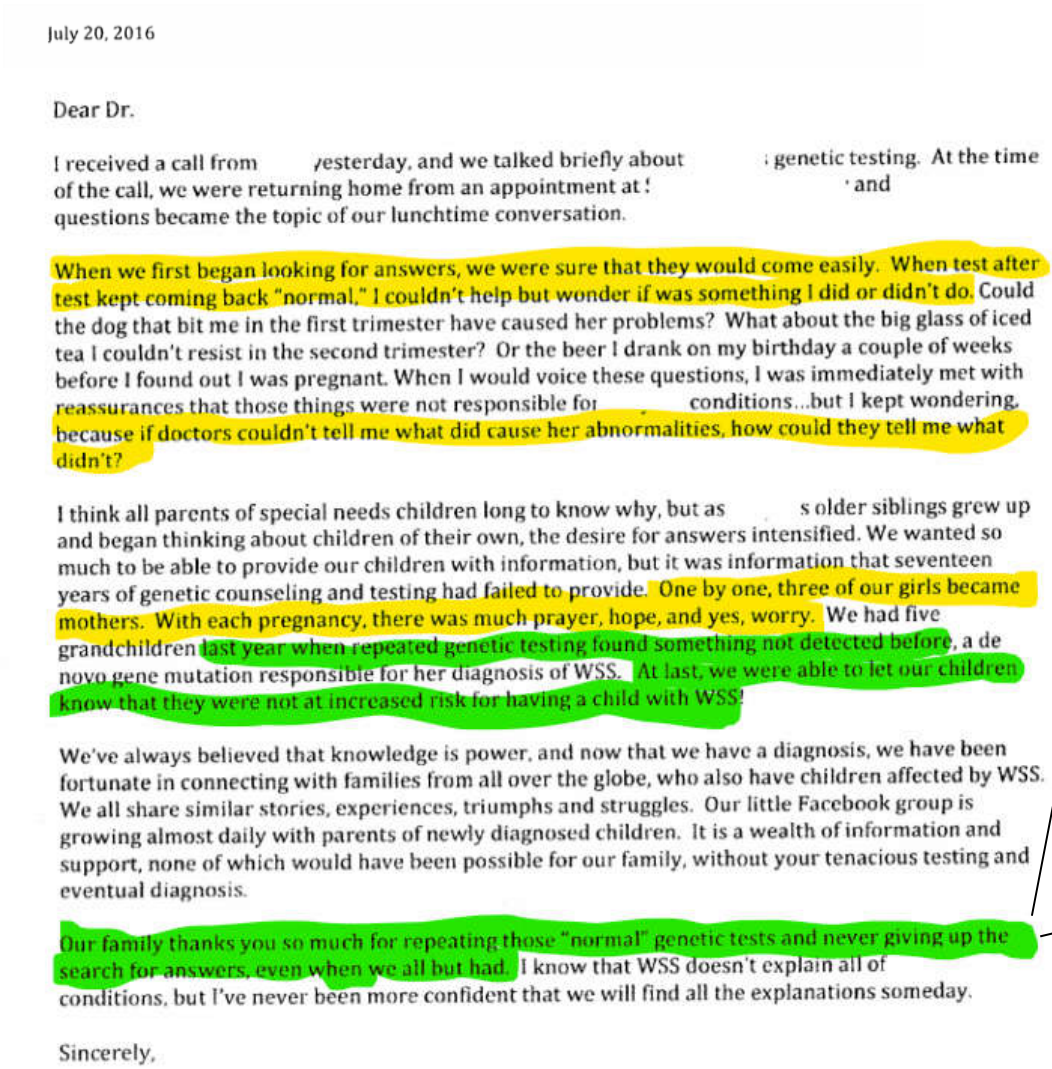
Goals

1. Diagnose (thousands of) *real* patients.

2. Understand amazing species adaptations.

3. Understand gene regulation.

We Change People's Lives



July 20, 2016

Dear Dr.

I received a call from yesterday, and we talked briefly about genetic testing. At the time of the call, we were returning home from an appointment at ! and questions became the topic of our lunchtime conversation.

When we first began looking for answers, we were sure that they would come easily. When test after test kept coming back "normal," I couldn't help but wonder if was something I did or didn't do. Could the dog that bit me in the first trimester have caused her problems? What about the big glass of iced tea I couldn't resist in the second trimester? Or the beer I drank on my birthday a couple of weeks before I found out I was pregnant. When I would voice these questions, I was immediately met with reassurances that those things were not responsible for conditions...but I kept wondering, because if doctors couldn't tell me what did cause her abnormalities, how could they tell me what didn't?

I think all parents of special needs children long to know why, but as s older siblings grew up and began thinking about children of their own, the desire for answers intensified. We wanted so much to be able to provide our children with information, but it was information that seventeen years of genetic counseling and testing had failed to provide. One by one, three of our girls became mothers. With each pregnancy, there was much prayer, hope, and yes, worry. We had five grandchildren last year when repeated genetic testing found something not detected before, a de novo gene mutation responsible for her diagnosis of WSS. At last, we were able to let our children know that they were not at increased risk for having a child with WSS!

We've always believed that knowledge is power, and now that we have a diagnosis, we have been fortunate in connecting with families from all over the globe, who also have children affected by WSS. We all share similar stories, experiences, triumphs and struggles. Our little Facebook group is growing almost daily with parents of newly diagnosed children. It is a wealth of information and support, none of which would have been possible for our family, without your tenacious testing and eventual diagnosis.

Our family thanks you so much for repeating those "normal" genetic tests and never giving up the search for answers, even when we all but had. I know that WSS doesn't explain all of conditions, but I've never been more confident that we will find all the explanations someday.

Sincerely,

Our family thanks you so much for repeating those "normal" genetics tests and never giving up the search for answers, even when we all but had.

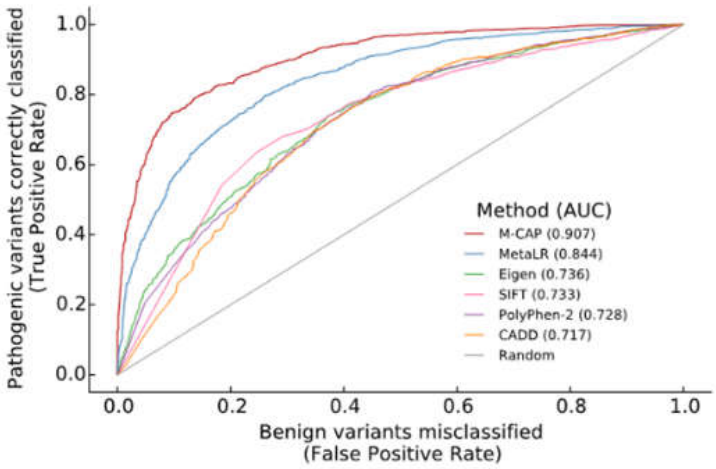
How?

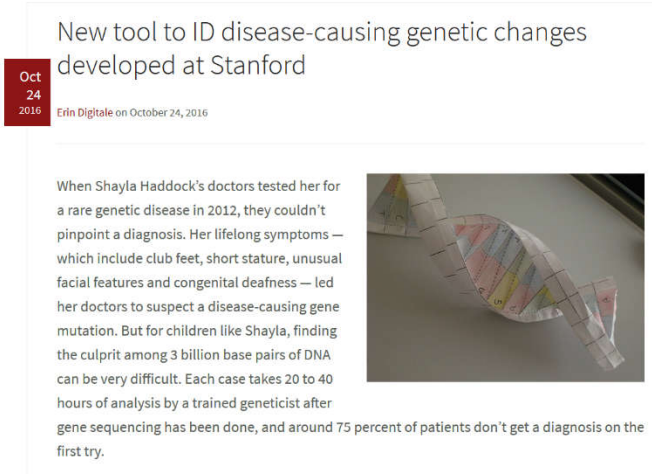
We Use Machine Learning

nature genetics

M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity

Karthik A Jagadeesh[1,5], Aaron M Wenger[2,5], Mark J Berger[1], Harendra Guturu[2], Peter D Stenson[3], David N Cooper[3], Jonathan A Bernstein[2] & Gill Bejerano[1,2,4]





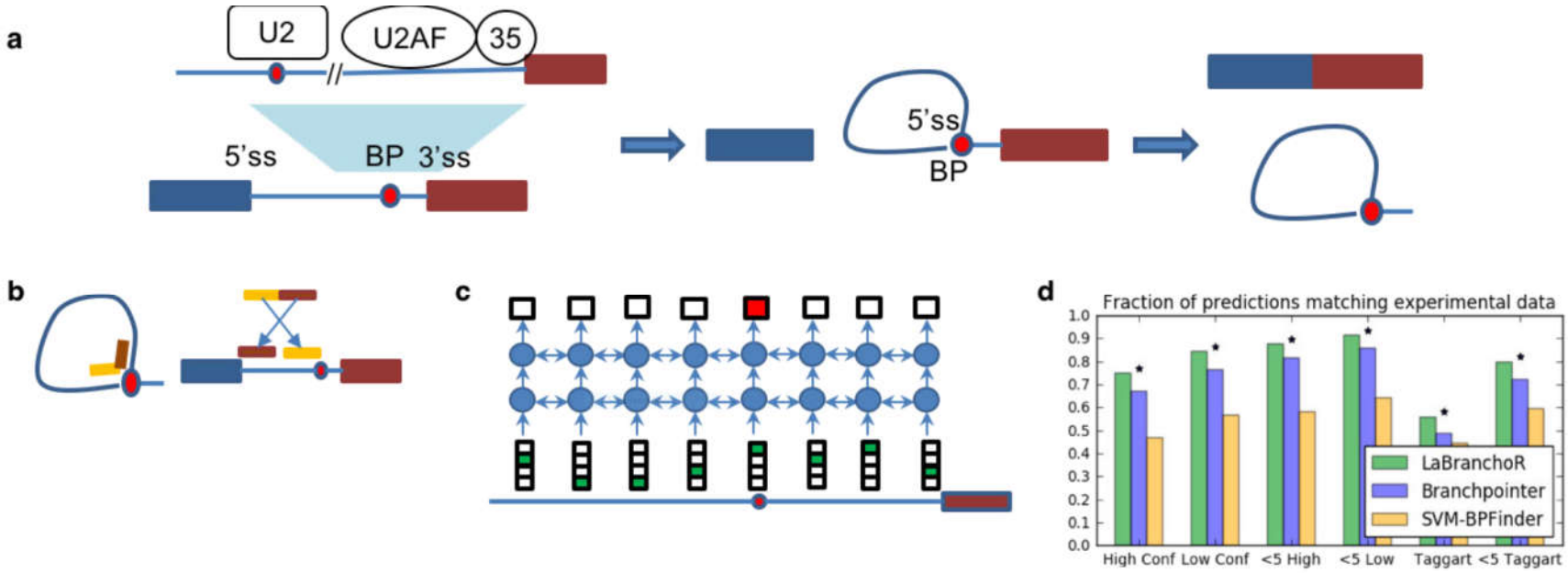
Oct 24 2016

New tool to ID disease-causing genetic changes developed at Stanford

Erin Digitale on October 24, 2016

When Shayla Haddock's doctors tested her for a rare genetic disease in 2012, they couldn't pinpoint a diagnosis. Her lifelong symptoms — which include club feet, short stature, unusual facial features and congenital deafness — led her doctors to suspect a disease-causing gene mutation. But for children like Shayla, finding the culprit among 3 billion base pairs of DNA can be very difficult. Each case takes 20 to 40 hours of analysis by a trained geneticist after gene sequencing has been done, and around 75 percent of patients don't get a diagnosis on the first try.

We Do Deep Learning

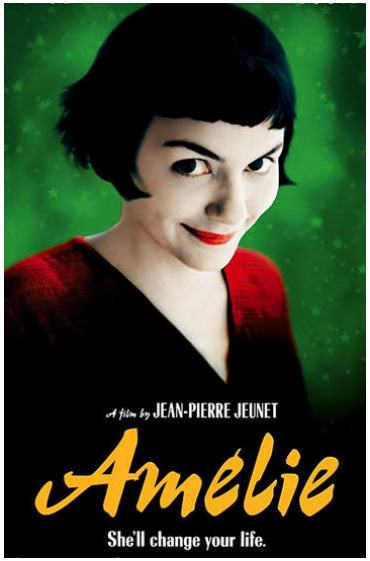


Natural Language Processing

Check out http://amelie.stanford.edu/

- Parse 27 million documents:

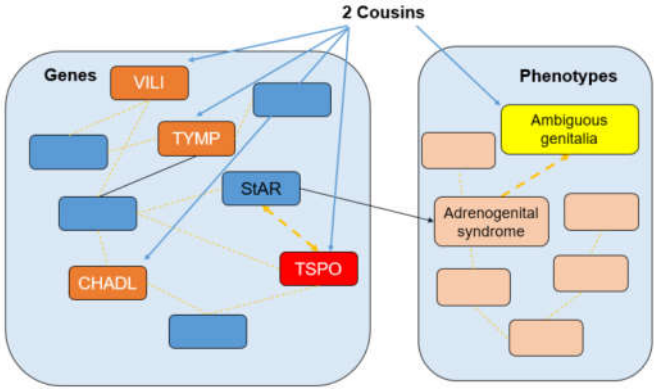
> Mutations affecting the SAND domain of **DEAF1** cause **intellectual disability** with severe **speech impairment** and **behavioral problems**.

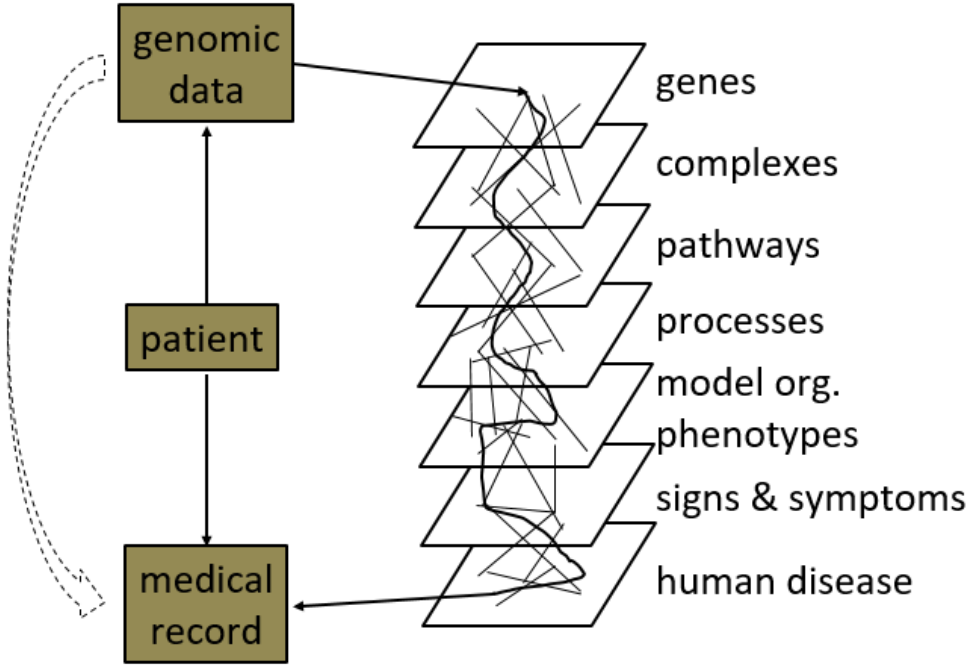


- Patient: **intellectual disability** with **impaired speech development** and **aggressive behavior**

AC018470.1, ACAP3, ADAP1, AMPD1, ASPM, ASXL2, BAZ1B, BHLHE22, BTBD9, C17orf104, C17orf74, C19orf26, C1orf87, C2orf81, CCNL2, CDH10, CHD6, CNOT3, COL6A5, DCHS2, **DEAF1**, DNM1, FAM216B, FAM73B, FAM83H, FAM84B, FAT3, FBXO25, FCRLB, FLJ00104, FRS2, GRK7, HEPHL1, HOXD11, IL19, INSRR, IQCC, KIAA0825, LAMA5, LAMC3, LGR6, MAST4, MBD6, MBLAC2, MCM10, MDH2, METRN, MSL2, N4BP3, NCKAP5, NUP50, NYNRIN, ORC3, PDCD2L, PDXP, PLEKHG1, PLIN2, POU3F2, PXMP2, RAB11FIP1, RASSF1, RIMS1, RTKN2, SASS6, SERPINA3, SH3BP1, SHB, SLC2A9, SLC38A8, SON, SP8, SPTBN5, SRRM2, TAAR1, TARSL2, TET2, TRIM72, TSPAN15, TSPYL4, WDR20, XPNPEP1, ZFYVE16, ZNF469, ZSCAN29

Network Analysis





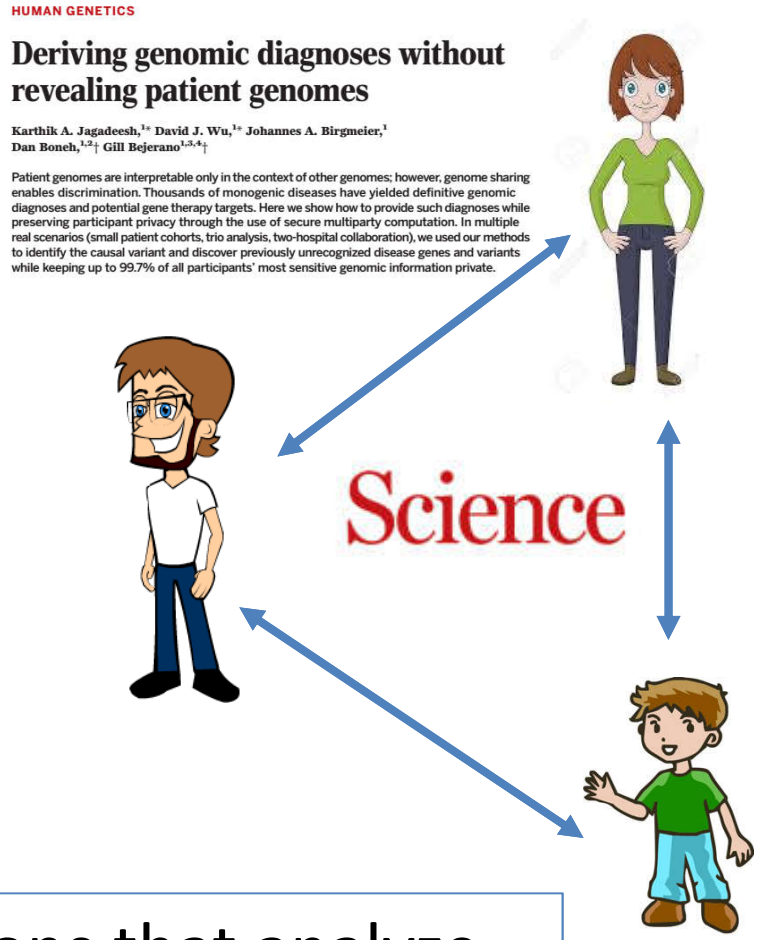
Even Cryptography

The problem:

- Our genomes are best understood in light of each other.
- But our genomes tell so much about us , sharing may lead to discrimination.

Goal:

Find ways to share genomes without revealing genomes

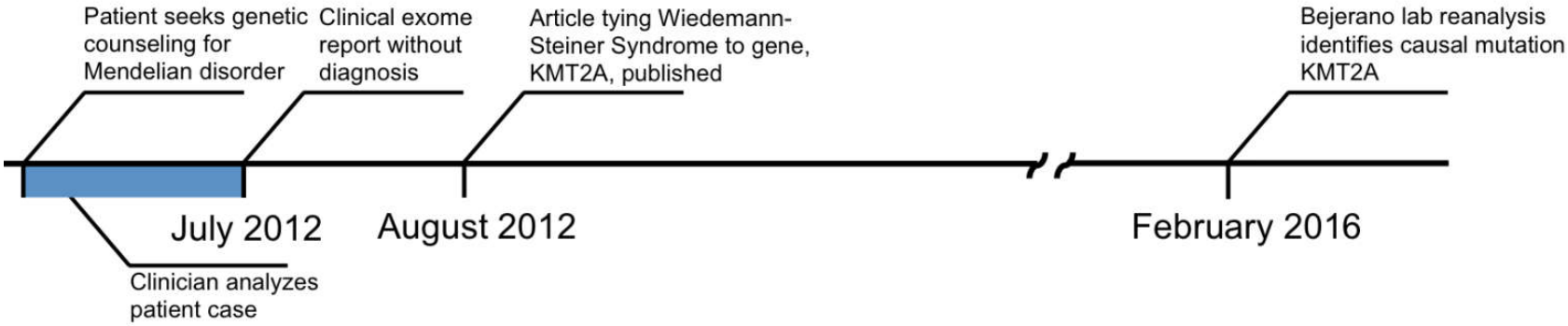


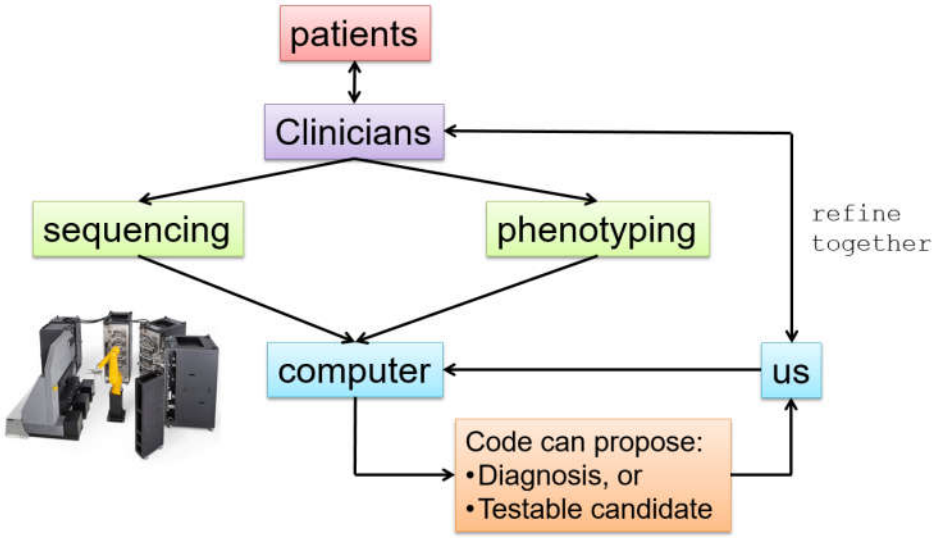
Design distributed computations that analyze, while revealing nothing else about genomes

HCI Opportunities

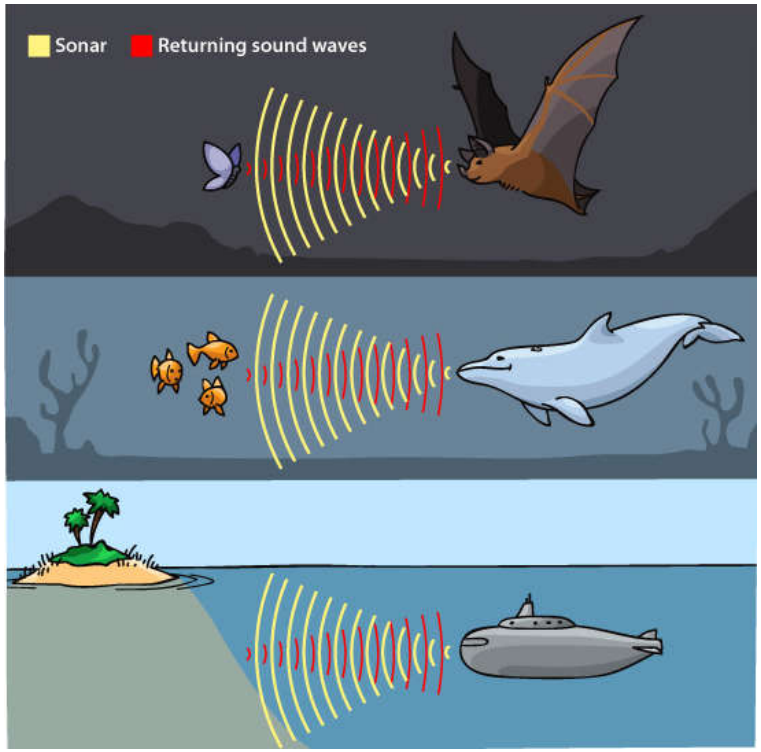
¾ million job submissions. >1,000 citations.

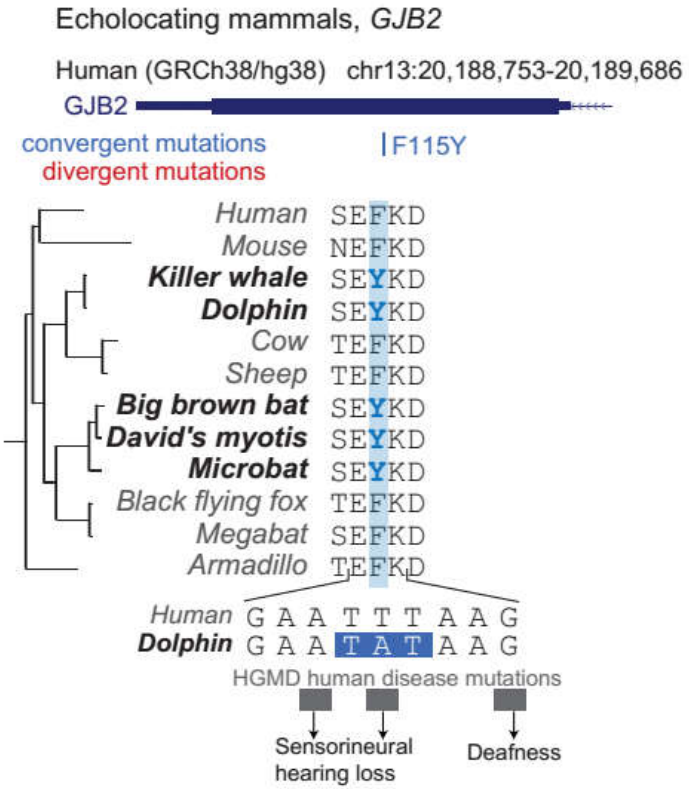
Build Perpetual Systems





Genomics Too!





LETTER

nature
International weekly journal of science

Human-specific loss of regulatory DNA and the evolution of human-specific traits

Collaborate

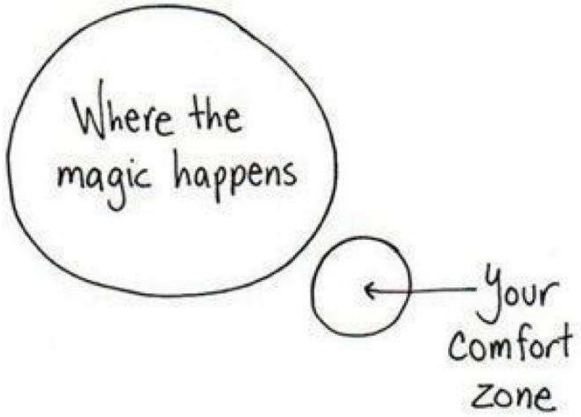
We have wonderful collaborators in

- Computer Science
- Electrical Engineering

- Evo Devo
- Functional Genomics
- Neurobiology

- Medical Genetics
- Gene Therapy

Interested?



Reach out for a chat.

Plenty of ways to get started.

Weekly open office hours
(posted on lab website)

OR

bejerano@stanford.edu

Questions, please?