Next-generation human genomics

Tim Yu, MD, PhD

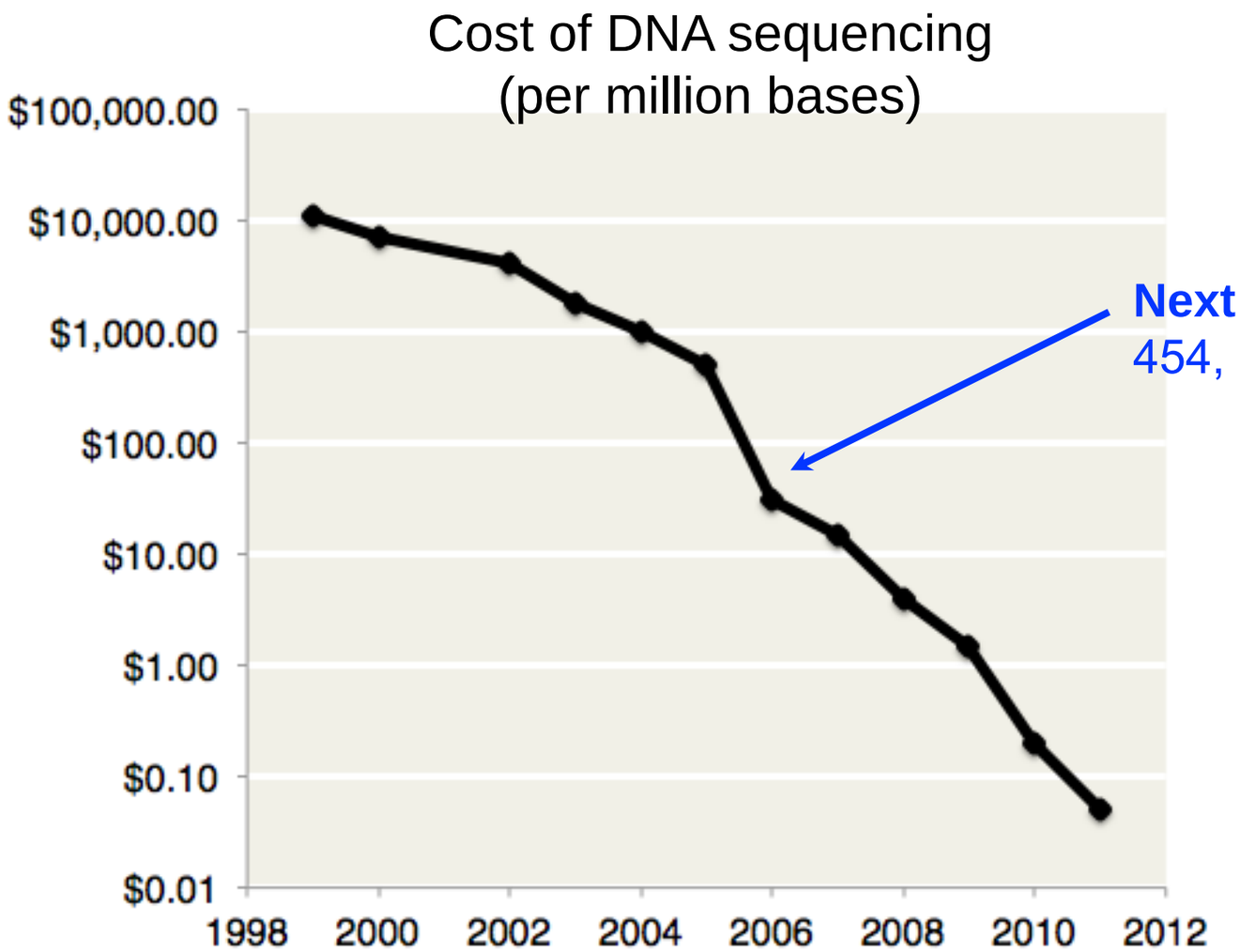
Division of Genetics & Genomics, Boston Children's Hospital
Claritas Genomics ($, stock)
Dept of Neurology, MGH
Harvard Medical School & the Broad Institute

The personalized genome

- Amy, age 21 years, visits with her physician and elects to have **complete genome sequencing**.
- At a follow-up visit, Amy chooses to learn of her genetic risk factors for heart disease, diabetes, breast cancer, and colon cancer. Amy's physician provides her with risk scores for those disorders, and with suggestions for lifestyle modifications. Specifically, Amy is alerted to her particularly **high risk of developing type 2 diabetes, and her physician recommends a rigorous program of diet and exercise** that had been shown in a controlled study to delay or prevent disease onset.
- The next year, Amy develops mild asthma and **her physician selects an optimal therapy based on Amy's genetic profile**.
- Five years later, Amy informs her physician that she and her husband are planning to start a family, and they request information regarding the risk of having a child affected by a serious genetic disease, based on their genome sequence data. She learns that both she and her husband are **carriers for the recessive lethal childhood disorder spinal muscular atrophy, and they seek further counseling**.
- When Amy turns 40, she begins **colorectal cancer screening** based on her higher-than-average risk factors, and at age 45 a precancerous polyp is detected in her colon and is successfully removed.

Predictions for 2020
Feero, Guttmacher, and Collins, *JAMA* 2008

Sequencing costs



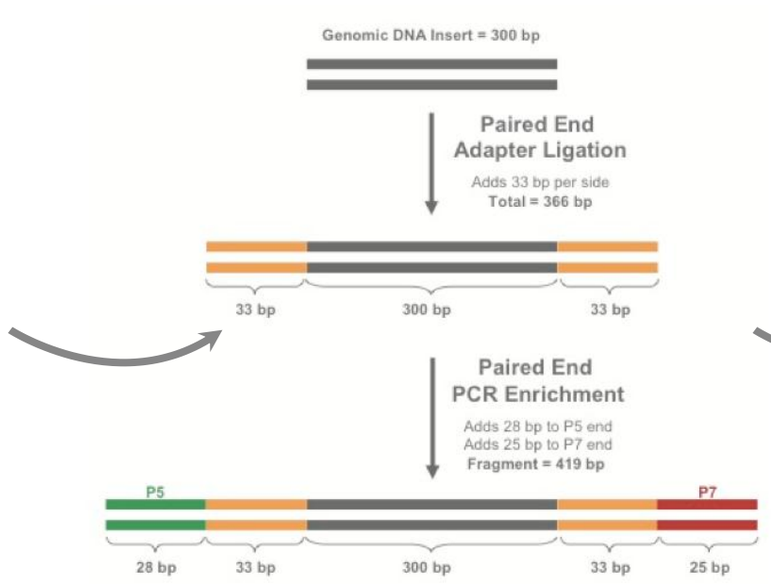
NCBI

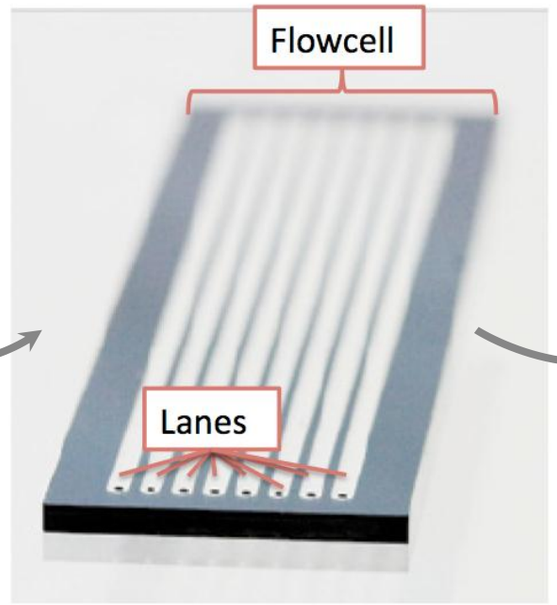
Cost of the first human genome (1990-2003):	$2.7 Billion
Cost of a human genome in 2009	$20,000
Cost of a human genome in 2013	$2,500
Cost of a human “exome” in 2013	$800 (< brain MRI)

Advances in the last 5 years have made it easy to generate whole genome sequencing data.

The challenge is interpretation.

Next-generation sequencing





ATGTACCAGAGGATGC
TCCGGTGCGGC…

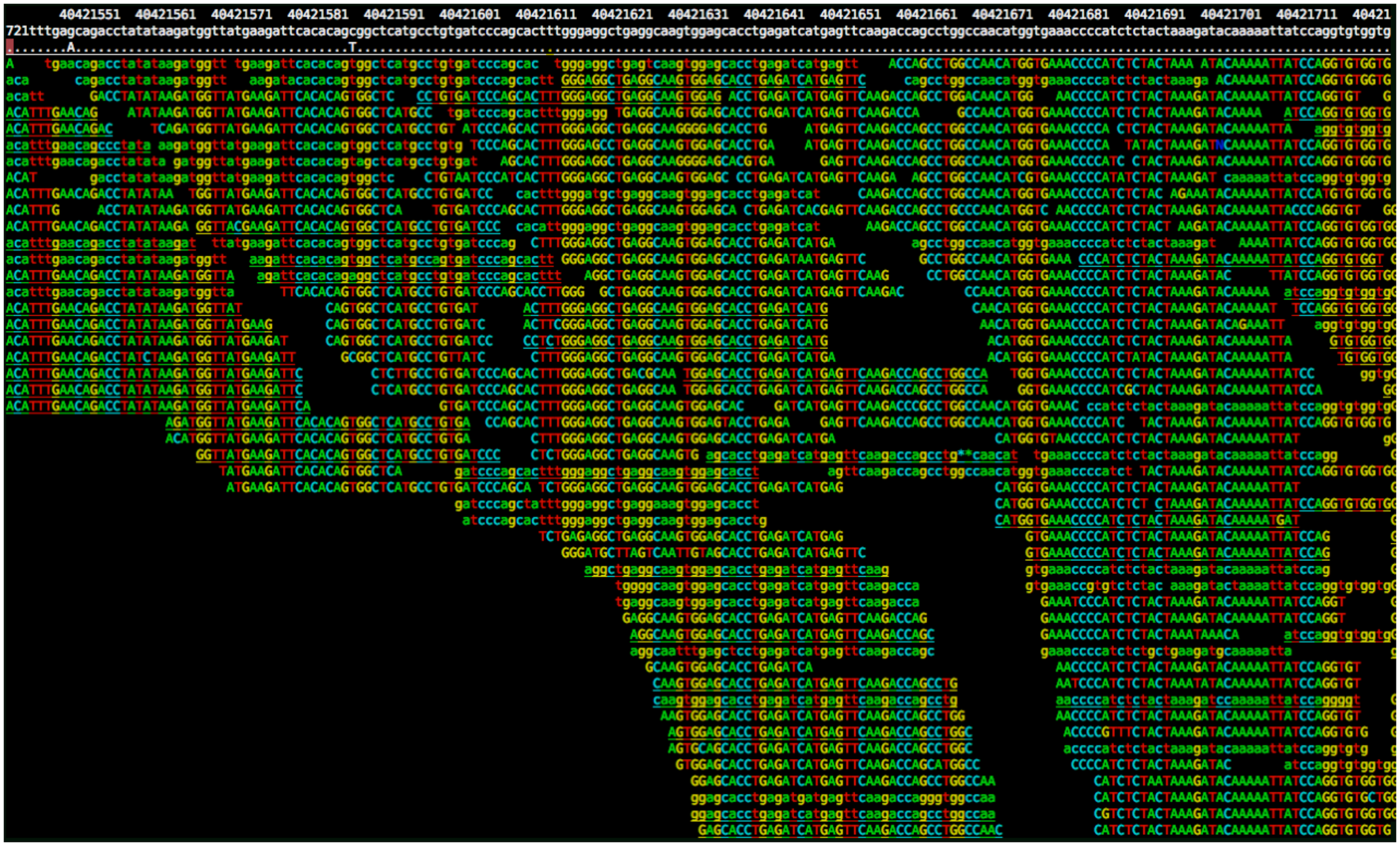
Blood samples

DNA Libraries

Flowcells

Sequenc
e

Millions of reads are mapped en masse to a reference genome

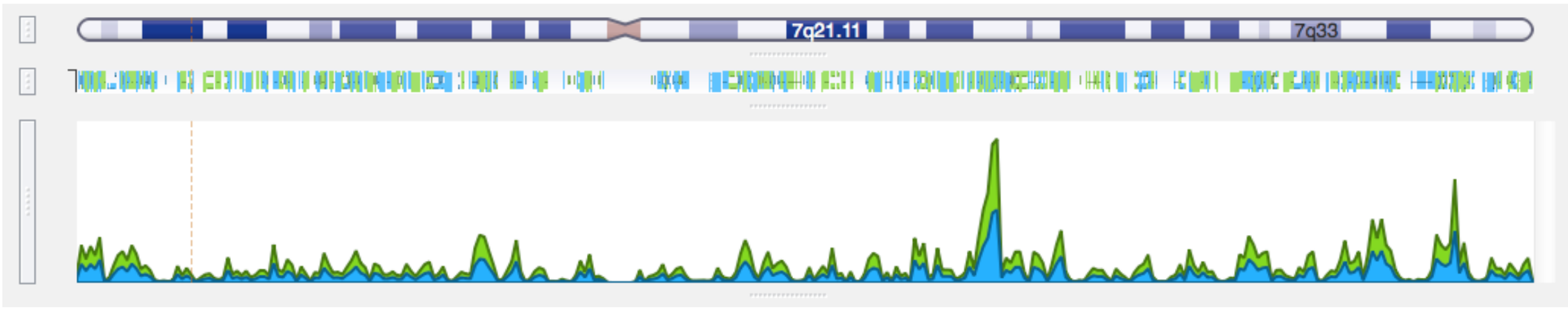


Reads can effectively cover 95% of the genome or exome

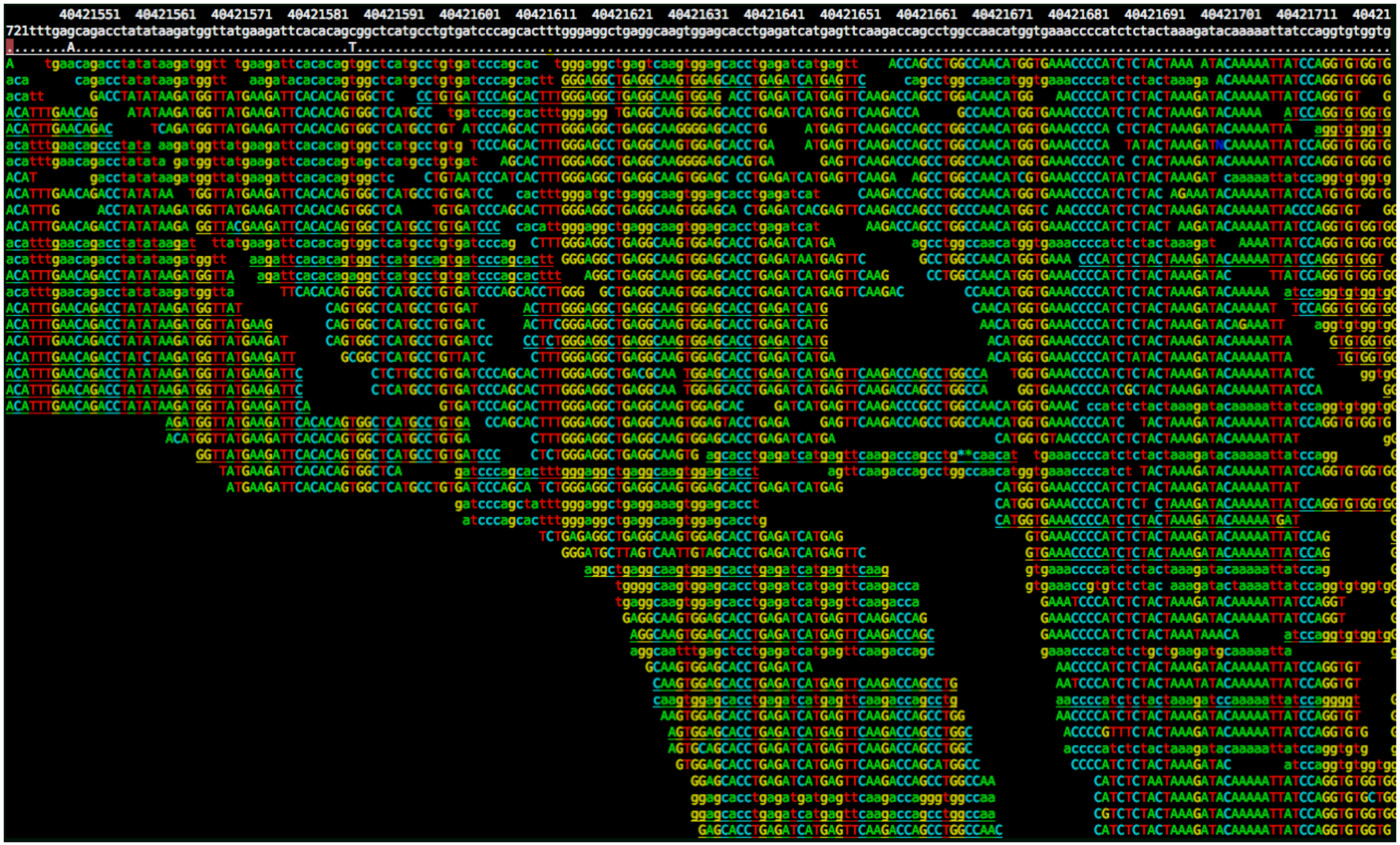
Coverage of chr7 in a typical
whole exome sequencing experiment

Genes

Read depth



Variants are detected when enough reads disagree with reference



Variants are detected when enough reads disagree with reference

Variants are detected when enough reads disagree with reference

The human genome is big

3 billion basepairs

X 0.1%

---

3 million sites of variation
between any two individuals

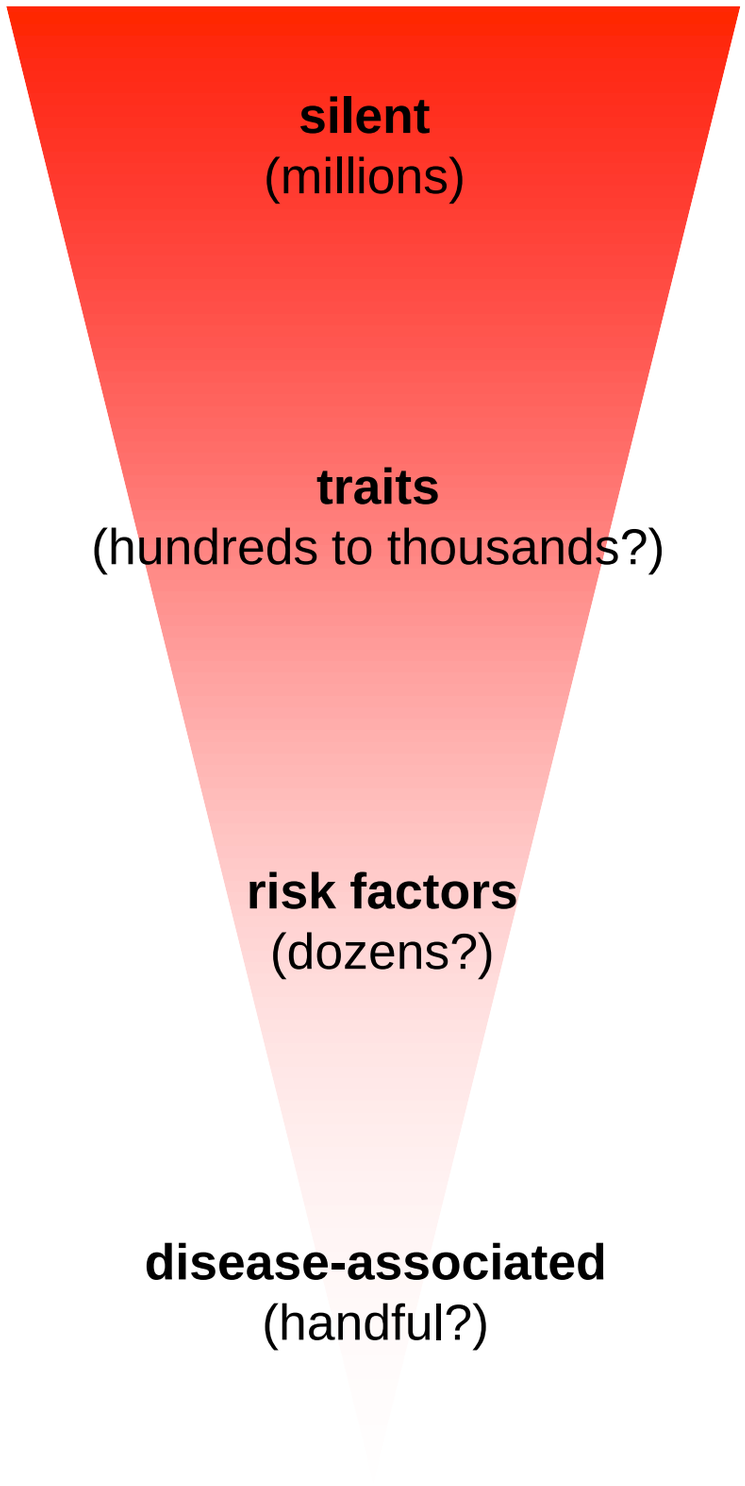
A parts list of variants from one individual

Table 2. SNPs Identified through Whole-Genome Sequencing of DNA from the Proband.*

SNP Type	No. of SNPs
Nongene	2,255,102
Gene	1,165,204
Intron	1,064,655
Promoter	60,075
3′ UTR	16,350
5′ UTR	3,517
Splice regulatory site	2,089
Splice site	112
Synonymous	9,337
Stop→stop	17
Nonsynonymous	9,069
Stop→gain	121
Stop→loss	27
Total	3,420,306

Lupski et al, *NEJM* 2010

Of these 3 million, which are medically relevant?



How do you interpret 3,000,000 variants?

3,000,000

Commonly encountered?

5-10% of variants have not been seen frequently in controls

In a gene coding region?

0.25% of rare variants lie within a gene coding region

Alter protein (or affect splicing)?

~50% of variants in genes alter protein or affect splicing

Affect a known disease gene?

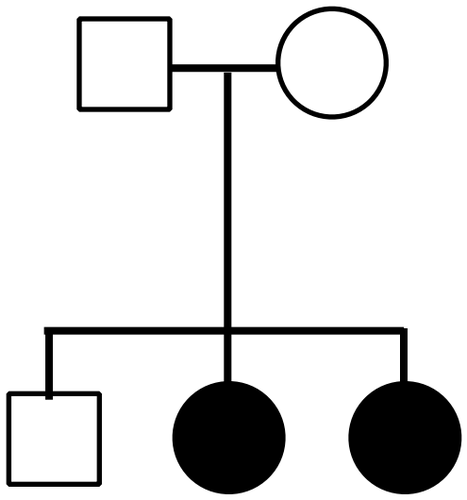
~10% of genes have been associated with human disease

~20-40

WGS has revolutionized Mendelian genetics

~20-40

Inheritance pattern



Sequencing parents & available siblings

Single disease genes!

Three siblings w/intellectual disability, cataracts and seizures

3,655,904 variants

9,521 protein-altering

*MAP3K5,**
*LFDH,**
*NHSL1,**
PEX7

935 rare and protein-altering

Only 4 shared by all affected children in the family

Rhizomelic chondrodysplasia punctata

Challenges to clinical adoption

- A “data deluge”
- Many results are of uncertain medical significance
- Insufficient numbers of geneticists and genetic counselors to handle the flood of clinical data
- Clinicians and patients will need education and training
- Collaborative efforts will be required to amass and organize data
- Birth of a new specialty?

Challenges to clinical adoption

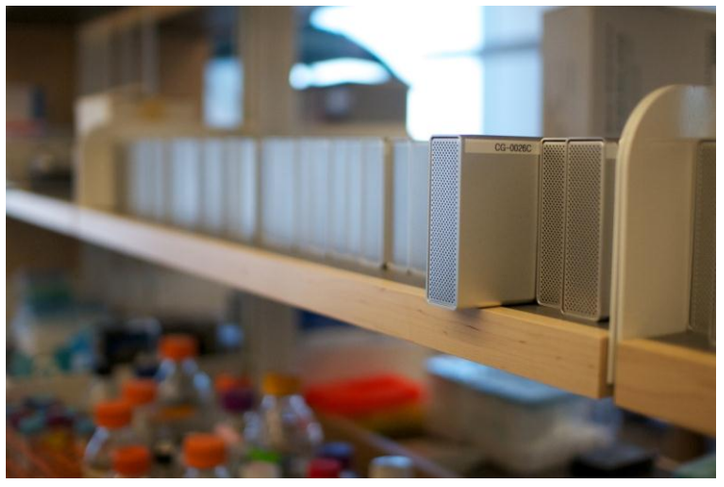
- Genomic medicine requires cross-disciplinary skills
  - Genetics
  - Genomics
  - Lab Medicine
  - Informatics

- Example:
  - Processing time for 1 human genome: **one week**
  - Requires highly specialized computer systems and expertise that **stresses most hospital infrastructures**

Challenges to clinical adoption

- 3 billion bases & 3 million variants **don't fit in a patient chart**
- 1 patient's whole genome data = **1 terabyte hard drive**
- Costs of secure storage!
- **Electronic Medical Records** are not yet equipped to handle this

40 patient whole genomes, on a shelf



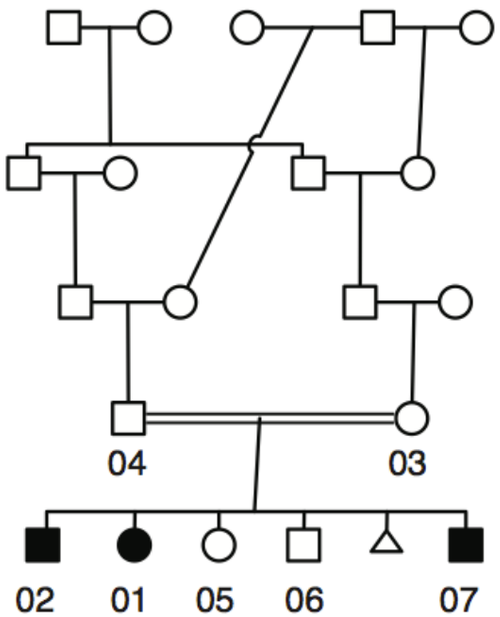
Ethical, legal, & social issues

- Privacy and confidentiality
- Stigmatization, and discrimination
- Results of uncertain medical significance, emotional distress, follow-up tests, and cost
- Genetic paternalism vs. right to control your genomic information

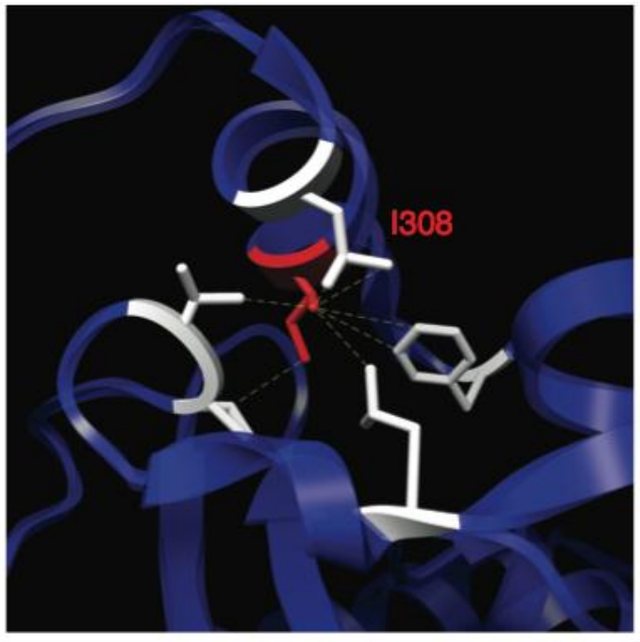
Standards of evidence

- Need to be careful applying what we think we know
- Many currently reported disease associations are based on old studies of with small sample sizes, or using outdated technologies
- These will need to be re-evaluated as we get better at interpreting genetic results
- We also need to be cost-effective, and demonstrate that this leads to better care

The importance of functional follow-up

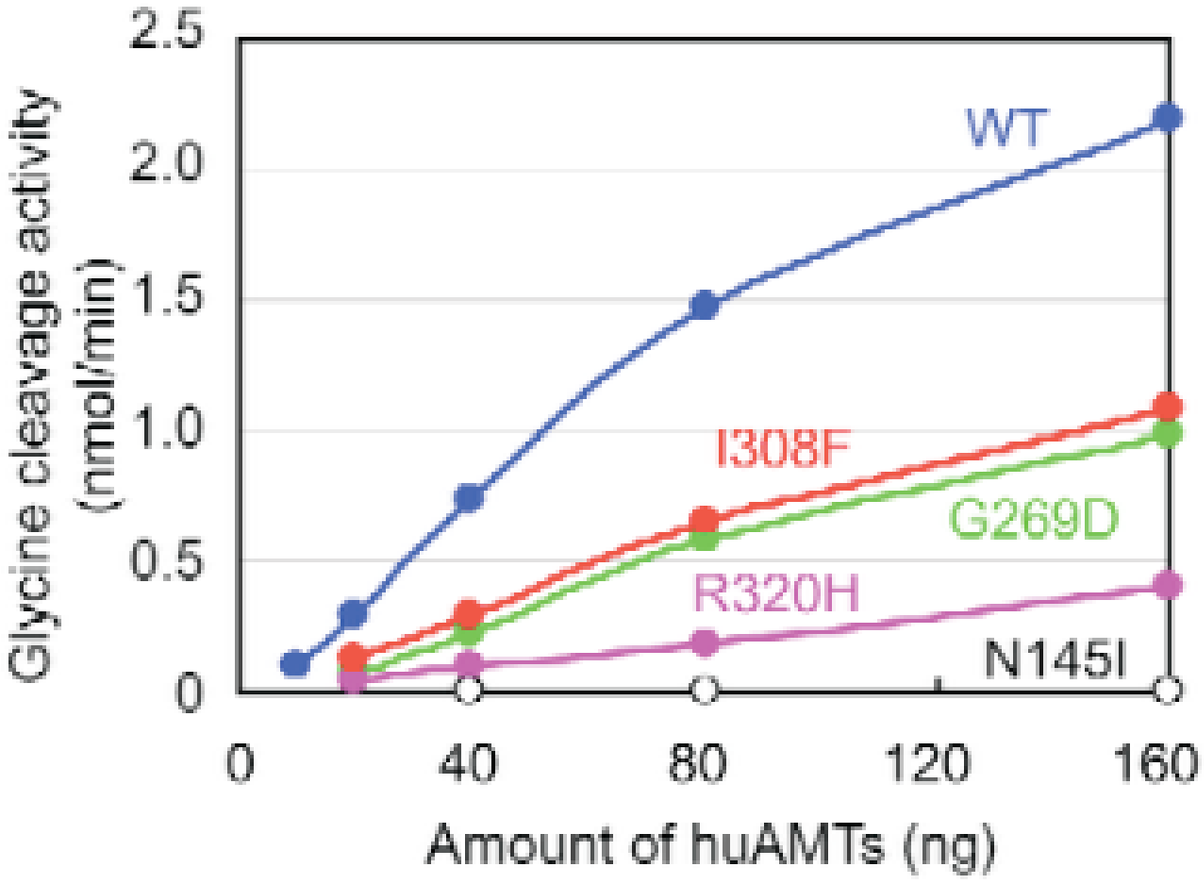


- ✓ Family with three children with autism, intellectual disability, and seizures
- ✓ WES performed in 3 children + parents
- ✓ Homozygous I308F mutation in **AMT**, a cause of **glycine encephalopathy**



- ✓ Rare (never before seen in controls)
- ✓ Alters a highly conserved residue, “looks deleterious”
- ✓ But patients with glycine encephalopathy often die in the neonatal period

AMT p.I308F is a hypomorphic LOF mutation



- **I308F: defects in protein folding and enzyme function, but retains some residual activity**

✓ These children had a **mild version** of glycine encephalopathy with autism and epilepsy

✓ Had been **previously undiagnosed** in the family (requires lumbar puncture, liver biopsy)

[T. Yu, M. Chahrour, et al, Neuron 2013]

Grappling with “incidental findings”

The Incidentalome
A Threat to Genomic Medicine

Isaac S. Kohane, MD, PhD
Daniel R. Masys, MD
Russ B. Altman, MD, PhD

GENOMIC MEDICINE IS POISED TO OFFER A BROAD ARray of new genome-scale screening tests. However, these tests may lead to a phenomenon in which multiple abnormal genomic findings are discovered, analogous to the “incidentalomas” that are often discovered in radiological studies. If practitioners pursue these unexpected genomic findings without thought, there may be disastrous consequences. First, physicians will

There is a rich literature in radiology on the “incidentaloma,” which is a finding (most commonly a mass) found on computed tomography or magnetic resonance imaging studies ordered for symptoms or concerns totally unrelated to the gland in which the mass is found. The workup of an incidentaloma is complicated by concerns that it may be associated with malignant disease and, at least initially, the lack of good data on the prevalence of malignant disease in the general population. Incidentalomas occur because imaging modes do not only report on the areas of direct clinical concern but, incidentally, report on all organs in the field of view.[1]

This phenomenon of possible incidental genomic find-

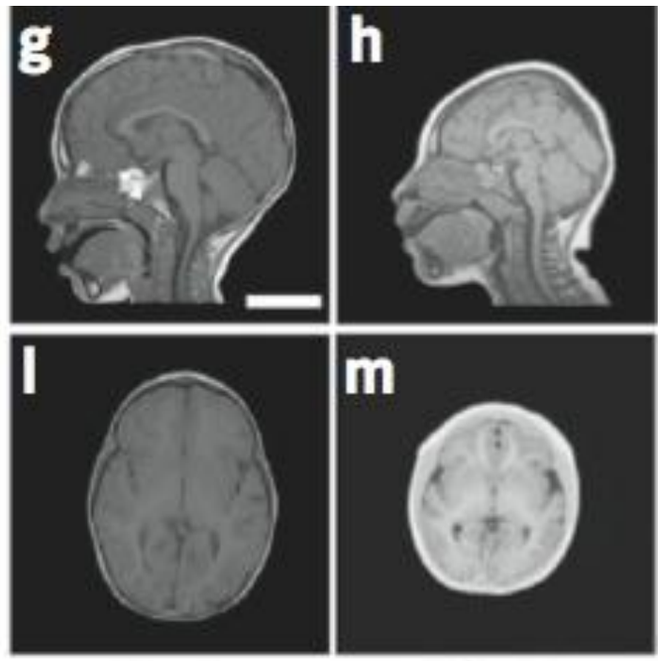
Other open issues in genomic medicine

- Primary vs. Secondary Findings
- Genes of Uncertain Significance
- Research vs. Clinical boundaries
- Stretching our definitions of disease
- Collaborative resources and standards

If you're interested in this, you're in a pretty good place

Mutations in *WDR62*, encoding a centrosome-associated protein, cause microcephaly with simplified gyri and abnormal cortical architecture

Timothy W Yu[1–7], Ganeshwaran H Mochida[1–7], David J Tischfield[1–5], Sema K Sgaier[1–5,8], Laura Flores-Sarnat[9], Consolato M Sergi[10,11], Meral Topçu[12], Marie T McDonald[13], Brenda J Barry[1–5], Jillian M Felie[1–5], Christine Sunu[1–5], William B Dobyns[14], Rebecca D Folkerth[15], A James Barkovich[16] & Christopher A Walsh[1–6]



One of the first demonstrations of using NGS in humans
Identification of a new gene for human microcephaly from 6 families

[Yu et al, *Nat Genetics*, 2010]

If you're interested in this, you're in a pretty good place

Genetic Defect in *CYP24A1*, the Vitamin D 24-Hydroxylase Gene, in a Patient with Severe Infantile Hypercalcemia

Andrew Dauber, Thutrang T. Nguyen, Etienne Sochett, David E. C. Cole, Ronald Horst, Steven A. Abrams, Thomas O. Carpenter, and Joel N. Hirschhorn

Novel Microcephalic Primordial Dwarfism Disorder Associated with Variants in the Centrosomal Protein Ninein

Andrew Dauber, Stephen H. LaFranchi, Zoltan Maliga, Julian C. Lui, Jennifer E. Moon, Cailin McDeed, Katrin Henke, Jonathan Zonana, Garrett A. Kingman, Tune H. Pers, Jeffrey Baron, Ron G. Rosenfeld, Joel N. Hirschhorn, Matthew P. Harris, and Vivian Hwa

Whole exome sequencing of children with endocrine disorders: new genes for pediatric hypercalcemia and dwarfism

[Dauber et al, *JCEM* 2012]

If you're interested in this, you're in a pretty good place

REPORT

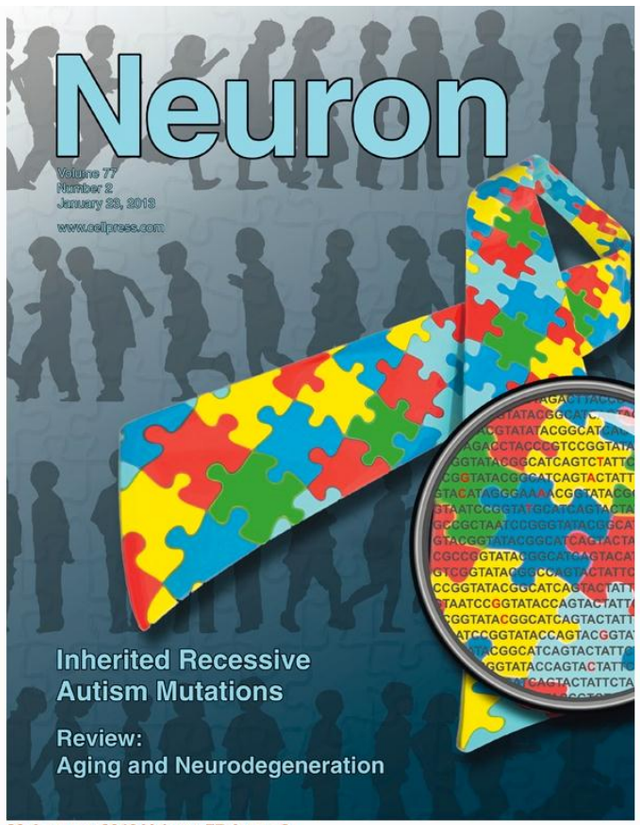
Exome Sequencing and Functional Validation in Zebrafish Identify *GTDC2* Mutations as a Cause of Walker-Warburg Syndrome

M. Chiara Manzini,[1,2] Dimira E. Tambunan,[1,2] R. Sean Hill,[1,2] Tim W. Yu,[1,2] Thomas M. Maynard,[3] Erin L. Heinzen,[4] Kevin V. Shianna,[4] Christine R. Stevens,[5] Jennifer N. Partlow,[1,2] Brenda J. Barry,[1,2] Jacqueline Rodriguez,[1,2] Vandana A. Gupta,[1,6] Abdel-Karim Al-Qudah,[7] Wafaa M. Eyaid,[8] Jan M. Friedman,[9,10] Mustafa A. Salih,[11] Robin Clark,[12] Isabella Moroni,[13] Marina Mora,[14] Alan H. Beggs,[1,6] Stacey B. Gabriel,[5] and Christopher A. Walsh[1,2,5,*]

Larger cohorts:
WES in 19 families identifies a novel gene responsible for a well-known neuromuscular disorder

[Manzini et al, *AJHG*, 2012]

If you're interested in this, you're in a pretty good place



Neuron

Article

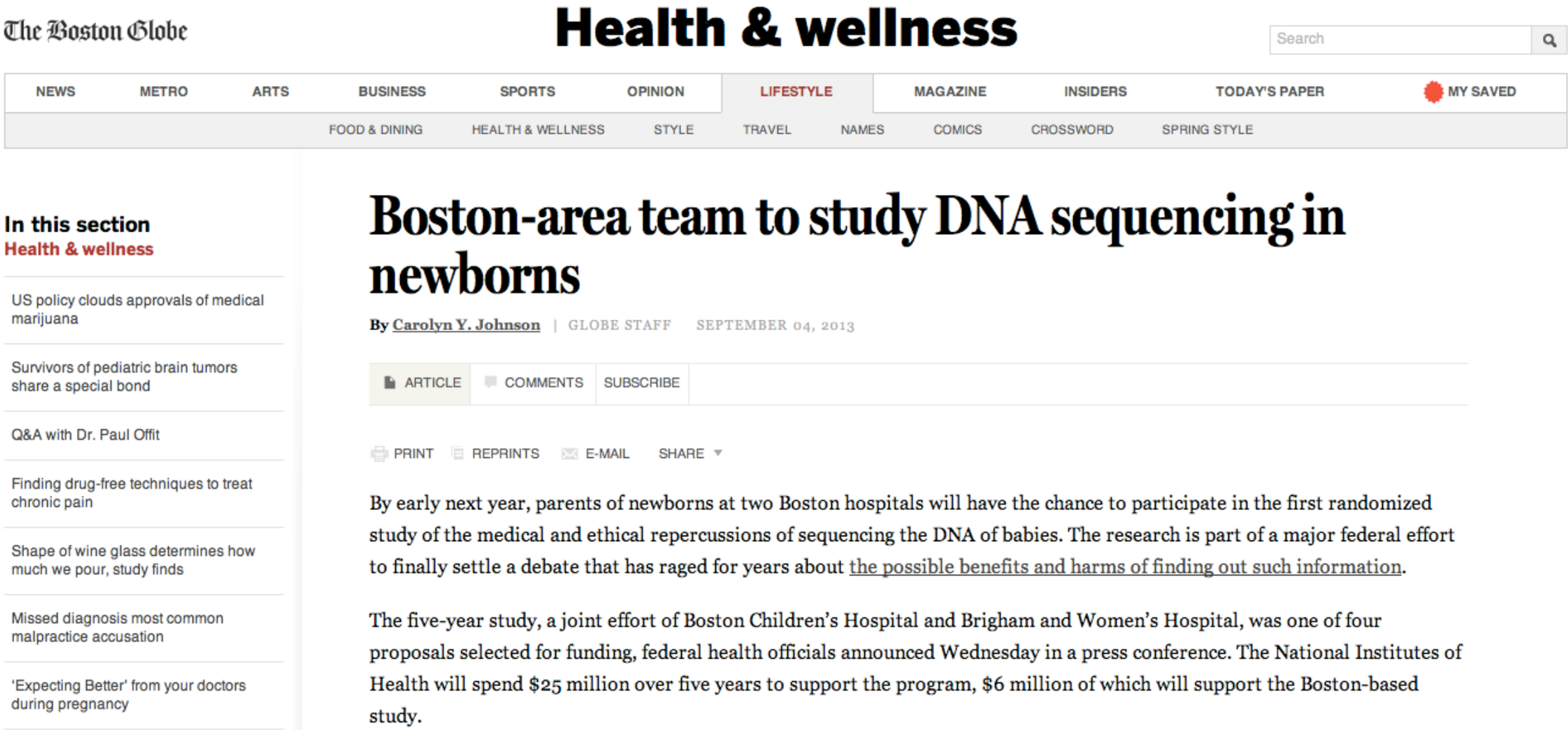
Using Whole-Exome Sequencing to Identify Inherited Causes of Autism

Timothy W. Yu,[1,2,3,4,5,6,7,32,*] Maria H. Chahrour,[1,2,3,4,5,7,32] Michael E. Coulter,[1,2,3,5] Sarn Jiralerspong,[8] Kazuko Okamura-Ikeda,[9] Bulent Ataman,[10] Klaus Schmitz-Abe,[1,2,5] David A. Harmin,[10] Mazhar Adli,[11] Athar N. Malik,[10] Alissa M. D'Gama,[5] Elaine T. Lim,[12] Stephan J. Sanders,[13] Ganesh H. Mochida,[1,2,3,5,6] Jennifer N. Partlow,[1,2,3] Christine M. Sunu,[1,2,3] Jillian M. Felie,[1,2,3] Jacqueline Rodriguez,[1,2,3] Ramzi H. Nasir,[5,14] Janice Ware,[5,14] Robert M. Joseph,[4,15] R. Sean Hill,[1,2,3,5] Benjamin Y. Kwan,[16] Muna Al-Saffar,[1,2,17] Nahit M. Mukaddes,[18] Asif Hashmi,[19] Soher Balkhy,[20] Generoso G. Gascon,[6,18,21] Fuki M. Hisama,[22] Elaine LeClair,[5,14] Annapurna Poduri,[5,23] Ozgur Oner,[24] Samira Al-Saad,[25] Sadika A. Al-Awadi,[26] Laila Bastaki,[26] Tawfeg Ben-Omran,[27,28] Ahmad S. Teebi,[27,28] Lihadh Al-Gazali,[17] Valsamma Eapen,[29] Christine R. Stevens,[7] Leonard Rappaport,[4,5,14] Stacey B. Gabriel,[7] Kyriacos Markianos,[1,2,5] Matthew W. State,[13] Michael E. Greenberg,[10] Hisaaki Taniguchi,[9] Nancy E. Braverman,[8] Eric M. Morrow,[4,30,31] and Christopher A. Walsh[1,2,3,4,5,7,*]

Larger cohorts:
WES in >150 consanguineous families
to find new recessive autism genes

[Yu et al, *Neuron* 2013]

If you're interested in this, you're in a pretty good place



BabySeq:

WES and WGS on 120 normal newborns and 120 NICU patients

-> Clinical outcomes, healthcare utilization, and safety

If you're interested in this, you're in a pretty good place

What is Claritas Genomics?

CLARITAS GENOMICS

A Diagnostic Testing Company

Arose out of Boston Children's Hospital

- CLIA licensed molecular lab for genomic medicine
- Specialty testing based on Boston Children's research and clinical knowledge

Partnership with major pediatric hospitals in the US

Partnering with industry (Life Technologies, Cerner)

Partnering with country health systems (Saudi Genome Project, US Million Veterans Project)

Starting an Interpretive Genomics Service at Boston Children's Hospital


40421571 40421581 40421591 40421601
ttatgaagattcacacagcggctcatgcctgtgatccca

Questions!