UNIVERSITY OF UTAH
HEALTH SCIENCES

What's New in Genomics?

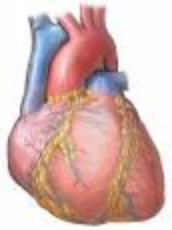
Lynn Jorde
Mark and Kathie Miller Presidential Endowed Professor
Chair, Department of Human Genetics
Executive Director, Utah Genome Project

May 16, 2018

DNA: Our body's instruction manual

DNA sequencing reveals a person's instruction manual (the **genome**, 3 billion pairs of DNA bases).

Each of us has a **unique instruction manual**.



Two very different instruction manuals

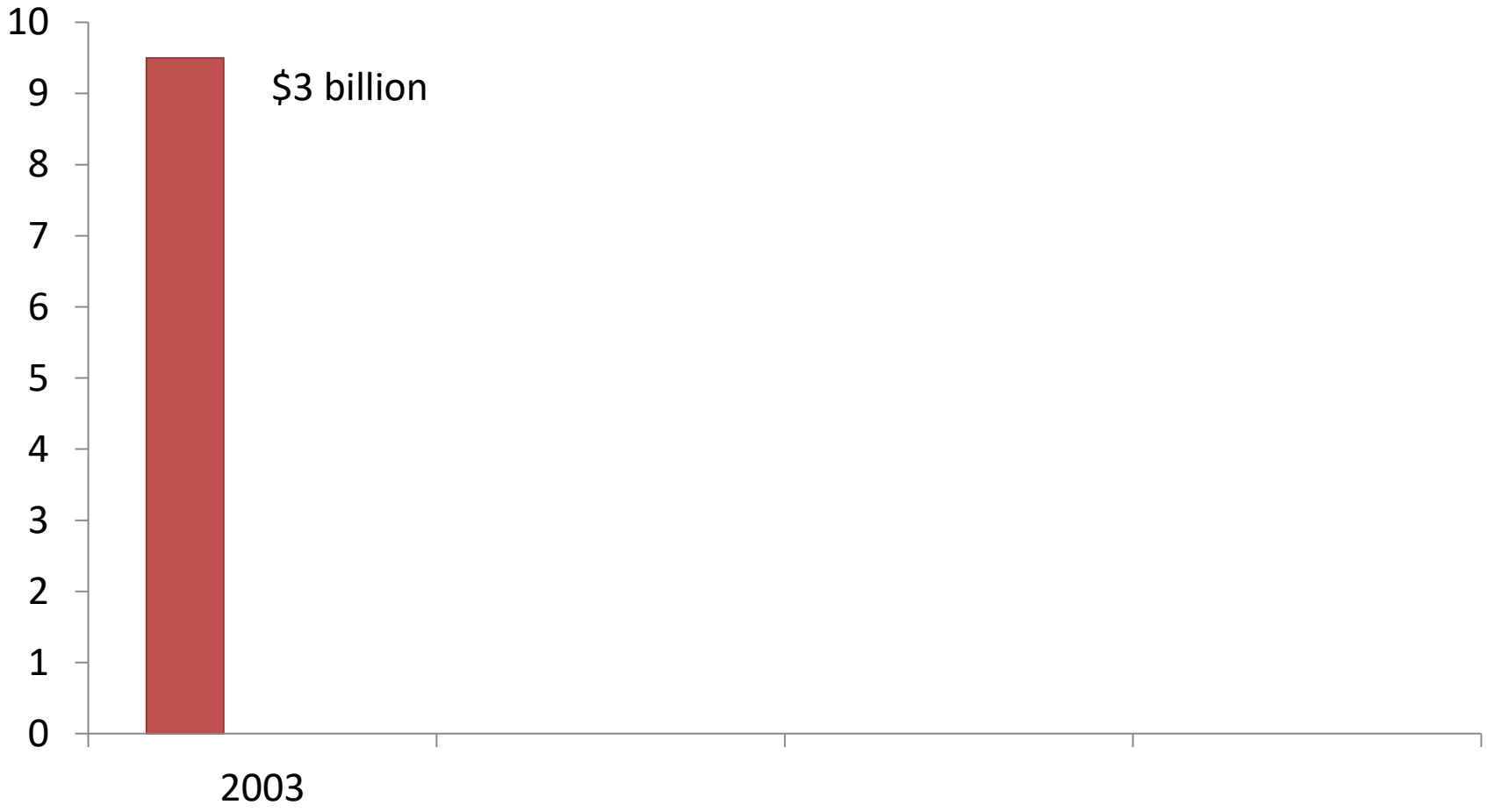
Jim Fixx, 1932-1984

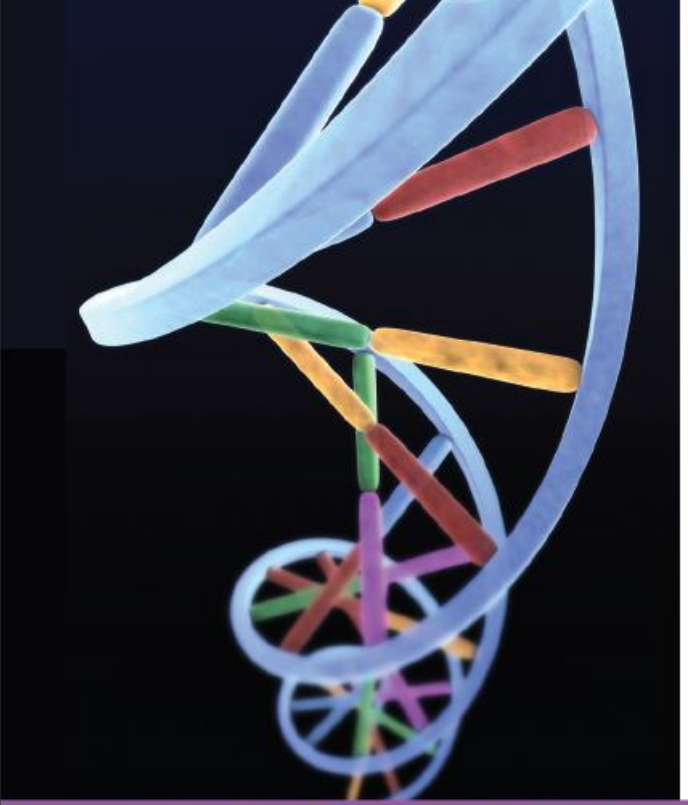
"The Complete Book of Running"

Winston Churchill, 1874-1965

Why not sequence everybody's DNA?

Cost (log_{10} scale) of whole-genome DNA sequencing





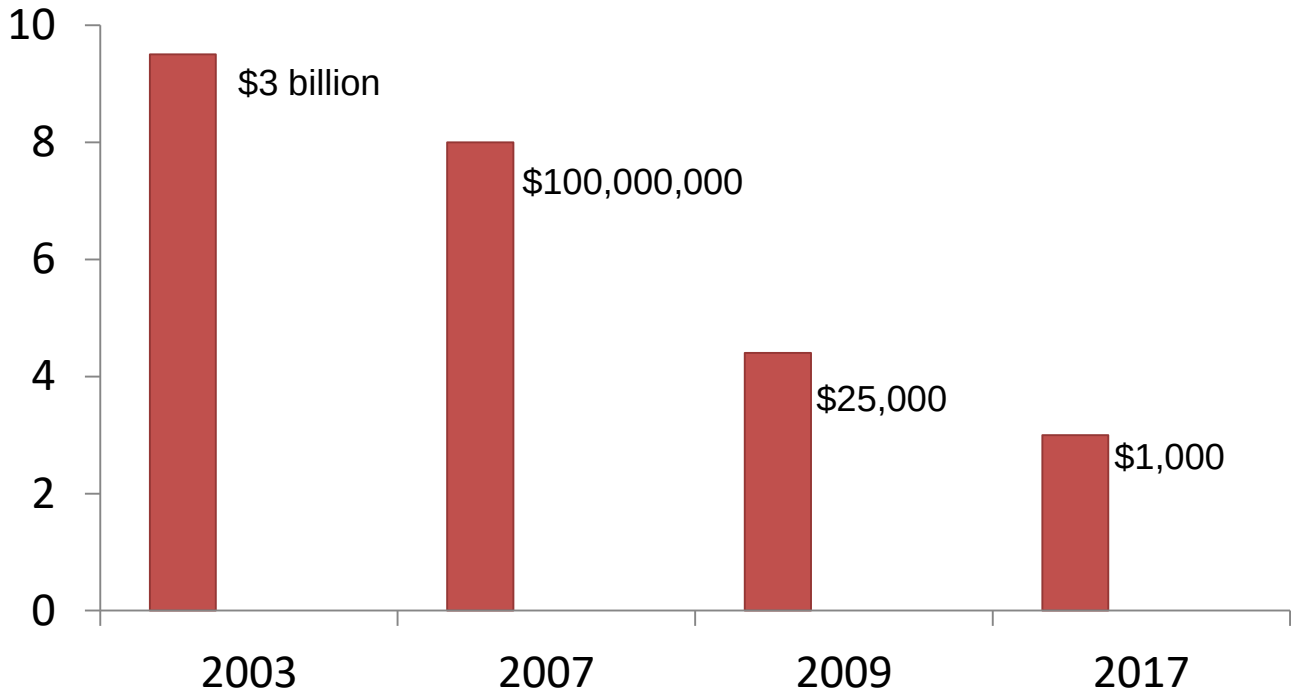
Economic Impact of the Human Genome Project

How a $3.8 billion investment drove $796 billion in economic impact, created 310,000 jobs and launched the genomic revolution

Prepared by Battelle Technology Partnership Practice

May 2011

Cost (log_{10} scale) of whole-genome DNA sequencing



HUMAN GENOME

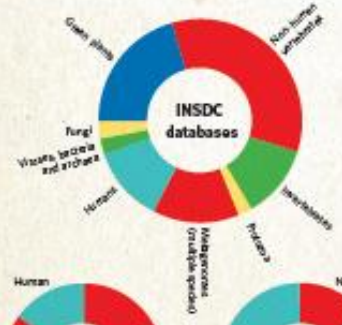
THE SEQUENCE EXPLOSION

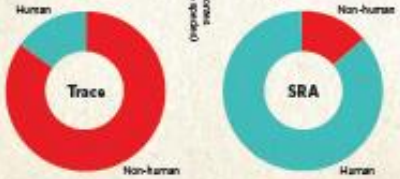
At the time of the announcement of the first drafts of the human genome in 2000, there were 8 billion base pairs of sequence in the three main databases for 'finished' sequence: GenBank, run by the US National Center for Biotechnology Information; the DNA Databank of Japan; and the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database. The databases share their data regularly as part of the International Nucleotide Sequence Database Collaboration (INSDC). In the subsequent first post-genome decade, they have added another 270 billion bases to the collection of finished sequence, doubling the size of the database roughly every 18 months. But this number is dwarfed by the amount of raw sequence that has been created and stored by researchers around the world in the Trace archive and Sequence Read Archive (SRA).

See Editorial, page 649, and human genome special at www.nature.com/humangenome

DNA SEQUENCES BY TAXONOMY

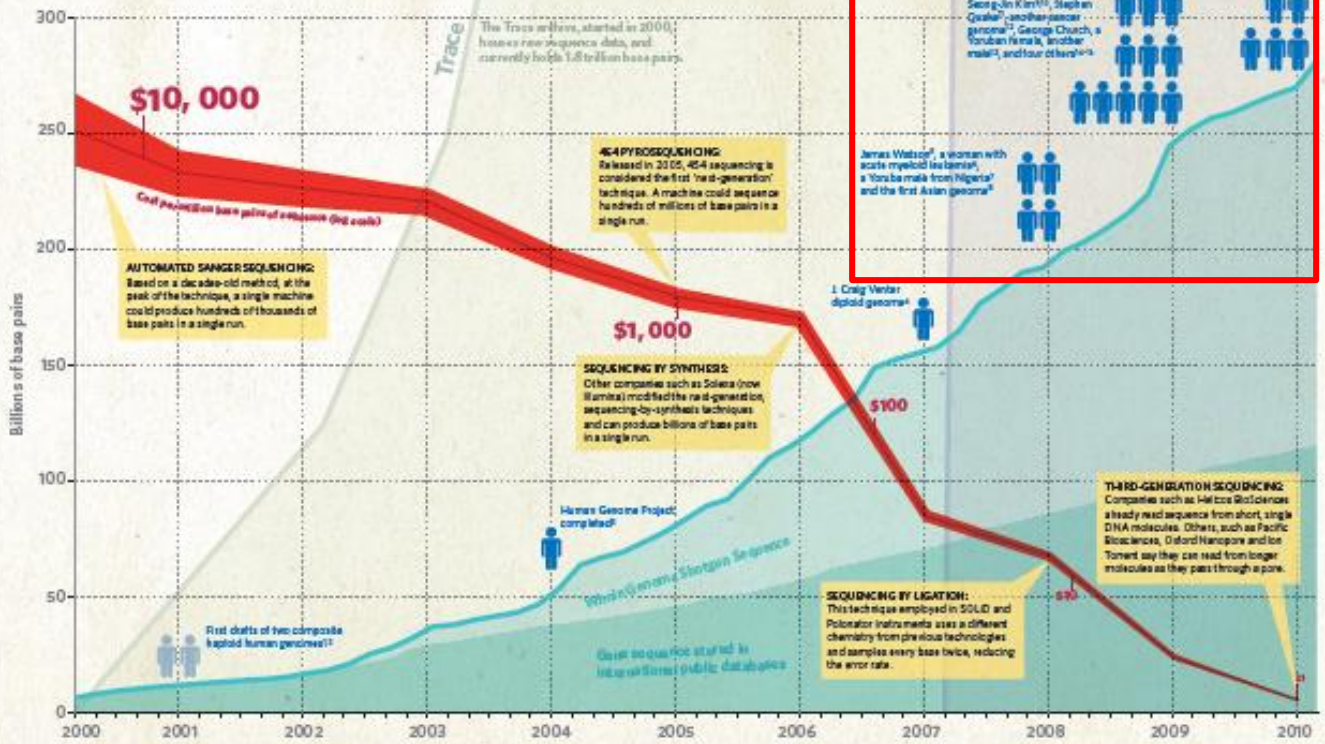
International Nucleotide Sequence Database Collaboration: The main repositories of 'finished' sequence span a wide range of organisms, representing the many priorities of scientists worldwide.





Trace Archive: Developed to house the raw output of high-throughput sequencers built in the late 1990s, the trace archive spans a wide range of taxa.

Sequence Read Archive: Houses raw data from next-generation sequencers. Dominated by human sequence, including multiple coverage for more than 170 people.



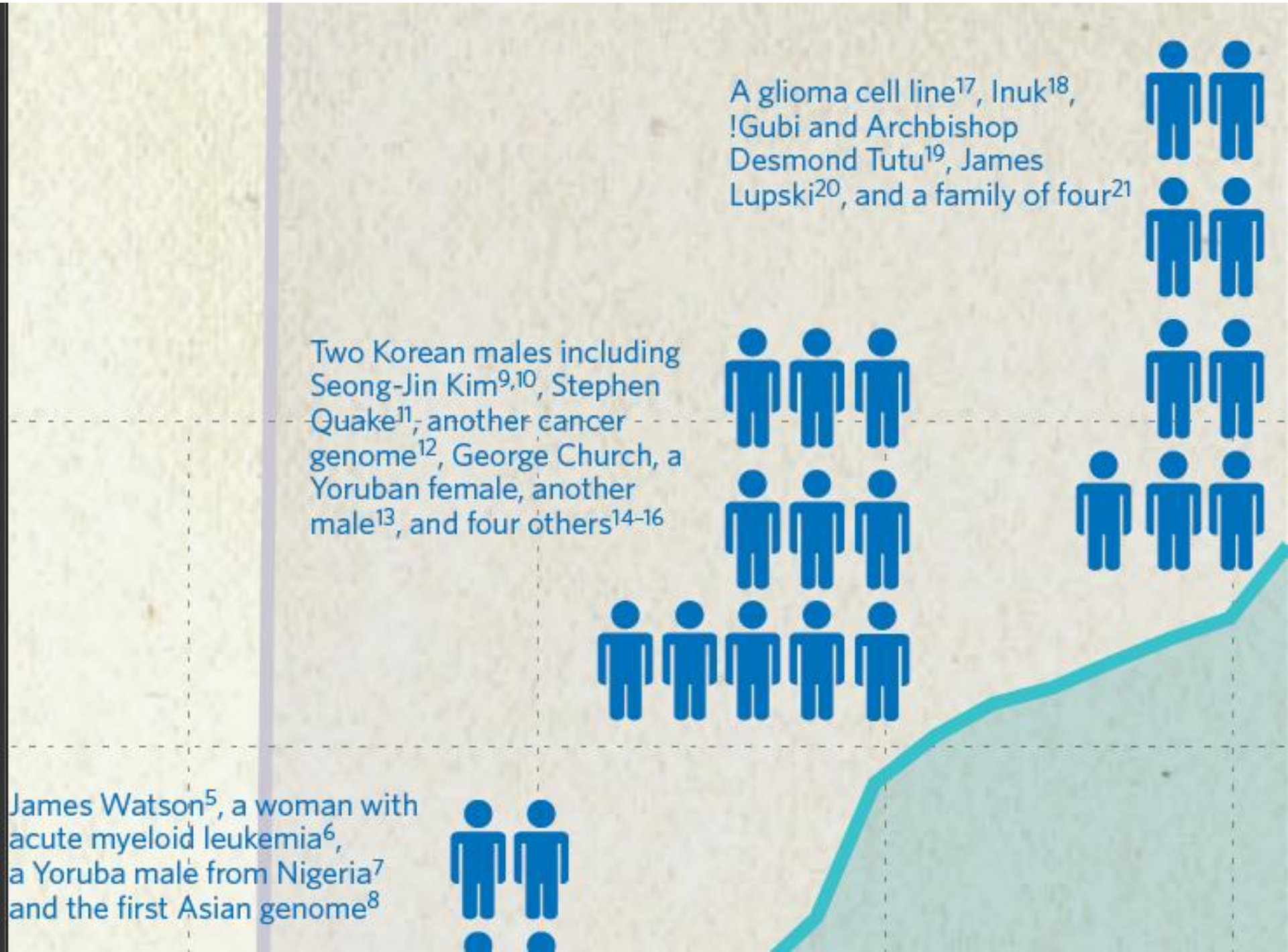
HOW MANY HUMAN GENOMES?

The graphic shows all published, fully sequenced human genomes since 2000, including nine from the first quarter of 2010. Some are resequencing efforts on the same person and the list does not include unpublished completed genomes.

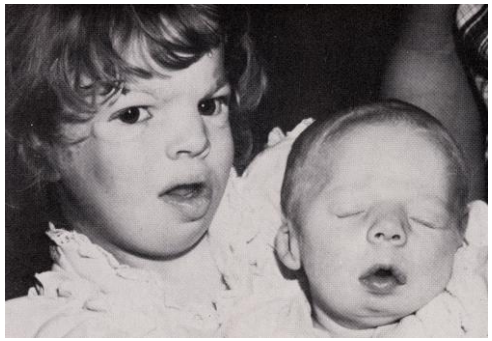
1. Venter, J. C. et al. Science 291, 1304–1351 (2001).
2. International Human Genome Sequencing Consortium Nature 409, 860–921 (2001).
3. International Human Genome Sequencing Consortium Nature 431, 931–945 (2004).
4. Levy, S. et al. PLoS Biol. 5, e254 (2007).
5. Wheeler, D. A. et al. Nature 452, 872–876 (2008).
6. Lay, T. J. et al. Nature 456, 66–72 (2008).
7. Bentley, D. R. et al. Nature 456, 53–59 (2008).
8. Wang, J. et al. Nature 456, 60–65 (2008).
9. Ahn, S.-M. et al. Genome Res. 19, 1622–1629 (2009).
10. Kim, J.-I. et al. Nature 460, 1011–1015 (2009).
11. Pushkarev, D., Neff, N. F. & Quake, S. R. Nature Biotechnol. 27, 847–850 (2009).
12. Wang, J. et al. N. Engl. J. Med. 360, 1058–1066 (2009).
13. Ley, T. J. et al. N. Engl. J. Med. 361, 1580–1592 (2009).
14. McKernan, K. J. et al. Genome Res. 19, 1527–1541 (2009).
15. Pleasance, E. D. et al. Nature 463, 191–196 (2010).
16. Pleasance, E. D. et al. Nature 463, 184–190 (2010).
17. Clark, M. J. et al. PLoS Genet. 6, e1000832 (2010).
18. Rasmussen, M. et al. Nature 463, 757–762 (2010).
19. Schuster, S. C. et al. Nature 463, 943–947 (2010).
20. Lupski, J. R. et al. N. Engl. J. Med. doi:10.1056/NEJMoa0908094 (2010).
21. Roach, J. C. et al. Science doi:10.1126/science.1186802 (2010).

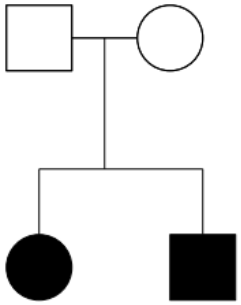
Page size by comparison

Nature, April 1, 2010


A glioma cell line[17], Inuk[18], !Gubi and Archbishop Desmond Tutu[19], James Lupski[20], and a family of four[21]
Two Korean males including Seong-Jin Kim[9,10], Stephen Quake[11], another cancer genome[12], George Church, a Yoruban female, another male[13], and four others[14–16]
James Watson[5], a woman with acute myeloid leukemia[6], a Yoruba male from Nigeria[7] and the first Asian genome[8]

The first sequenced family: Miller syndrome (postaxial acrofacial dysostosis)

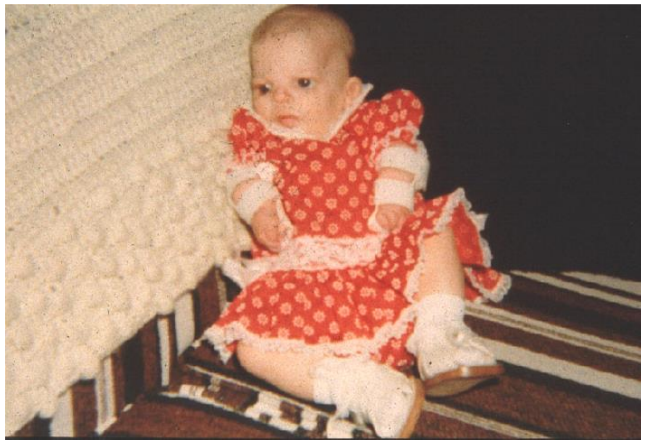




Fineman, 1981, *Journal of Pediatrics* 98: 87-8

Craniofacial malformations

Limb malformations



Cause of Miller syndrome remained unknown for three decades. Genome sequencing revealed the answer.

Exome and whole-genome sequencing (WGS) identifies two independent autosomal recessive conditions

- Miller syndrome: *DHODH* -- dihydroorotate dehydrogenase
  - involved in *de novo* synthesis of pyrimidines
- Primary ciliary dyskinesia: *DNAH5* -- dynein heavy chain needed for normal ciliary function
  - pulmonary infections, bronchiectasis
  - situs inversus
- WGS yields first direct estimate of the human mutation rate (10^{-8}/nucleotide/generation)

Mike Bamshad, MD

Chad Huff, PhD

Ng et al., 2010, *Nature Genetics*
Roach et al., 2010, *Science*

Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing

Jared C. Roach,[1*] Gustavo Glusman,[1*] Arian F. A. Smit,[1*] Chad D. Huff,[1,2*] Robert Hubley,[1] Paul T. Shannon,[1] Lee Rowen,[1] Krishna P. Pant,[3] Nathan Goodman,[1] Michael Bamshad,[4] Jay Shendure,[5] Radoje Drmanac,[3] Lynn B. Jorde,[2] Leroy Hood,[1]† David J. Galas[1]†

We analyzed the whole-genome sequences of a family of four, consisting of two siblings and their parents. Family-based sequencing allowed us to delineate recombination sites precisely, identify 70% of the sequencing errors (resulting in >99.999% accuracy), and identify very rare single-nucleotide polymorphisms. We also directly estimated a human intergeneration mutation rate of $\sim 1.1 \times 10^{-8}$ per position per haploid genome. Both offspring in this family have two recessive disorders: Miller syndrome, for which the gene was concurrently identified, and primary ciliary dyskinesia, for which causative genes have been previously identified. Family-based genome analysis enabled us to narrow the candidate genes for both of these Mendelian disorders to only four. Our results demonstrate the value of complete genome sequencing in families.

11 March 2010 VOL 328 *SCIENCE*

"All the News That's Fit to Print"

The New York Times

National Edition

Clouds yield to sun in the east. Mostly sunny elsewhere. Highs in 40s, but 50s in the far south. Mostly clear tonight. Mostly sunny, milder tomorro

VOL. CLIX . . . . No. 54,976 © 2010 The New York Times THURSDAY, MARCH 11, 2010

PANEL PROPOSES SINGLE STANDARD FOR ALL SCHOOLS

VISION FOR THE NATION

Goals for Every Grade Could Change Texts

Disease Cause Is Pinpointed With Genome

Decoding Cost Drops to $50,000 a Patient

By NICHOLAS WADE

Two research teams have inde-pendently decoded the entire ge-

POOL PHOTO BY SORKHABI

ighborhood

President Mahmoud Ahmadinejad of Iran to ed American actions in the region. Page A6.

f Their Bloody Revenge

Tuesday, 332 in a mass of Dogo Na Red Cross man rights government

lence in this fraught zone. The pattern is familiar and was seen as recently as January: uneasy coexistence suddenly explodes into killing, amplified for days by retaliation.

LEAD BLO TO

EFFOR

A Ba Spe P

B

WASH Democr day bar private tice tha dollars compan scandal

The step yet Congres use of a dividual nancing and bu campai

But quick r manshi Democr earmark just to publica party v called broken

Both claim th the issu budgeti

GENOMES ON PRESCRIPTION

The first clinical uses of whole-genome sequencing show just how challenging it can be.

BY BRENDAN MAHER

The first thing Debbie Jorde noticed about her newborn daughter was that her arms were bent at unnatural angles. She had other problems, too: a cleft palate, eight fingers, eight toes and no lower eyelids. She would eventually be diagnosed with Miller syndrome, a disease so rare that doctors have long assumed that each case arises through spontaneous mutation, rather than being passed down through families. Doctors told Jorde that her chances of having a second child with the syndrome were less than one in a million.

They were wrong. Jorde's son, born three years after his sister, had the same features. Lynn Jorde, Debbie's current husband and a geneticist at the University of Utah in Salt Lake City, still cringes when Debbie recounts what the doctors had told her. "The right answer for that situation is that there have been so few cases that we really can't predict the risk," he says.

Thanks to next-generation genome sequencing, Debbie and her children now know the family's genetic risks. Lynn and his collaborators had been talking about sequencing the genomes of an entire 'nuclear' family affected by a genetic disease, both to identify the mutation responsible and to investigate how genes are inherited in unprecedented detail. Debbie, her former husband and her now-adult children, Heather and Logan Madsen, were happy to be take part, and in 2009 became the first family in the world to have their genomes fully sequenced[1].

Over the course of six months, the research team cross-compared the whopping amount of DNA data from the four genomes. With the help of a parallel sequencing effort that included others

Nature, 6 Oct. 2011

Salt Lake Tribune, 11 March 2010

Top 100 Stories of 2010 #5: Family Genomics Links DNA to Disease

by Kathleen McAuliffe

From the January-February special issue; published online for subscribers only on December 16, 2010

An individual with Miller syndrome paints his own DNA.

Logan Madsen, David Galas, Gustavo Glusman/ISB

A decade ago, sequencing the dna in a person's entire genome cost up to $1 billion, a price so prohibitive that only a few genetics pioneers had the honor of having it done. In 2010 the cost per

The challenge of whole-genome sequences

- ~3-4 million variants, ~10,000 amino acid changes, and ~100 loss-of-function variants in an average individual
- Goal is to identify a tiny subset of these
- Data analysis “bottleneck”

Mardis *Genome Medicine* 2010, 2:84
http://genomemedicine.com/content/2/11/84

MUSINGS

The $1,000 genome, the $100,000 analysis?

Elaine R Mardis*

An “all-in-one” tool is needed

TECHNOLOGY FEATURE

WHEN DISEASE STRIKES FROM NOWHERE

When healthy parents have a child with a genetic disorder, the cause is sometimes a new mutation. Tools are emerging to meet the challenge of finding such changes.

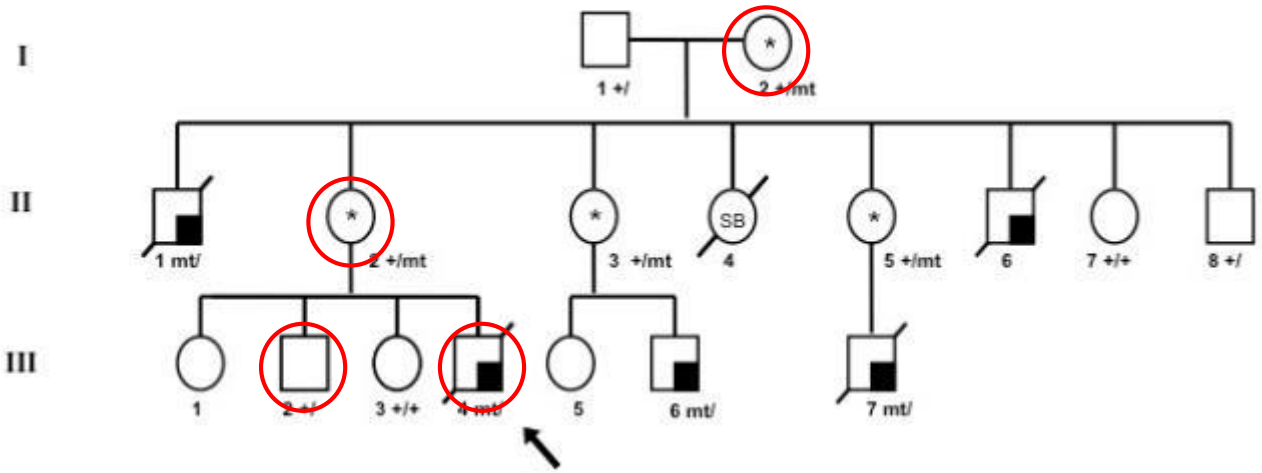
BIRKEDOSTOK/SHUTTERSTOCK

Children born with disorders not readily explained by standard tests can sometimes be diagnosed through genome sequencing and analysis.

Nature - Sept. 18, 2014

and non-coding genome regions is FastQForward, which integrates the software programs VAAST[3], pVAAST[4] and Phevor[5]. These tools were co-developed by Mark Yandell, a computational geneticist at the University of Utah in Salt Lake City who directs software development and computational analysis related to the Utah Genome Project. That project combines family histories from the Utah Population Database with medical records, which increasingly include DNA sequence information. The project includes family histories for more than 7 million people and medical records for around 4 million of them.

Sequencing identifies the cause of an undiagnosed lethal heart condition



X chromosome sequencing of four individuals

"Progeria-like" features; cardiac arrhythmia; death before age one year

Rope *et al., Am. J. Hum. Genet.* (July 2011)

nature news

nature news home | news archive | specials | opinion | features | news blo

Published online 23 June 2011 | Nature | doi:10.1038/news.2011.382

News

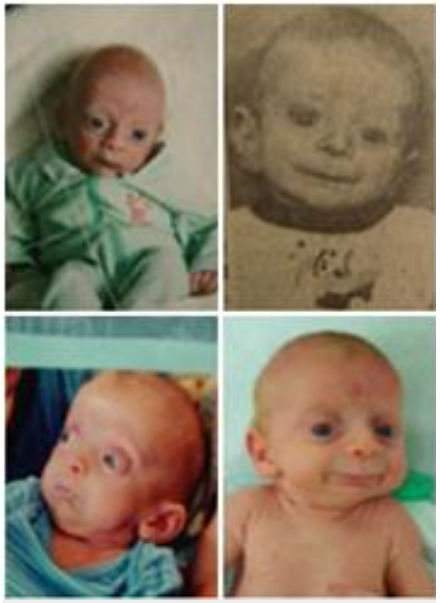
Software pinpoints cause of mystery genetic disorder

Genome analysis tools speedily track down previously unknown mutation.

Brendan Maher

Halena Black's first son, Kenny Rae, was born in November 1979. He struggled to put on weight, and had thin, wrinkled skin, big eyes and a broad mouth. In October the following year, he died from heart problems.

After Kenny Rae died, Black — a Mormon living in Ogden, Utah — had three healthy daughters before giving birth to another son in 1987. He had the same problem, and a similarly short lifespan. Her third son is healthy.



Four boys from the same

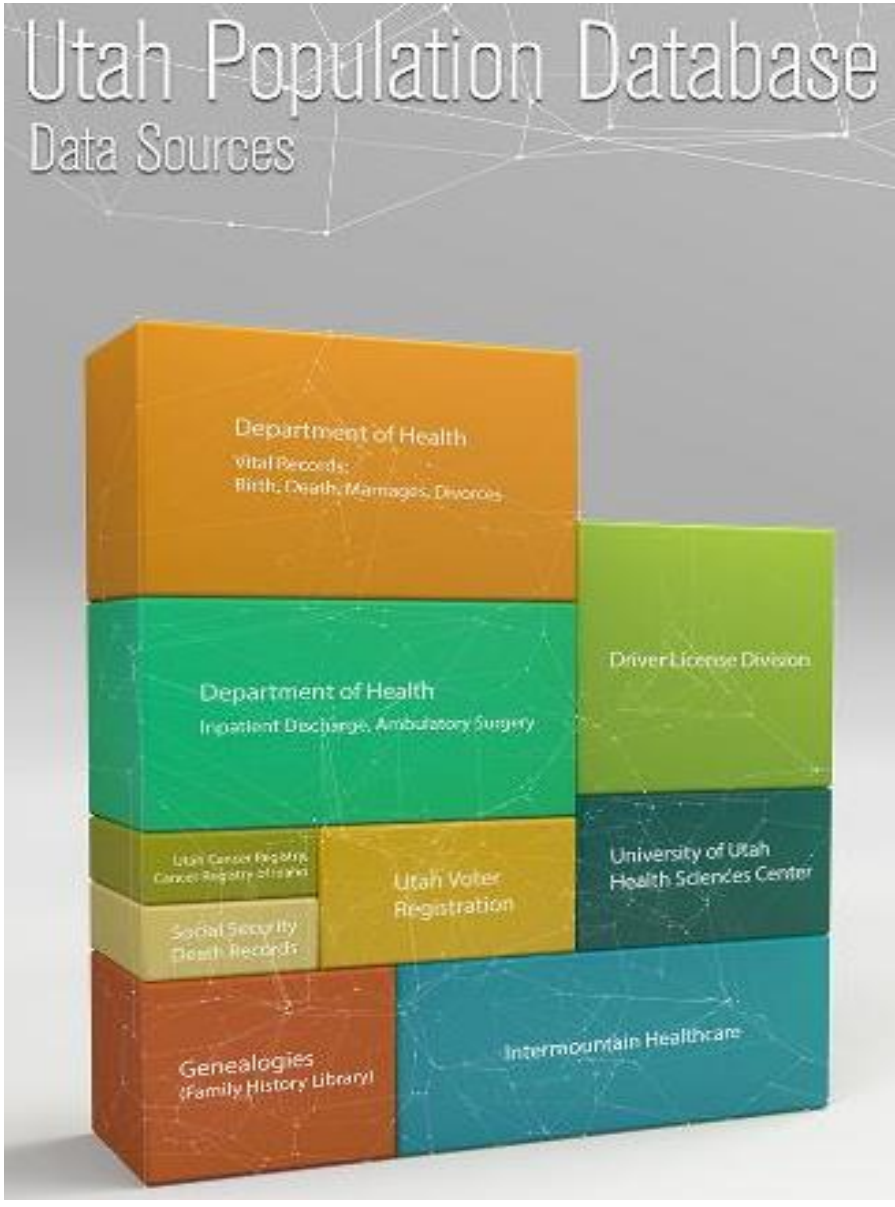
Genetic test developed by ARUP within one year of gene discovery.

IVF embryos can be genetically tested.

Why Utah?

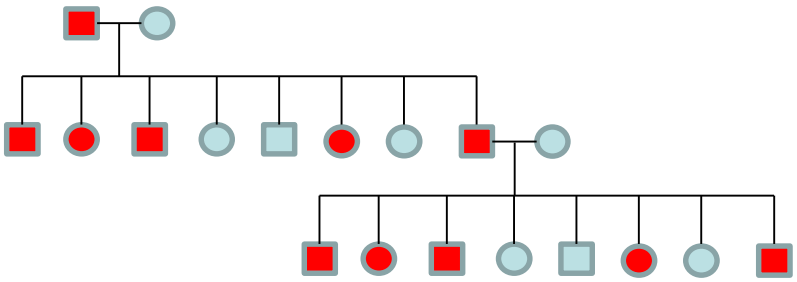
Utah Population Database (UPDB)

- University of Utah resource to support medical research, initiated in the 1970s
- 10 million people in large, multigenerational pedigrees are linked to more than 25 million medically relevant records
- Geocode information available
- Now expanded to 100 million people



Whole-genome sequencing and families: a game-changing combination

- Sequencing allows us to detect *all* disease-causing DNA variants: both rare and common
- Families are a genetic “magnifying glass”



Utah Genome Project: >7,000 cases now sequenced

- Amyotrophic lateral sclerosis **(Neurology)**
- Genetics of extreme thinness to develop obesity interventions **(Internal Medicine)**
- Spontaneous preterm birth **(Obstetrics/Gynecology)**
- Childhood cancers (Ewing sarcoma, Wilms tumor, germ cell tumors) **(Pediatrics, HCI)**
- DNA repair genes and cancer **(Oncological Sciences)**
- Hematologic cancers: CLL, CML, multiple myeloma **(Hematology, HCI)**
- Common cancers: breast, prostate, colorectal **(Internal Medicine, HCI)**
- Psoriasis and psoriatic arthritis **(Dermatology)**
- Juvenile idiopathic arthritis **(Pediatric Rheumatology)**
- Crohn disease **(Pediatric Gastroenterology)**
- Common variable immunodeficiency **(Pediatrics, Pathology)**
- Chronic obstructive pulmonary disease **(Pulmonology)**
- Idiopathic pulmonary fibrosis **(Pulmonology)**
- Familial cardiac arrhythmia **(Pediatric Cardiology)**
- Autism **(Psychiatry)**
- Primary ovarian insufficiency **(Endocrinology, Internal Medicine)**

Utah Genome Project: New genes discovered!

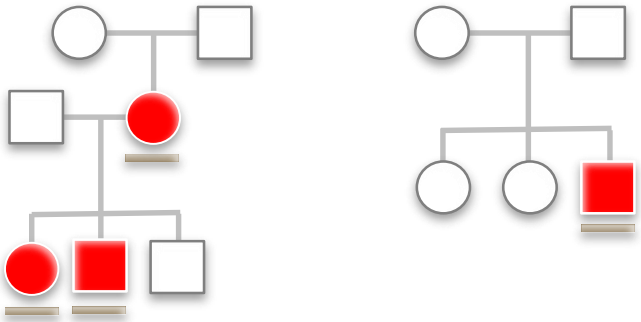
- Amyotrophic lateral sclerosis (Neurology)
- Genetics of extreme thinness to develop obesity interventions
- Spontaneous preterm birth
- Childhood cancers (Ewing sarcoma, Wilms tumor, germ cell tumors)
- DNA repair genes and cancer
- Hematologic cancers: CLL, CML, multiple myeloma
- Common cancers: breast, prostate, colorectal
- Psoriasis and psoriatic arthritis
- Juvenile idiopathic arthritis
- Crohn disease
- Common variable immunodeficiency
- Chronic obstructive pulmonary disease
- Idiopathic pulmonary fibrosis
- Familial cardiac arrhythmia
- Autism
- Primary ovarian insufficiency

UGP Discoveries: Examples

- Two new ALS candidate genes (*TP73, MFN2*)
  - Gibson, Downie*, et al.,* 2017, *Neurology;* additional MSs in prep.
- *POLR2C* is a cause of primary ovarian insufficiency
  - Moriwaki *et al*., 2017, *J Endocr Soc*
- *RIPK2* is a cause of familial early-onset osteoarthritis
  - Jurynec *et al.,* 2018, *Hum. Mol. Genet.*
- New candidate loci for early infantile epileptic encephalopathy (Ostrander, Butterfield, Quinlan)
  - *Journal of Clinical and Translational Science* (submitted)
- *NKX3-2* causes Treacher-Collins syndrome
  - Velinder *et al.* (MS in preparation)
- *USP45* and *ARID1A* cause inherited multiple myeloma (Waller et al., 2018, *PLoS Genet.*)

Identification of *NFKB2* mutations that cause common variable immunodeficiency (CVID)

NFKB2 added to multi-gene testing panel to increase diagnostic rate



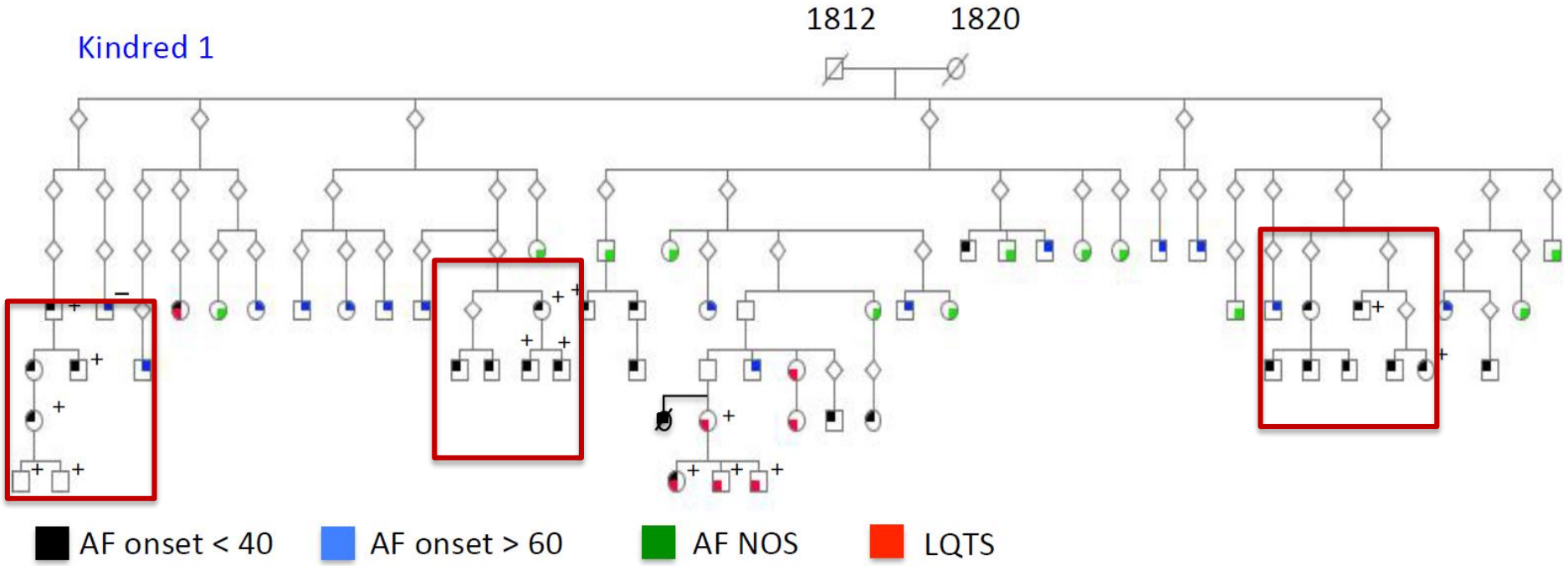
Karin Chen, MD
Pediatrics

Attila Kumanovics, MD, Pathlogy

Chen, Coonrod, *et al*., 2013, *Am. J. Hum. Genet.*
Collaborators in Depts. of Pediatrics, Pathology, Human Genetics, Molecular Medicine, and ARUP

Family data can identify at-risk pedigree members: Young-onset atrial fibrillation

Martin Tristani-Firouzi, MD



Using DNA Sequencing to Target Drug Therapy for Spontaneous Preterm Birth

Tracy Manuck, MD Ob/Gyn

Case-control study of responders and non-responders to progesterone (17p) treatment to prevent recurrent spontaneous preterm birth

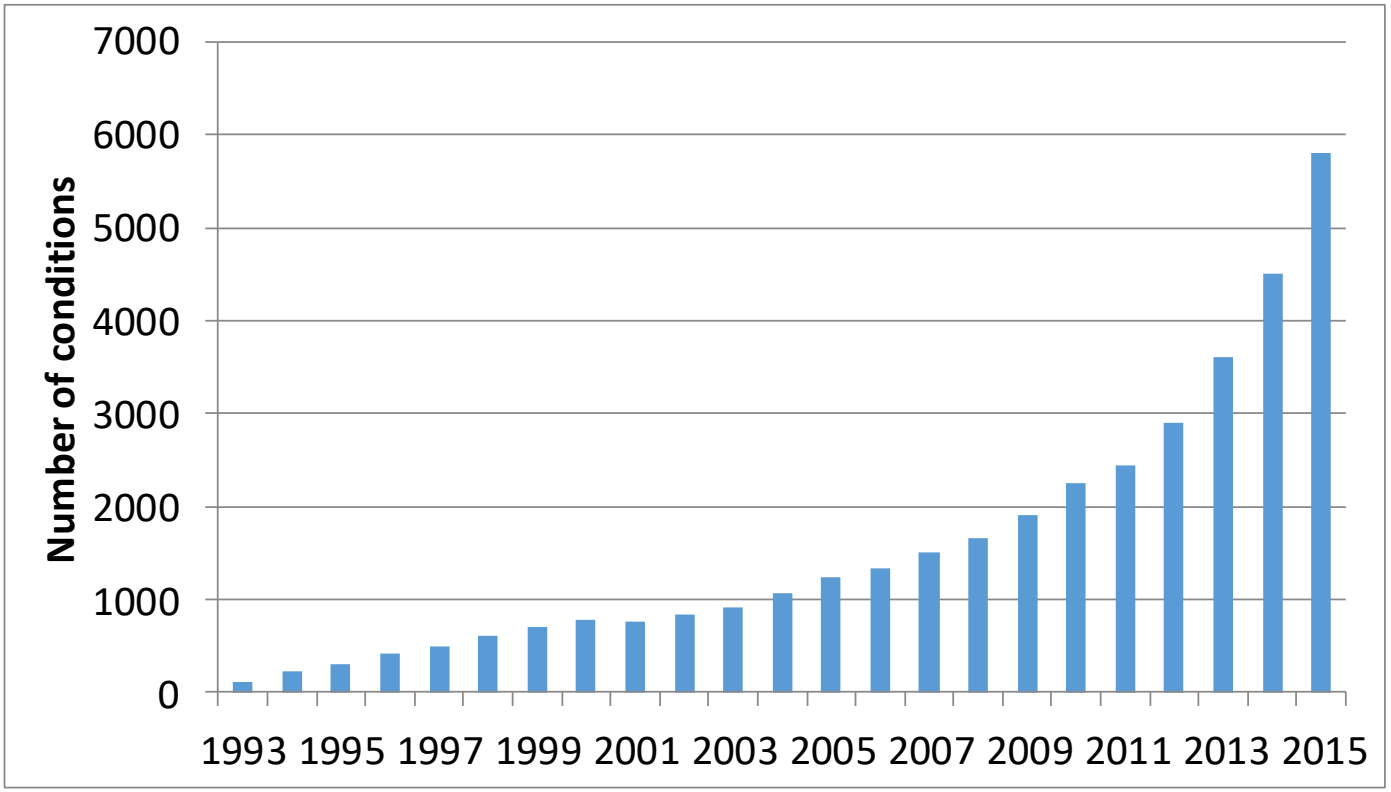
U of U analysis software identifies *NOS1* (nitric oxide synthase) as a mediator of response to 17P therapy

"This is the first step in using pharmacogenomics to prevent preterm birth," Edward McCabe, MD, March of Dimes chief medical officer, 3 February 2014

Manuck *et al., Am. J. Obstet. Gynecol.,* 2014
Manuck *et al., Brit. J. Obstet. Gynaecol.,* 2017

Gene discovery leads to genetic testing

2014: $5 billion market; 2020: at least $20 billion

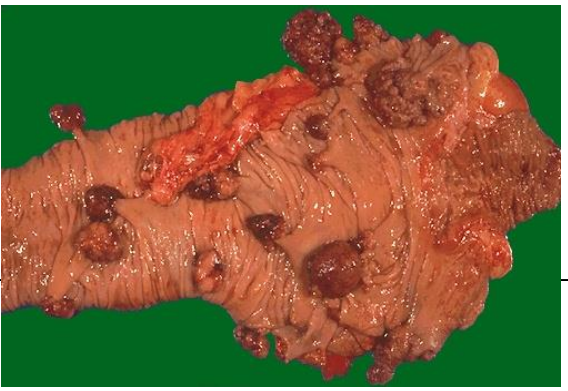


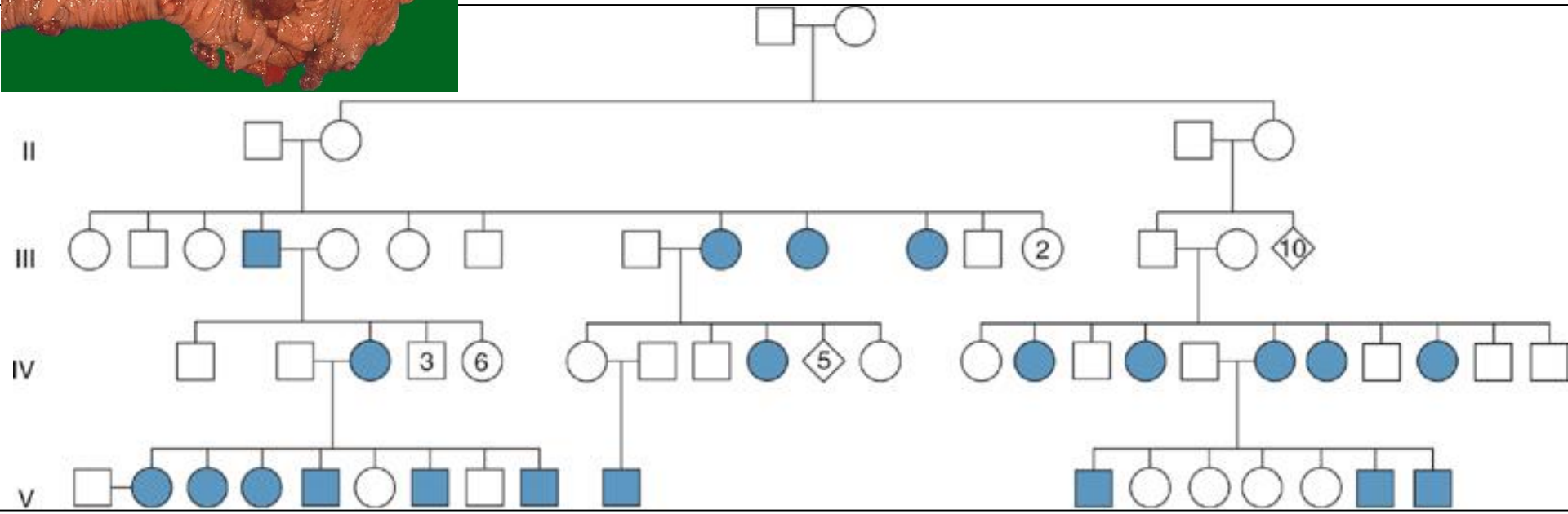
Genetic Testing Registry: http://www.ncbi.nlm.nih.gov/gtr/

Examples of diseases for which mutation testing is available (single-gene diseases)

- Cystic fibrosis
- Duchenne/Becker muscular dystrophy
- Hemophilia A, B
- PKU
- Fragile X syndrome
- Hemochromatosis
- Sickle cell disease
- Neurofibromatosis 1, 2
- Familial hypercholesterolemia
- Huntington disease
- Familial breast cancer, colon cancer
- Familial Alzheimer disease

Familial adenomatous polyposis (FAP): Dominant inheritance of an *APC* mutation in a Utah family





Each filled symbol represents a person diagnosed with colorectal cancer

Genetic discoveries can lead to disease prevention: *BRCA1* mutation carrier Angelina Jolie

- Family history of breast and ovarian cancer
- *BRCA* genes can be tested to reveal genetic risk

Examples of drugs for which FDA recommends genetic testing (7% of 1200 FDA-approved drugs)

Drug	Indication	Associated gene
Abacavir	HIV	*HLA-B5701*
Carbamazepine	Seizures, bipolar disorder	*HLA-B1502*
Azathioprine	Rheumatoid arthritis	*TPMT*
Thioguanine	Acute myeloid leukemia	*TPMT*
Clopidogrel	Atherothrombosis	*CYP2C19*
Warfarin	Anticoagulation management	*KVORC1/CYP2C9*
Imatinib	Chronic myeloid leukemia	*BCR-ABL* translocation
	GIST	*cKIT*
Lapatinib	Breast cancer	*HER2*
Tamoxifen	Breast cancer	ER/PR expression
Erlotinib	Non-small cell lung cancer	*EGFR*
Lanalidomide	Myelodysplastic syndrome	5q deletion

Korf and Rehm, 2013, *JAMA*

Successful one-dose gene therapy for spinal muscular atrophy type 1 (SMA1)

Evelyn Villarreal, treated for spinal muscular atrophy type 1 with a new gene therapy, is nearly 3. Few children with her condition reach age 2.

"As of the data cutoff on August 7, 2017, all 15 patients were alive and event-free at 20 months of age, as compared with a rate of survival of 8% in a historical cohort." (Mendell *et al., NEJM,* 2 November 2017)

SMA1 (autosomal recessive disease affecting 1/10,000 births) is now screened in all Utah newborns; treatment can begin at ~2 months of age

Gene therapy: examples of successes

Condition	Disease type	Patients benefiting
X-linked SCID	Immunodeficiency	17/20
ADA-SCID	Immunodeficiency	26/37
Adrenoleukodystrophy	Neurologic disease	2/4
Leber congenital amaurosis	Eye disease/blindness due to retinal dystrophy	28/30
Wiskott-Aldrich syndrome	Immunodeficiency	13/14
Metachromatic leukodystrophy	Myelinopathy	20/20
Beta-thalassemia major	Hemoglobinopathy	9/10
Hemophilia A	Coagulation deficiency	7/7
Hemophilia B	Coagulation deficiency	11/17
B-cell lymphoma; ALL	Hematologic cancer	58/71
Spinal muscular atrophy 1	Neurodegeneration	15/15

2016-17 FDA approvals: antisense therapy for SMA1;CAR T-cell therapy for B-cell malignancies; AAV therapy for retinal dystrophy

Kaiser, 2011, *Science* 334: 29-30
Naldini, 2015, *Nature* 526: 351-60

Summary

- Whole genome sequencing is now accessible and affordable
- Combined with powerful analytic techniques, WGS or WES can:
  - Discover new disease-causing genes
  - Improve diagnosis and patient management
  - Provide new drug targets and guidelines for targeted drug therapy
- “Genomic medicine” is real!