

# *At The Forefront of Genomics:* Making Genomic Medicine a Reality

**Eric Green, M.D., Ph.D.**  
Director, NHGRI

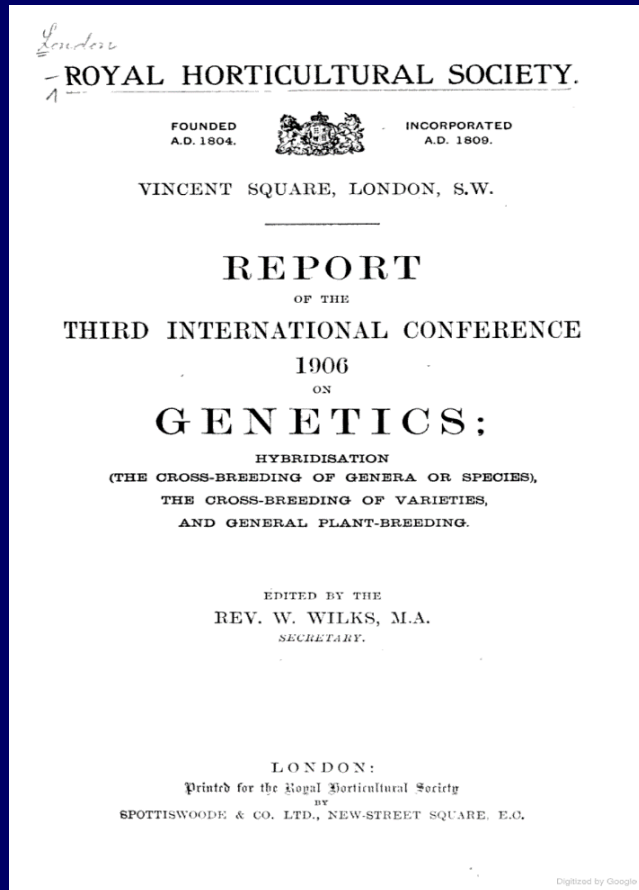


National Human Genome  
Research Institute

The **Forefront**  
of **Genomics**<sup>®</sup>

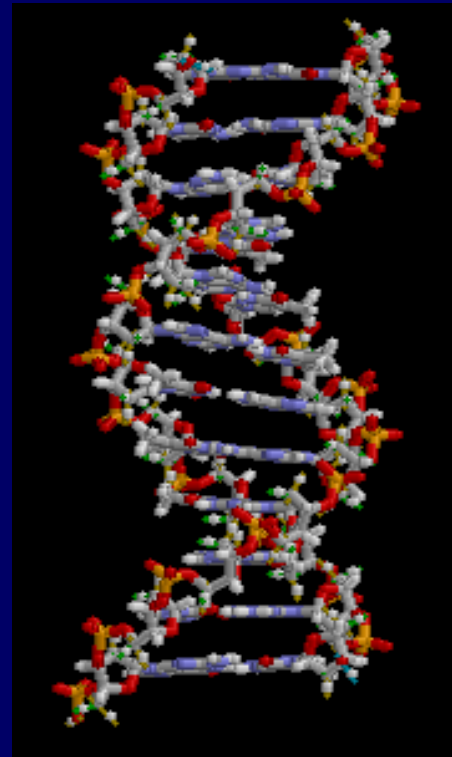
# Two Scientific Fields Launched Last Century Are Changing Medicine This Century

## “Genetics”



1907

## DNA's Double Helix



1953

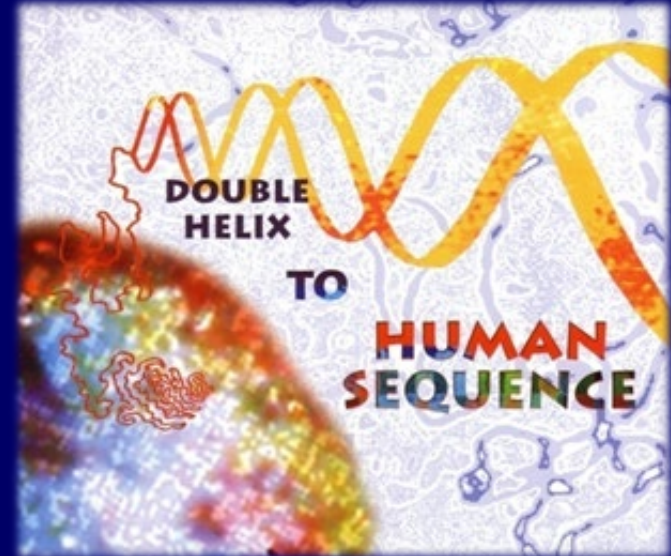
## “Genomics”



1987



# Human Genome Project: 1990-2003



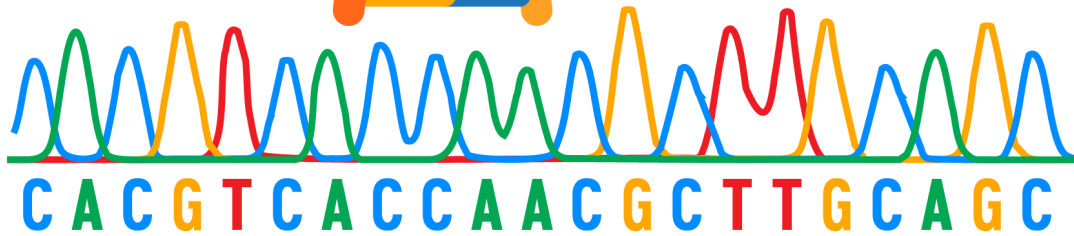
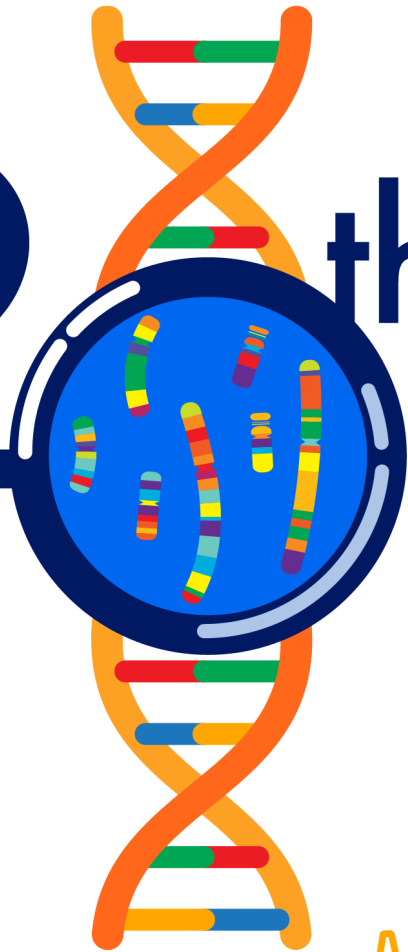
[genome.gov/HGP](http://genome.gov/HGP)

2003

2023

20<sup>th</sup>

# Anniversary of the Human Genome Project Completion





D.T. Max: What Darwin Can Teach Us About Jane Austen. Alex Witchel: An Erma Bombeck for Military Wives

# The New York Times Magazine

NOVEMBER 6, 2005 • SECTION 6

## The Dawn of Genomic Medicine

How a pediatrician working with the Amish is changing what it means to diagnose and treat disease.

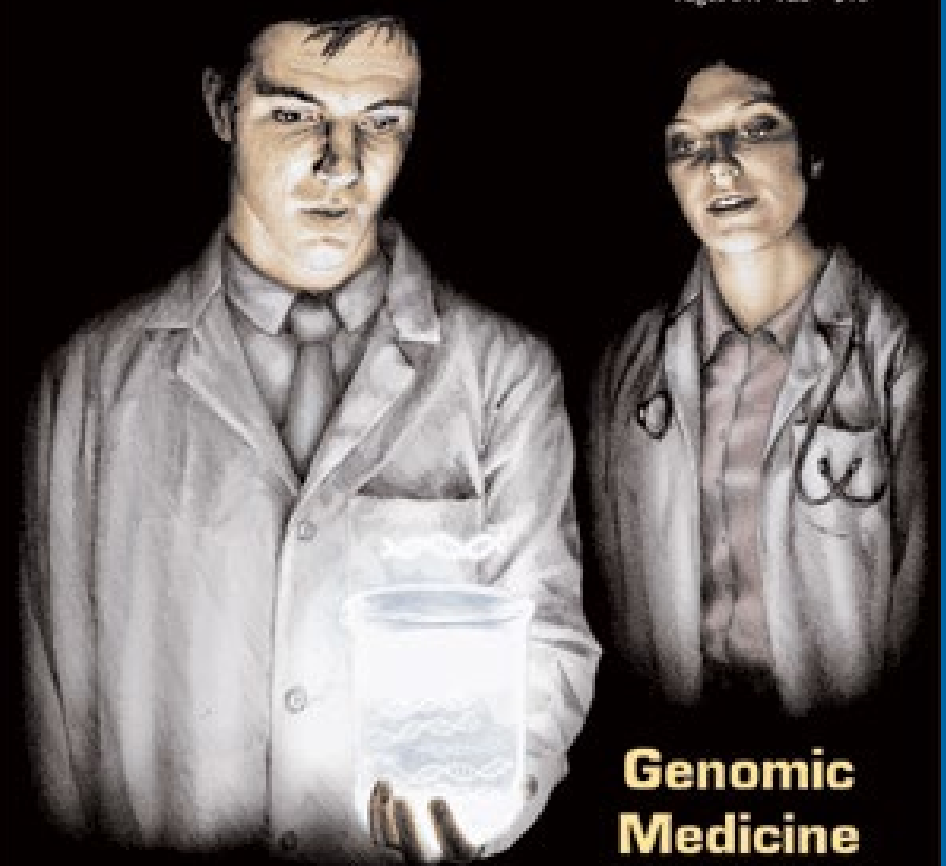
By Lisa Belkin

Why Waste Your Time Voting? (See Freakonomics, Page 30)

# Science

24 October 2003

Vol. 302 No. 5645  
Pages 517-728 510

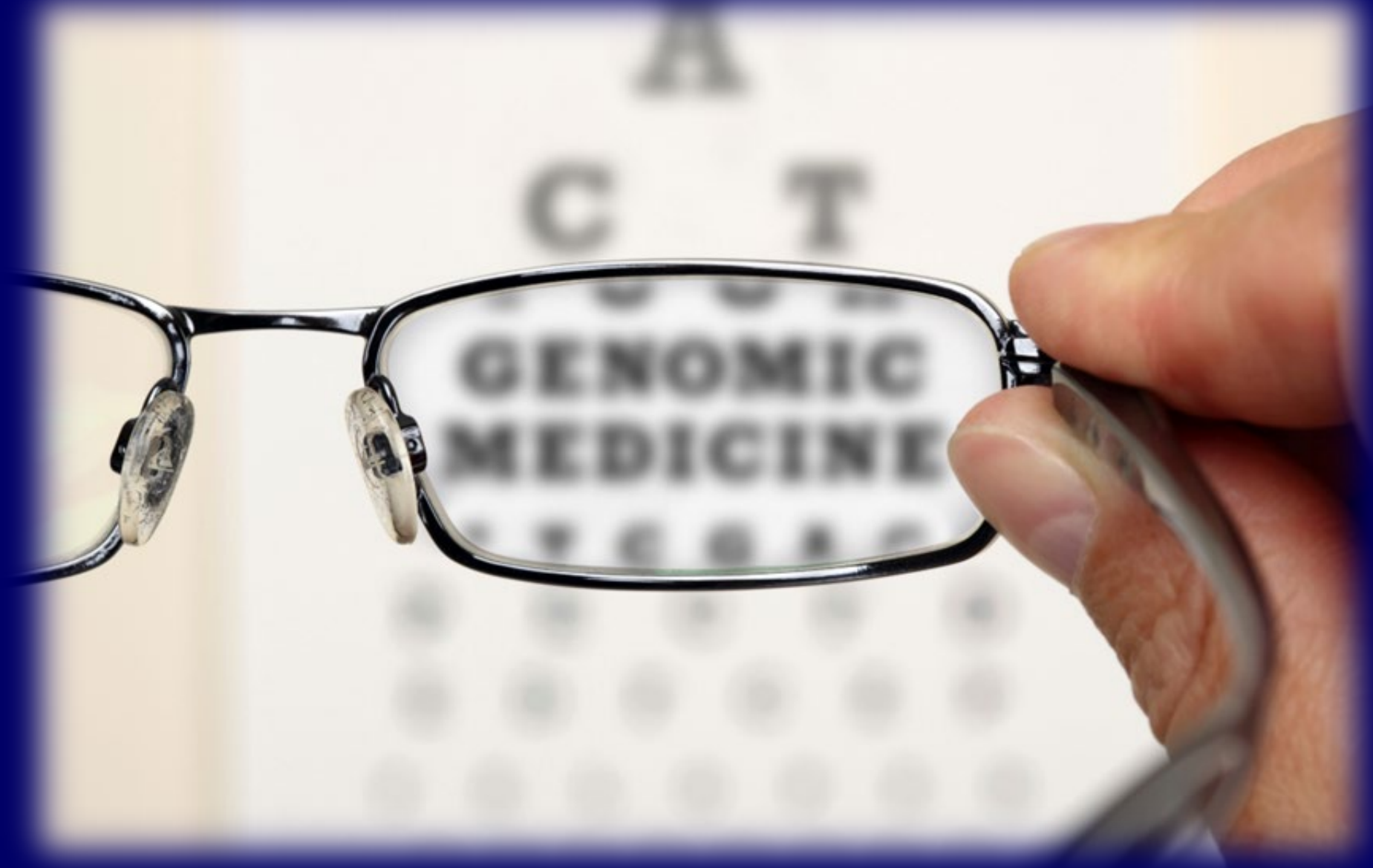


## Genomic Medicine



AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

# Bringing Genomic Medicine Into Focus





# Genomic Medicine

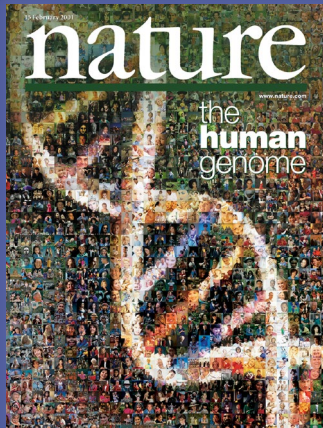
An emerging medical discipline that involves using genomic information about an individual as part of their clinical care (e.g., for diagnostic or therapeutic decision-making) and the other implications of that clinical use



Related (but not identical) terms:

- Personalized medicine
- Individualized medicine
- Precision medicine

# The Pivot to Genomic Medicine



Human Genome Project



Realization of Genomic Medicine



# En Route to Genomic Medicine



**Human Genome Sequenced for First Time  
by the Human Genome Project**



**Cost of Sequencing a Human Genome  
Reduced >1 Million-Fold**



**Millions of Human  
Genomes Sequenced**



**Profound Advances in Understanding  
How the Human Genome Functions**



**Significant Advances in Unraveling the  
Genomic Bases of Human Disease**



# Making Genomic Medicine a Reality





# Analyzing a Patient's Genome is Becoming More Routine

TTGAACACCATTGGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC  
GGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC  
CACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC  
TCGAGGAAACTTGAACACCATTGGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATT  
TTGAACACCATTGGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC  
GGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC  
CACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC  
TCGAGGAAACTTGAACACCATTGGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATT  
TTGAACACCATTGGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC  
GGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC  
CACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC  
TCGAGGAAACTTGAACACCATTGGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATT  
TTGAACACCATTGGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC  
GGCACGATGCTCCGTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACCATTGGGTTCGAGGAAACTTGAACACC

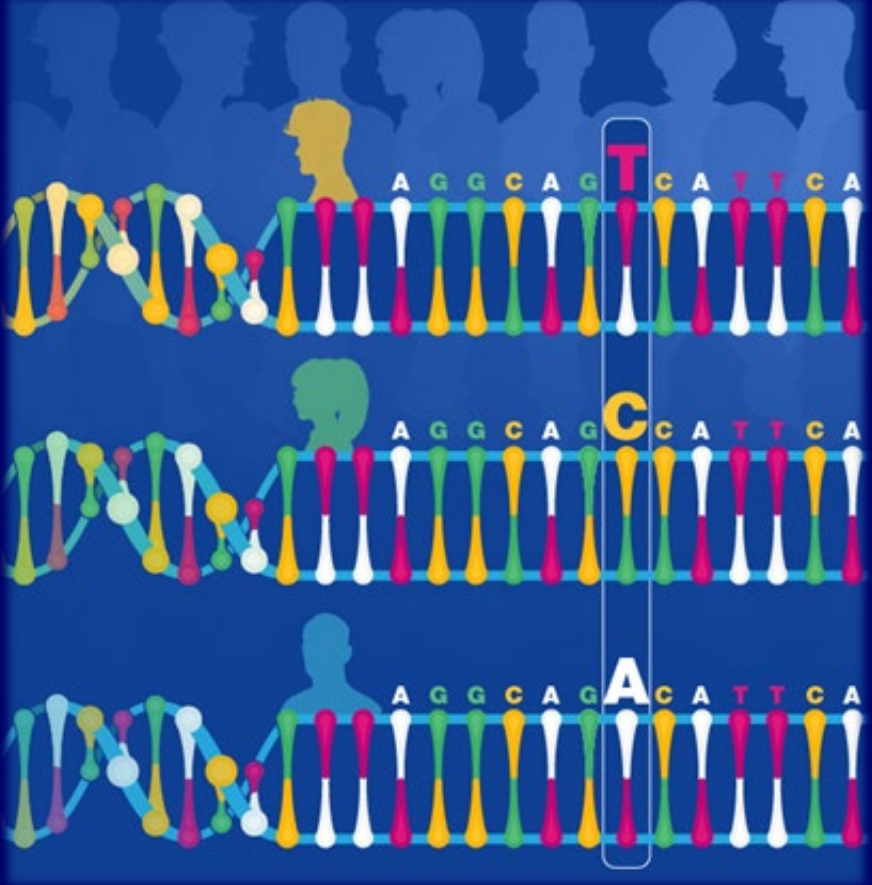
# Genome Sequencing in Medicine

Reference  
Genome Sequence

Patient's  
Genome Sequence

Reference Genome Sequence: `TTAACTTGAGTGGTTAATAACGGTATGCTCGGGATTAAGTGTAGTATAGTGTGATTATCGGAGA`

Patient's Genome Sequence: `TTAACTTGAGTGGATTACTAACGGTATTTCTGGGATTA...GTATAGTGTGAGTATCGGAGA`



List of  
Genomic Variants



Implement  
Genomic Medicine



# Genomic Medicine Implementation

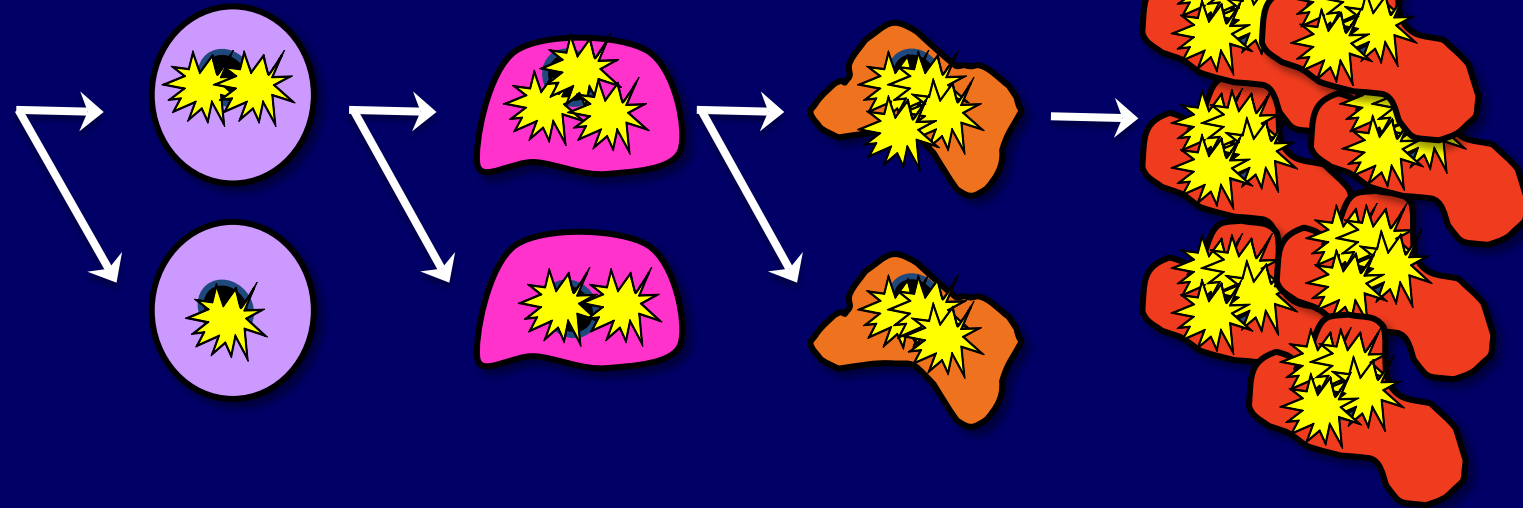
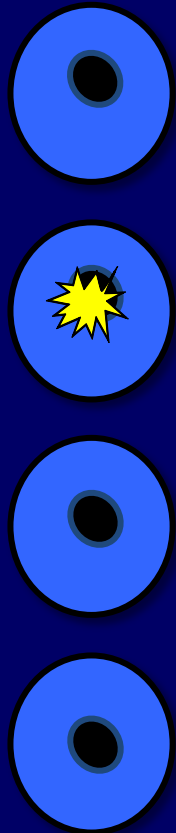


Cancer  
Genomics



# Cancer is a Disease of the Genome

Normal  
Cells



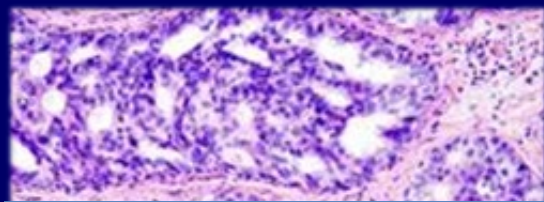
Tumor  
Cells

It Takes Multiple Mutations to Make a Cell Malignant

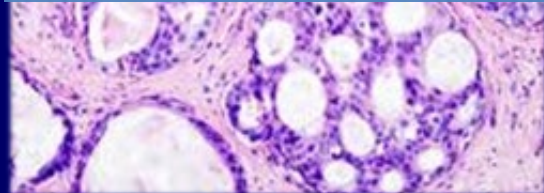


# Routine Cancer Diagnostic Tools

## Cancer Histopathology



**Morphology**



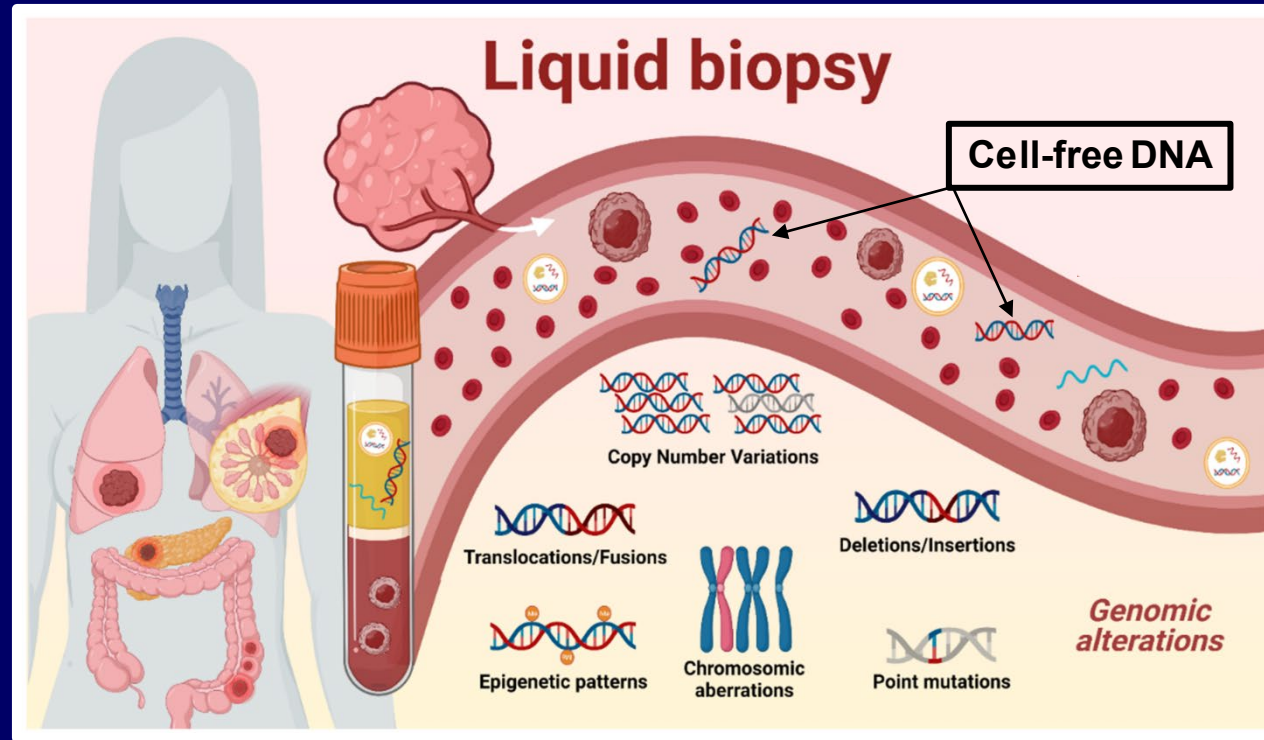
## Cancer Genome Sequencing



**Genomic Signature**

**Paradigm Change: Genomic signature of a tumor often (perhaps almost always) provides more valuable clinical information than the tissue of origin.**

# Liquid Biopsy for Detecting Cancer



- Standard biopsies of human tissues are invasive and can be dangerous
- Tumor cells frequently die and release their DNA into the bloodstream
- Highly sensitive DNA-sequencing methods can detect and analyze that cell-free tumor DNA (accessed through simple blood draw)



# Cancer Genomics Today

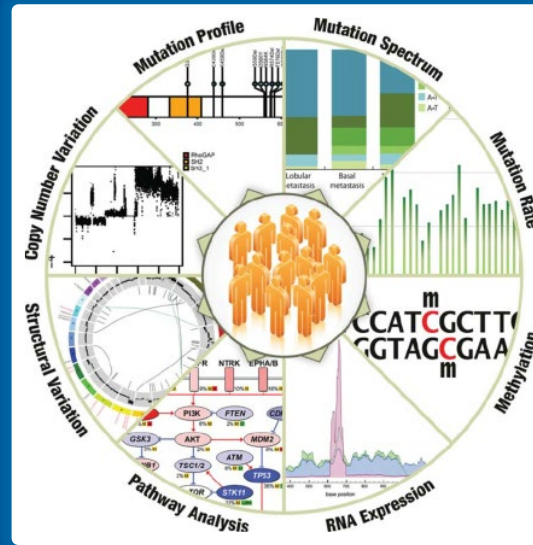
In retrospect

## 15 years after a giant leap for cancer genomics

Sheng F. Cai & Ross L. Levine

In 2008, the first comprehensive sequence of a cancer genome was reported, ushering in a new era of molecular diagnostic, prognostic and therapeutic advances informed by an essential framework to understand cancer's complexities.

*Nature, 2023*



*“Today, genomic sequencing as part of clinical care has transformed cancer diagnostics, clinical trials, and the use of new therapies to improve outcomes for people with cancer. Our unprecedented view of the cancer genome empowers clinicians, computational biologists, and bench scientists alike to define biologically relevant groups of people with cancer, direct genomic inquiry, and ultimately identify new therapies and biomarkers.”*

# Genomic Medicine Implementation



**Cancer  
Genomics**

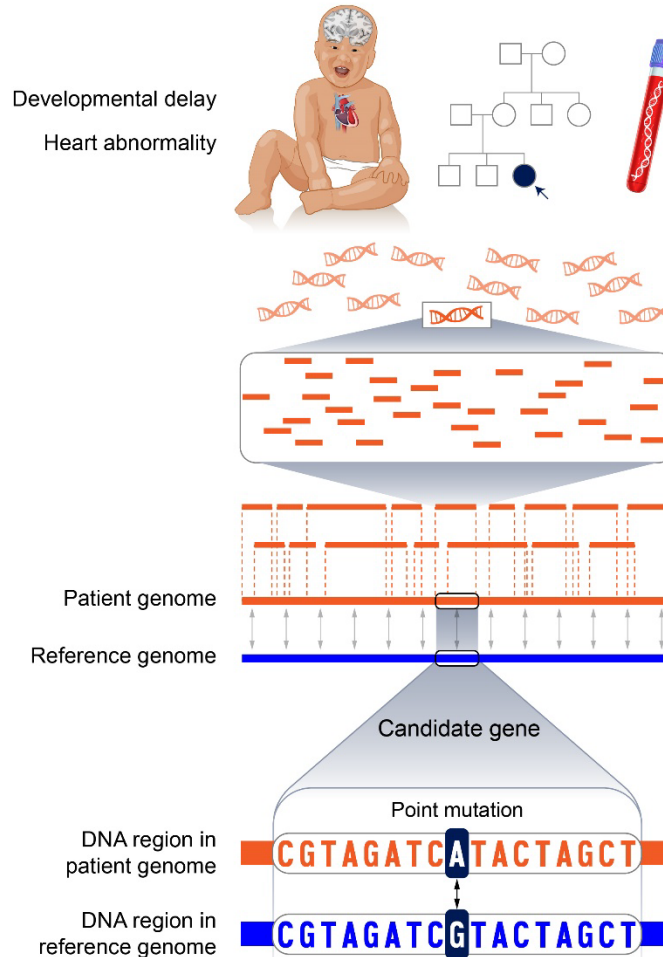


**Rare Genetic Disease  
Diagnostics**





# Genome Sequencing is Becoming a Standard Diagnostic Tool in Medicine



- Gather medical and family history
- Collect DNA sample



- Sequence patient's genomic DNA
- Align patient's genome sequence to reference human genome sequence



- Identify and characterize genomic variant(s) potentially associated with disease
- Analyze knowledge of candidate gene(s) relative to patient's clinical features
- If possible, make disease diagnosis

# Rare Disease Diagnostics

ORIGINAL ARTICLE

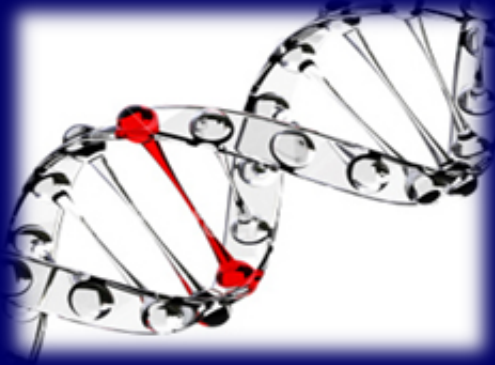
## Genome Sequencing for Diagnosing Rare Diseases

M.H. Wojcik, G. Lemire, E. Berger, M.S. Zaki, M. Wissmann, W. Win, S.M. White, B. Weisburd, D. Wiczorek, L.B. Waddell, J.M. Verboon, G.E. VanNoy, A. Töpf, T.Y. Tan, S. Syrbe, V. Strehlow, V. Straub, S.L. Stenton, H. Snow, M. Singer-Berk, J. Silver, S. Shril, E.G. Seaby, R. Schneider, V.G. Sankaran, A. Sanchis-Juan, K.A. Russell, K. Reinson, G. Ravenscroft, M. Radtke, D. Popp, T. Polster, K. Platzer, E.A. Pierce, E.M. Place, S. Pajusalu, L. Pais, K. Öunap, I. Osei-Owusu, H. Opperman, V. Okur, K.T. Oja, M. O'Leary, E. O'Heir, C.F. Morel, A. Merckenschlager, R.G. Marchant, B.E. Mangilog, J.A. Madden, D. MacArthur, A. Lovgren, J.P. Lerner-Ellis, J. Lin, N. Laing, F. Hildebrandt, J. Hentschel, E. Groopman, J. Goodrich, J.G. Gleeson, R. Ghaoui, C.A. Genetti, J. Gburek-Augustat, H.T. Gazda, V.S. Ganesh, M. Ganapathi, L. Gallacher, J.M. Fu, E. Evangelista, E. England, S. Donkervoort, S. DiTroia, S.T. Cooper, W.K. Chung, J. Christodoulou, K.R. Chao, L.D. Cato, K.M. Bujakowska, S.J. Bryen, H. Brand, C.G. Bönnemann, A.H. Beggs, S.M. Baxter, T. Bartolomaeus, P.B. Agrawal, M. Talkowski, C. Austin-Tse, R. Abou Jamra, H.L. Rehm, and A. O'Donnell-Luria

***NEJM, 2024***

**GENOME SEQUENCING: Yields a diagnosis for a rare genetic disease in ~30-50% of cases (and this % will increase over time!).**

# Undiagnosed Diseases





# Rapid Genome Sequencing of Sick Newborns



CONRAD BORNHAIN/GALLO/GETTY

The genomes of ill newborns can be sequenced in less than 24 hours to give clinicians a rapid diagnosis.

## GENOMICS

# Fast sequencing saves newborns

Rapid analysis of infant genomes is aiding diagnosis and treatment of inexplicably ill babies.

**Begin NGS**  
NEWBORN GENOMIC SEQUENCING  
to end the diagnostic odyssey

Newborn Screening via rapid Whole Genome Sequencing

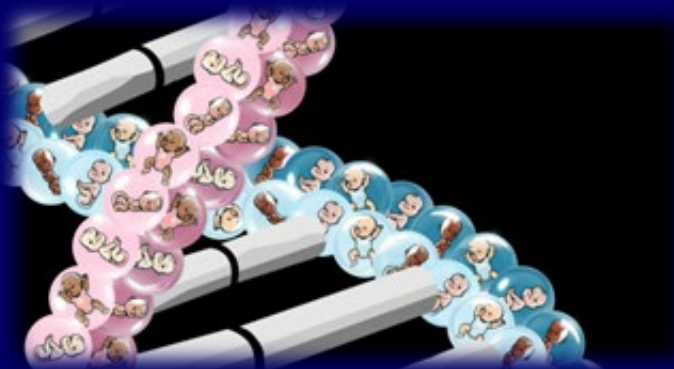
+  
TREATMENT GUIDANCE  
for clinical care team

Rady Children's Institute  
Genomic Medicine

[www.Begin-NGS.org](http://www.Begin-NGS.org)

*Nature, 2014*

# Rapid Genome Sequencing of Sick Newborns



npj | genomic medicine Review article

Published in partnership with CEGMR, King Abdulaziz University

<https://doi.org/10.1038/s41525-024-00404-0>

## Rapid genomic sequencing for genetic disease diagnosis and therapy in intensive care units: a review

Check for updates

Stephen F. Kingsmore , Russell Nofsinger & Kasia Ellsworth



*NPJ Genome Med, 2024*

***“In 44 studies of children in ICUs with diseases of unknown etiology, 37% received a genetic diagnosis, 26% had consequent changes in management, and net healthcare costs were reduced by \$14,265 per child tested...”***

***In five years, there is the potential for infant and childhood mortality in the US and UK to have been reduced by several percent through use of [rapid genome sequencing] as a first-tier, standard of care test for children in ICUs with diseases of uncertain etiology.”***

# Genomic Medicine Implementation



**Cancer  
Genomics**



**Rare Genetic Disease  
Diagnostics**



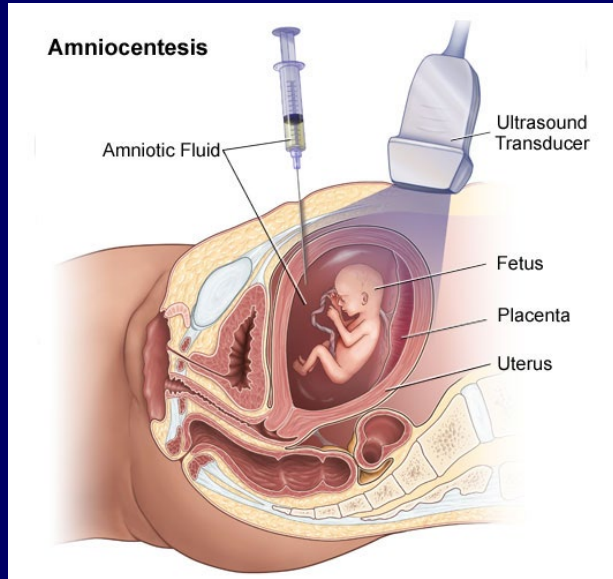
**Noninvasive Prenatal  
Genomic Testing**





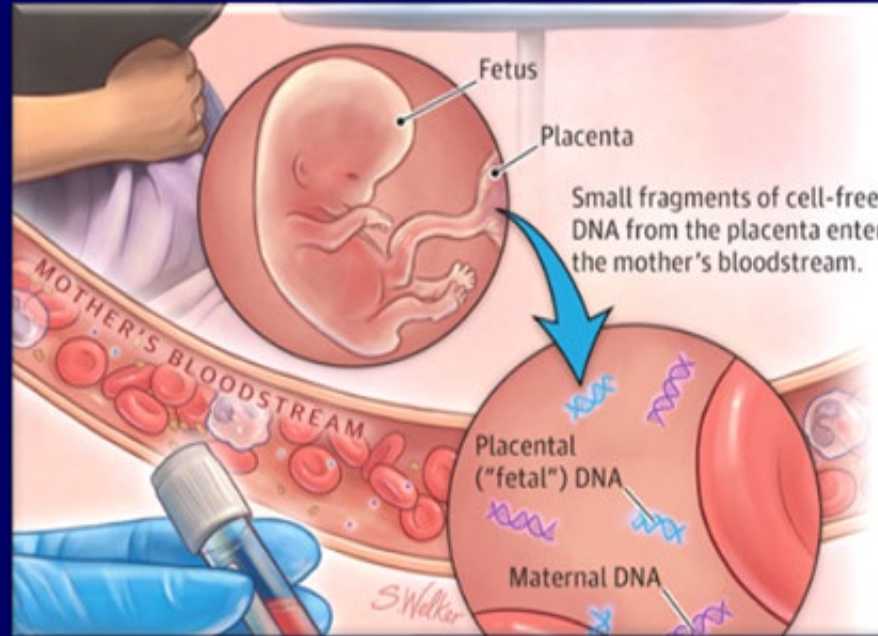
# Noninvasive Prenatal Testing (NIPT)

Then (before Genomics)



↓  
“Aneuploidy”

Now (with Genomics)



- Both mother and fetus release cell-free DNA from dying cells into the blood
- As an alternative to an invasive procedure, sequencing of cell-free DNA in maternal plasma now used to screen for aneuploidy
- #1 genomic medicine test worldwide

# Genomic Medicine Implementation



**Cancer  
Genomics**



**Rare Genetic Disease  
Diagnostics**



**Noninvasive Prenatal  
Genomic Testing**



**Pharmacogenomic  
Testing**






# People Respond Differently to Medications



Because Everyone Responds Differently.



All of these work.

Just not for everyone.

Perlegen may be able to help you sort out which medicine helps which patient.

Working with you, we can comprehensively analyze the DNA from thousands of patients taking your drug. Out of the millions of genetic variations between patients, we may be able to help you identify the ones that are associated with strong efficacy, poor efficacy, or side effects.

Perlegen's exceptional coverage of the genome and experienced team of analysts could help you get clinically relevant answers, not just data, in a matter of months.

We partner with the top pharmaceutical companies around the world. We also license late-stage drugs. If you have a drug that can benefit from our approach, please contact us.



# Prescribing Medications is Imprecise

## IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

1. **ABILIFY** (aripiprazole)  
Schizophrenia



2. **NEXIUM** (esomeprazole)  
Heartburn



3. **HUMIRA** (adalimumab)  
Arthritis



4. **CRESTOR** (rosuvastatin)  
High cholesterol



5. **CYMBALTA** (duloxetine)  
Depression



6. **ADVAIR DISKUS** (fluticasone propionate)  
Asthma



7. **ENBREL** (etanercept)  
Psoriasis



8. **REMICADE** (infliximab)  
Crohn's disease



9. **COPAXONE** (glatiramer acetate)  
Multiple sclerosis

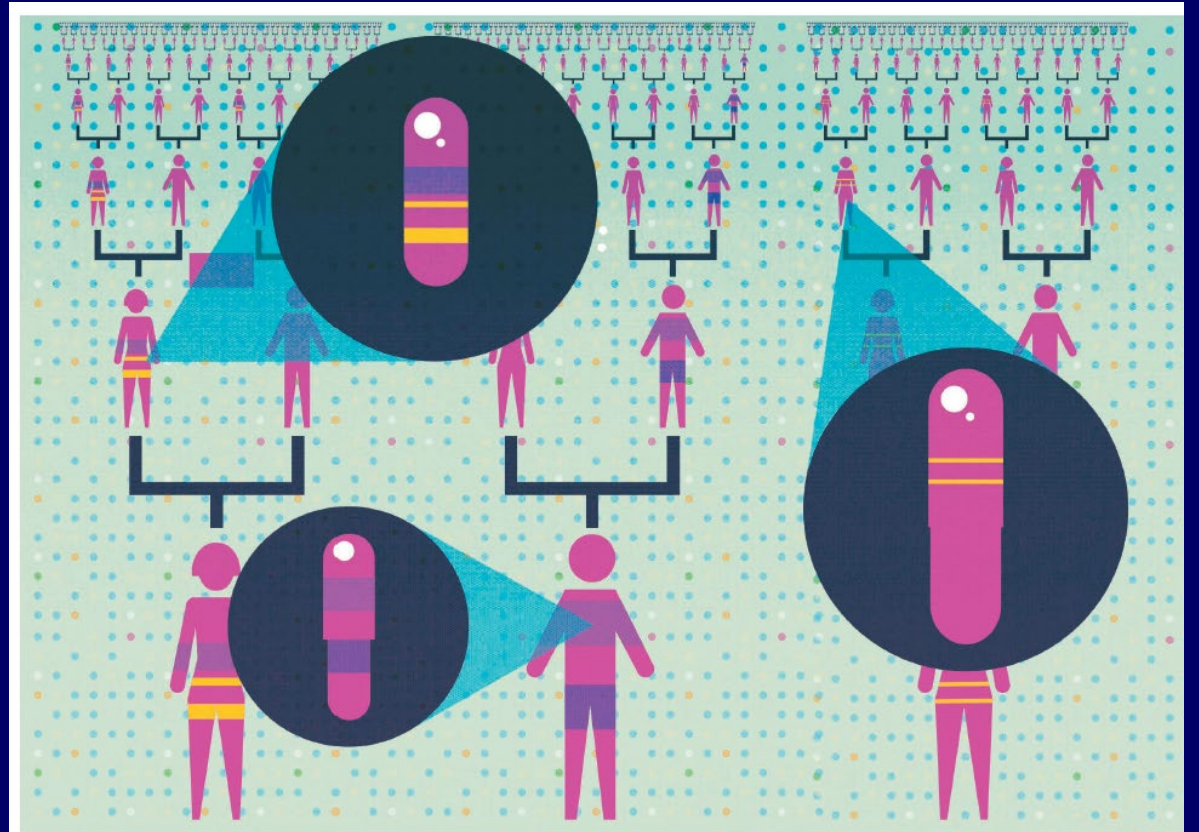


10. **NEULASTA** (pegfilgrastim)  
Neutropenia



Based on published number needed to treat (NNT) figures. For a full list of references, see Supplementary Information at [go.nature.com/4dr78f](http://go.nature.com/4dr78f).

*Nature* (2015)

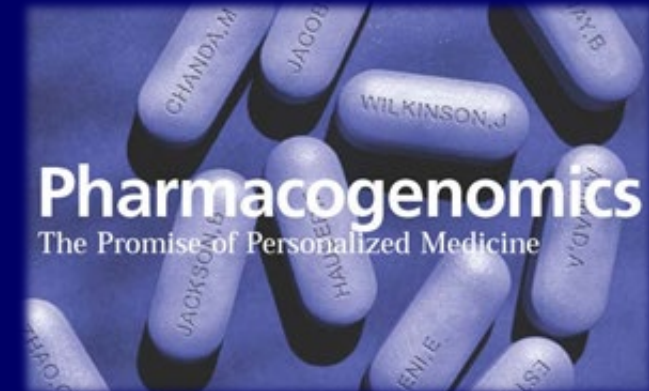
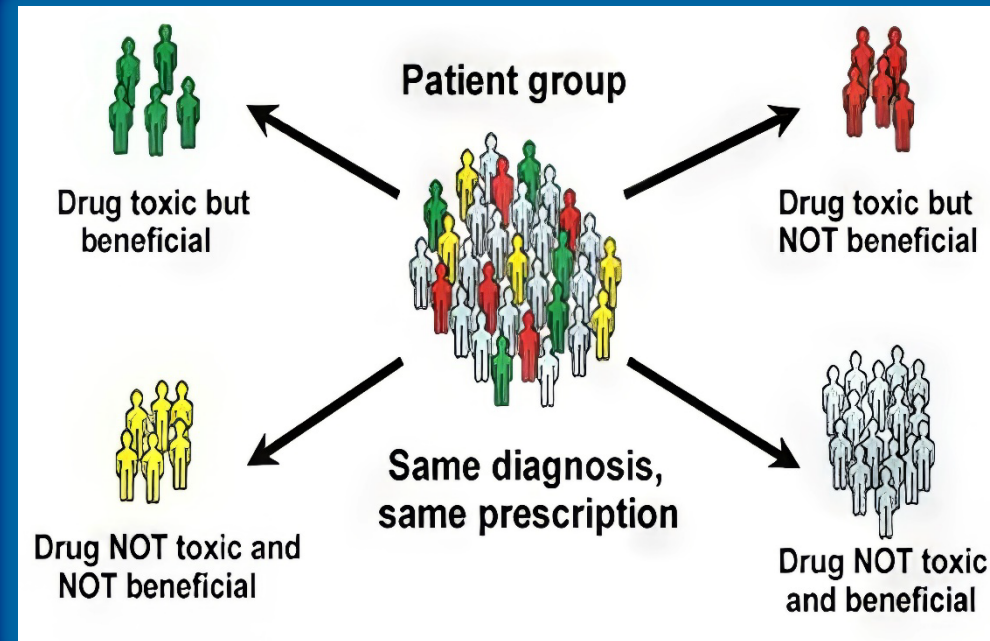


PHARMACOGENETICS

## The right drug for you

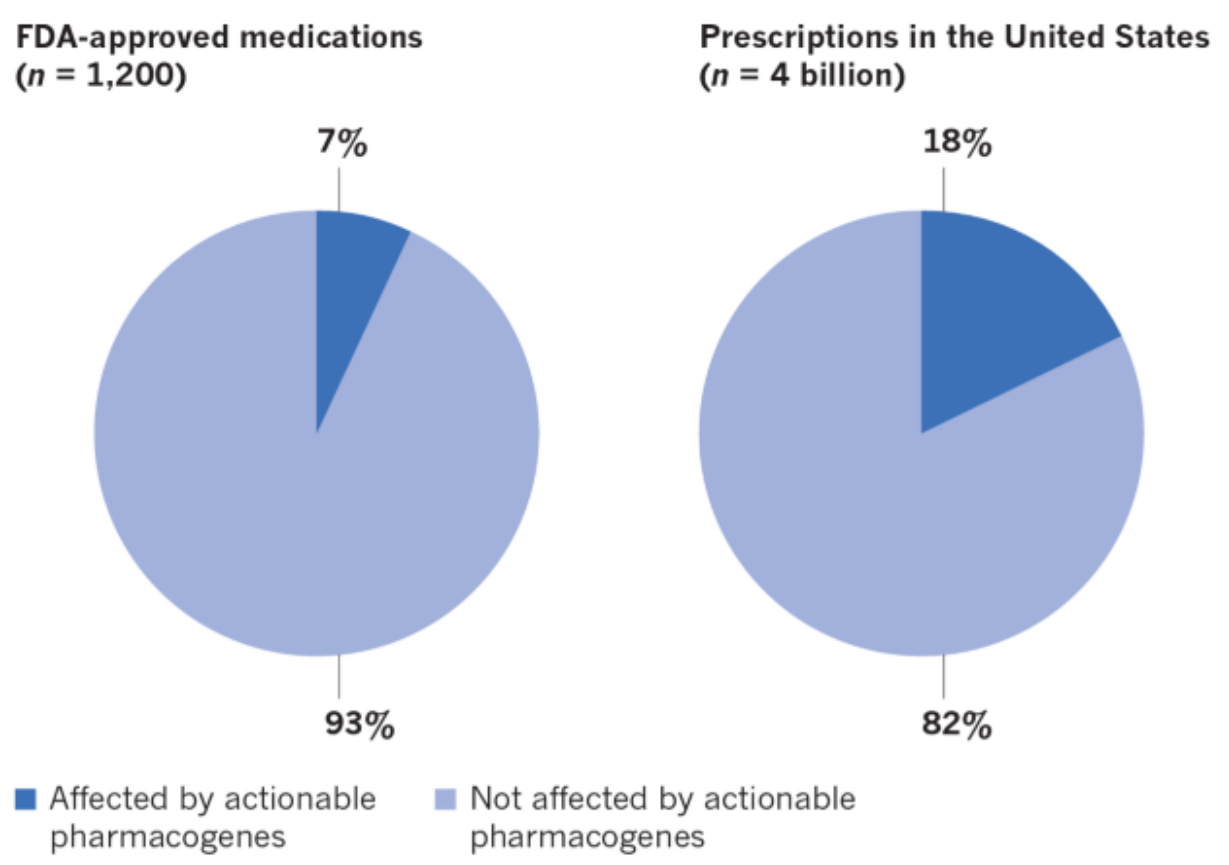
*Nature* (2016)

# Pharmacogenomics: Basic Rationale



- ‘One size does not fit all’
- Stratify patients based on detected genomic variants
- Use genomics to ‘get the right drug to the right patient at the right dose’

# Pharmacogenomics: Getting Real



## Drug:

Abacavir

Allopurinol

Azathioprine

Carbamazepine

Clopidogrel

Methotrexate

Phenytoin

## Used to Treat:

HIV

Gout

Autoimmune disease

Seizures

Blood clots

Autoimmune disease

Seizures

**Bottom Line: Pharmacogenomic testing now appropriate for a small subset of prescription medication, but that subset expected to grow in the future. Meanwhile, efforts to increase clinical usage are ongoing.**



# Genomic Medicine Implementation



**Cancer  
Genomics**



**Rare Genetic Disease  
Diagnostics**



**Noninvasive Prenatal  
Genomic Testing**



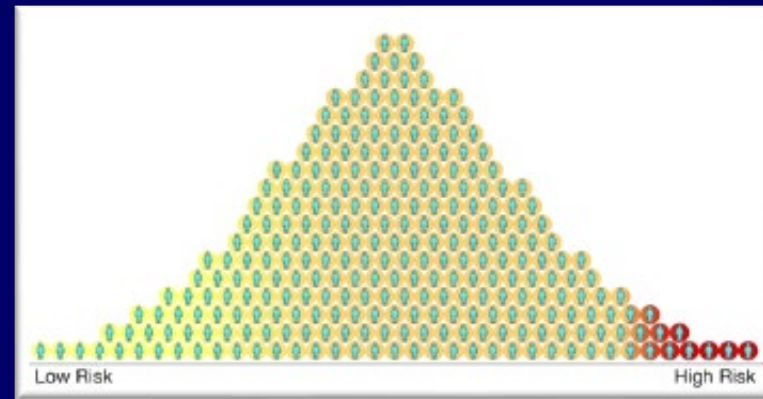
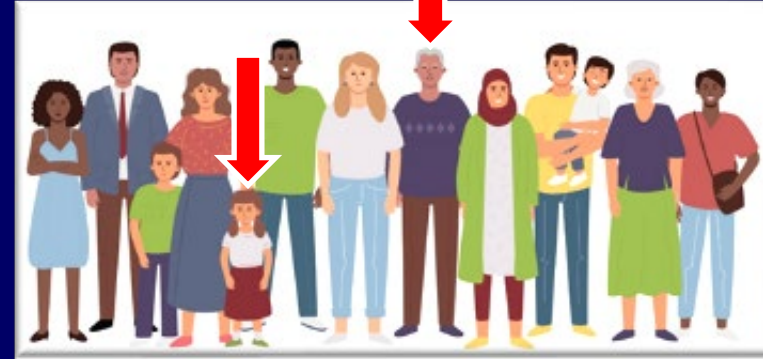
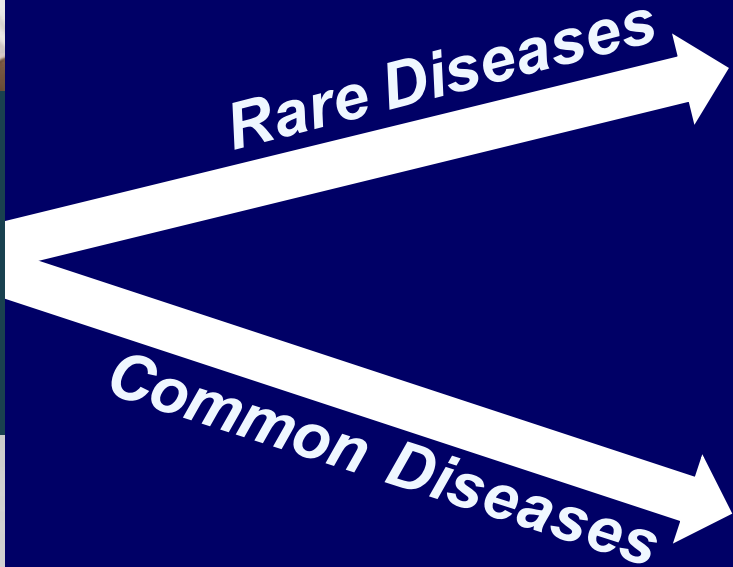
**Pharmacogenomic  
Testing**



**Genomics-based  
Prevention**



# Genomics-based Prevention



*Polygenic Risk Scores*

# En Route to Genomic Medicine

✓ Human Genome Sequenced for First Time  
by the Human Genome Project

✓ Cost of Sequencing a Human Genome  
Reduced >1 Million-Fold

✓ Millions of Human  
Genomes Sequenced

✓ Profound Advances in Understanding  
How the Human Genome Functions

✓ Significant Advances in Unraveling the  
Genomic Bases of Human Disease

✓ Vivid Examples of Genomic Medicine  
Now Emerging





Research  
^  
It's A Journey

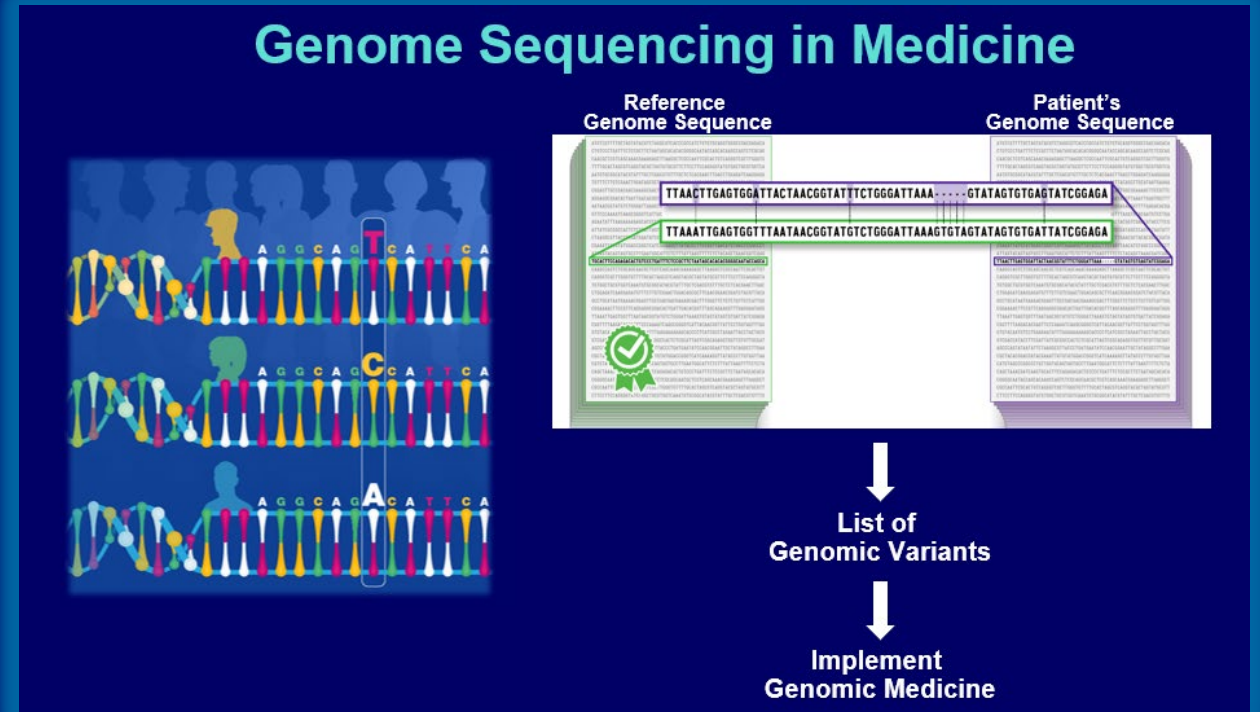




**Challenges  
ahead**

# Challenges of Analyzing a Patient's Genome Sequence

## An Intentional Oversimplification





# Analyzing a Person's Genome Sequence

## 1. Detecting all genomic variants in the generated genome sequence

Required: Reference Genome Sequence (or Pangenome Reference)

Examples: GRCh38.p13 (Build 38) or T2T-CHM13

# Reference vs. Routine Genome Sequences

## Reference genome sequence



## Routine genome sequence

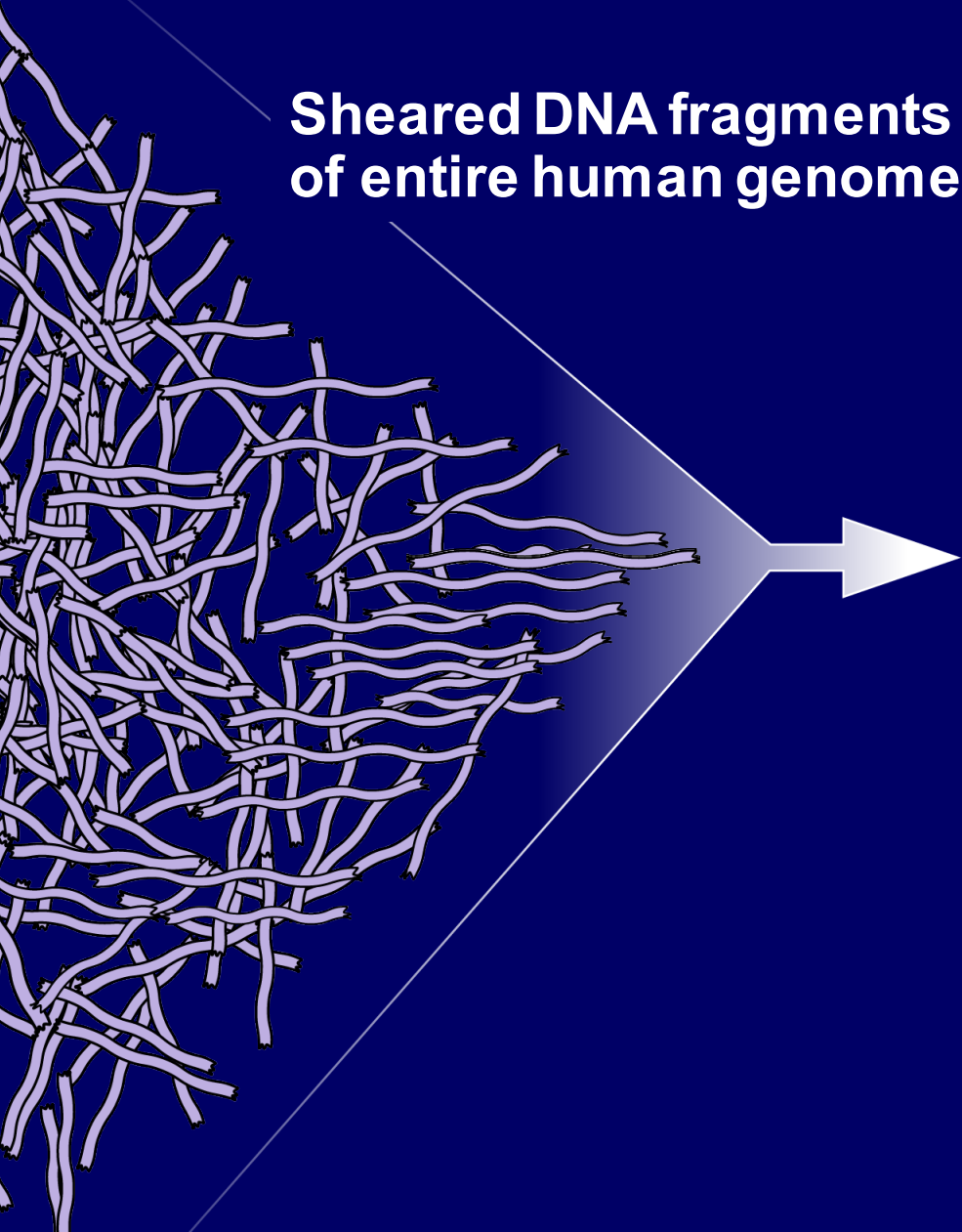


- Purpose: ‘High-quality’ representation
- Meticulously generated
- Multiple DNA-sequencing technologies
- No (or little) missing sequences
- Cost: ~\$10,000

- Purpose: Identify genomic variants
- High-throughput generated
- Single DNA-sequencing technology
- Always has ‘missing’ sequences
- Cost: <\$1,000

# Generating DNA Sequence Reads

Sheared DNA fragments  
of entire human genome

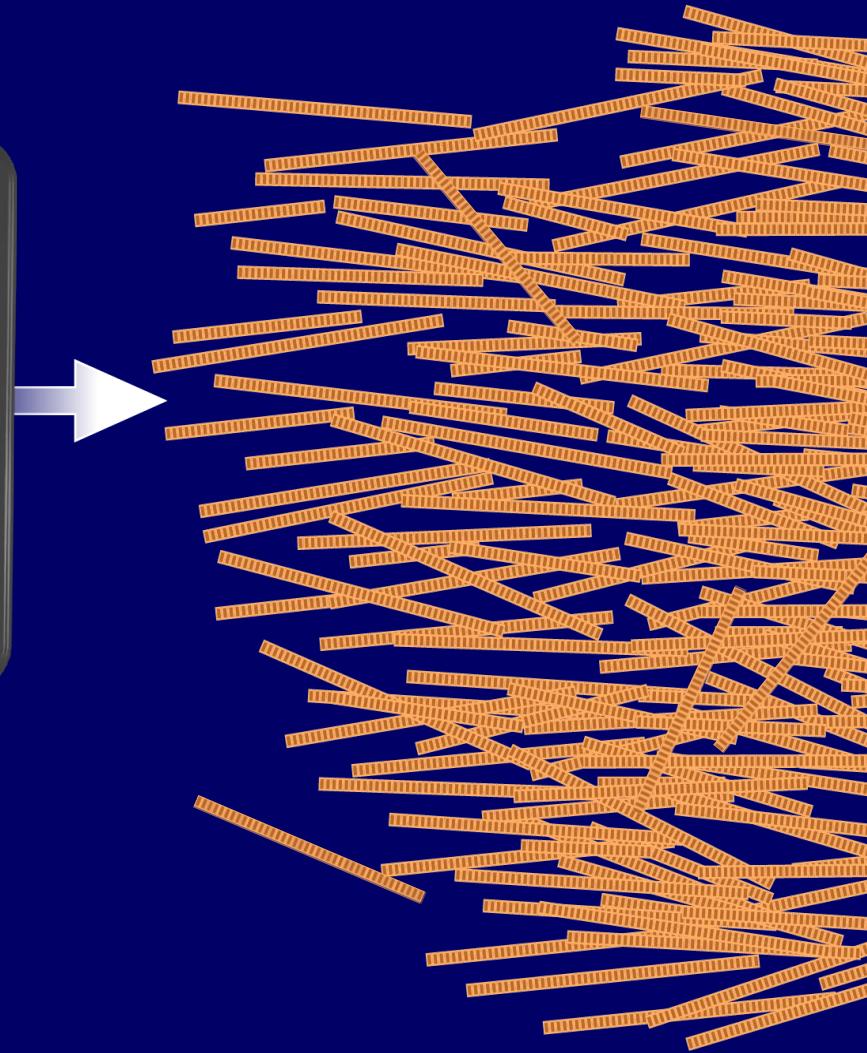


```
GCGGCTAGCGCTAGTGC  
ATATGCTAGCTAGCGGC  
GCGGCTAGATGCGGATG  
GTTATCATAACGGTCAC
```

DNA sequencing  
instrument



Generated DNA  
sequence reads





# Reference-Assisted Genome Sequence Generation



## Reference genome sequence

```
GACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAAC
TGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGTCATTGGCGGAA
A ACTTCGGTTCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTTGAG
GAATAGGTTAAATTGAGTGGTTAATAACGGTATGTCTGGGATTAAGTGTAGT
ATAGTGTGCTTATCGGAGACGGTTTTAAGACACGAGTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCCTGGTAGTTAGGTGTACAATGTCCTGAAGAATAT
TTAAGAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATAC
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
TAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
CAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTATTAGTACAGTAG
TTAACTTGAGTGGATTACTAACGGTATTTCTGGGATTAATGATTGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGGCTCGCC
AATTGCGACTGTCAAGGTCGCTTGGGTGTTTTGCACTAGCGTCAGGTACGCTAG
TATGCGTTCCTCCAGGGGTATGTGGCTGCGTGGTCAAATGTGCGGCATAC
GTATTTGCTCGACGTGTTGCTCTCACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAAACGGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGT
CATTGGCGGAAAAACTCCGTTCAAGGAGGCGGACACTGATTGACACGGTTTAGCA
GAAGGTTTGAGGAATAGGTTAAATTGAGTGGTTAATAACGGTATGTCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTTAAGACACGAGTCCCAA
AATCAAGCGGGGTCATTACAACGGTTATTCCTGGTAGTTAGGTGTACAATGTC
CTGAAGAATATTTAAGAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGG
TTGTGTTGCGATAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAAGGTTATACCCTTGTAGTTAACATGTAGCCCGGCCAT
```

## Sequence read 1

```
CTGAAGAATATTTAAGAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
```



Generated DNA sequence reads

# Reference-Assisted Genome Sequence Generation



## Reference genome sequence

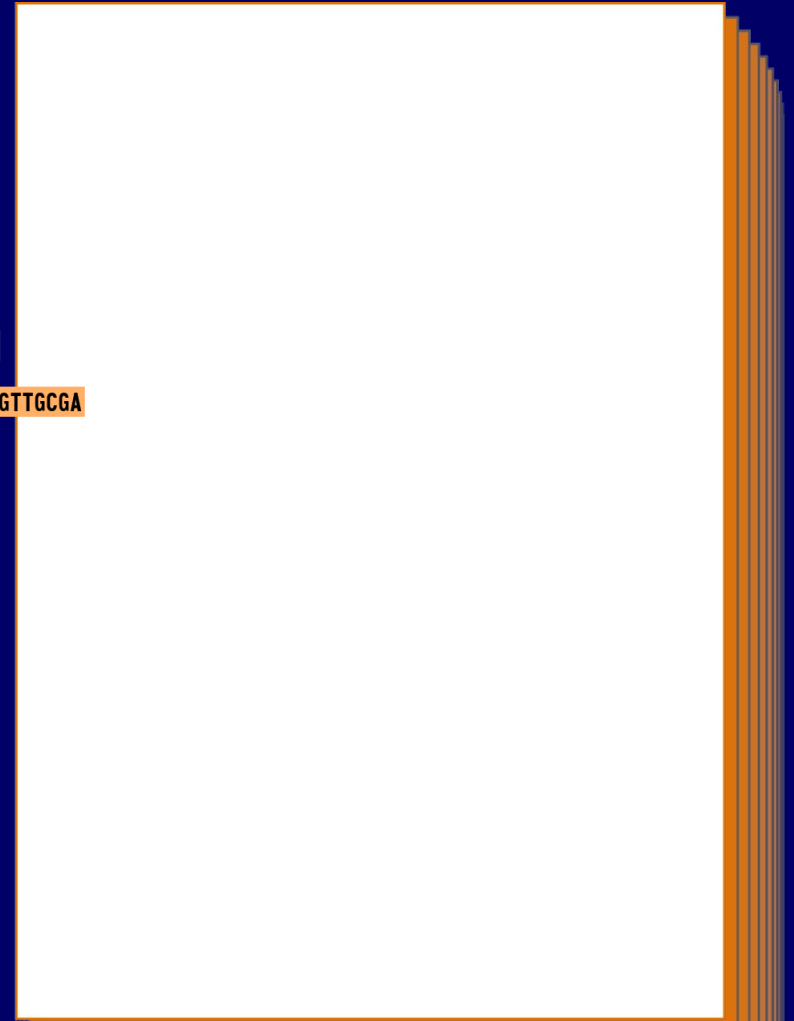
```
GACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAAC
TGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGTCATTGGCGGAA
AACTTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTGAG
GAATAGGTTAAATTGAGTGGTTAATAACGGTATGCTGCGGATTAAGTGTAGT
ATAGTGTGCTTATCGGAGACGGTTTTAAGACACGAGTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCCTGGTAGTTAGGTGTACAATGTCCTGAAGAATAT
TTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATAC
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTGTTGCGA
TAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
CAAAAGGTTCTGAAGAATATTTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTGTTGCGA
TTAACTTGAGTGGATTACTAACGGTATTTCTGGGATTAATGATTGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGGCTCGCC
AATTCGCACTGTCAAGGTCGCTTGGGTGTTTTGCACTAGCGTCAGGTACGCTAG
TATGCGTTCTTCCAGGGGTATGTGGCTGCGTGGTCAAATGTGCGGCATAC
GTATTTGCTCGAGTGTGTTGCTCTCACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAAACGGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGT
CATTGGCGGAAAACTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCA
GAAGGTTTGAGGAATAGGTTAAATTGAGTGGTTAATAACGGTATGCTGCGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTTAAGACACGAGTCCCAA
AATCAAGCGGGGTCATTACAACGGTTATTCCTGGTAGTTAGGTGTACAATGTC
CTGAAGAATATTTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGG
TTGTGTTGCGATAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAAGGTTATACCCTTGTAGTTAACATGTAGCCCGGCCCTAT
```

## Routine genome sequence

Sequence read 1  
CTGAAGAATATTTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTGTTGCGA

## Sequence read 1

CTGAAGAATATTTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTGTTGCGA



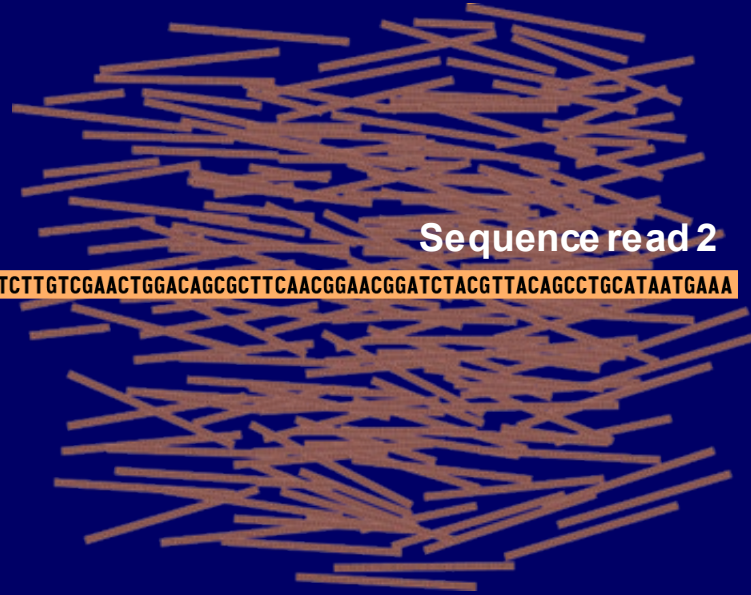
# Reference-Assisted Genome Sequence Generation



## Reference genome sequence

```
GACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAAC
TGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGCATTGGCGGAA
AACTTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGTTTGAG
GAATAGGTTAAATTGAGTGGTTTAATAACGGTATGTCTGGGATTAAGTGTAGT
ATAGTGTGCTTATCGGAGACGGTTTTAAGACACGAGTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCCTGGTAGTTAGGTGTACAATGTCCTGAAGAATAT
TTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATAC
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
TAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCCTTGAACGCTAC: GCTCTCACGAACTTGACCTGGAGATCAAGGAGATGTTCTTGTGCGAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAA
CAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTATTAGTACAGTAG
TTAACTTGAGTGGATTACTAACGGTATTTCTGGGATTAATGATTGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGGCTCGCC
AATTCGCACTGTCAAGGTCGCTTGGGTGTTTTGCACTAGCGTCAGGTACGCTAG
TATGCGTTCCTCCAGGGTATGTGGCTGCGTGGTCAAATGTGCGGCATAC
GTATTTGCTCGACGTGTTT GCTCTCACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCGAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAAACGGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGT
CATTGGCGGAAAACCTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCA
GAAGGTTTGAGGAATAGGTTAAATTGAGTGGTTTAATAACGGTATGTCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTTAAGACACGAGTCCCAA
AATCAAGCGGGGTCATTACAACGGTTATTCCTGGTAGTTAGGTGTACAATGTC
CTGAAGAAATTTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGG
TTGTGTTGCGATAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAGGTTATACCCTTGTAGTTAACATGTAGCCCGGCCCTAT
```

## Generated sequence reads



Sequence read 2

## Routine genome sequence

```
CTGAAGAATAT
TTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATAC
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
```



# Reference-Assisted Genome Sequence Generation



## Reference genome sequence

```
GACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAAC
TGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGCATTGGCGGAA
AACTTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTGAG
GAATAGGTTAAATTGAGTGGTTAAATAACGGTATGCTCTGGGATTAAGTGTAGT
ATAGTGTGCTTATCGGAGACGGTTTTAAGACACGAGTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCCTGGTAGTTAGGTGTACAATGTCCTGAAGAATAT
TTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATA
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
TAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
```

TAACATGTAGCCGGCCCTATTAGTACAGTAGTTAACTTTAAACATGTAGCCGGCCCTATTAGTACAGTAGTATTGATTGTATAGTGTCAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAG

TTAACTTGGTGGATTACKAACCGTATTTCTGGGATTAATGATTGTATAGTGT  
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGCTCGCC

```
AATTCGCACTGTCAAGGTCGCTTGGGTGTTTTGCACTAGCGTCAGGTACGCTAG
TATGCGTTCCTCCAGGGGTATGTGGCTGCGTGGTCAAATGTGCGGCATAC
GTATTTGCTCGACGTGTTTGTCTCACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAAACGGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGT
CATTGGCGGAAAACTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCA
GAAGGTTTGGGAATAGGTTAAATTGAGTGGTTAAATAACGGTATGCTCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTTAAGACACGAGTCCCAA
AATCAAGCGGGGTCATTACAACGGTTATTCCTGGTAGTTAGGTGTACAATGTC
CTGAAGAAATTTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGG
TTGTGTTGCGATAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTAT
```

## Generated sequence reads



Sequence read 3

## Routine genome sequence

```
CTGAAGAATAT
TTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATA
CCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
```

```
GCTCTACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAA
```

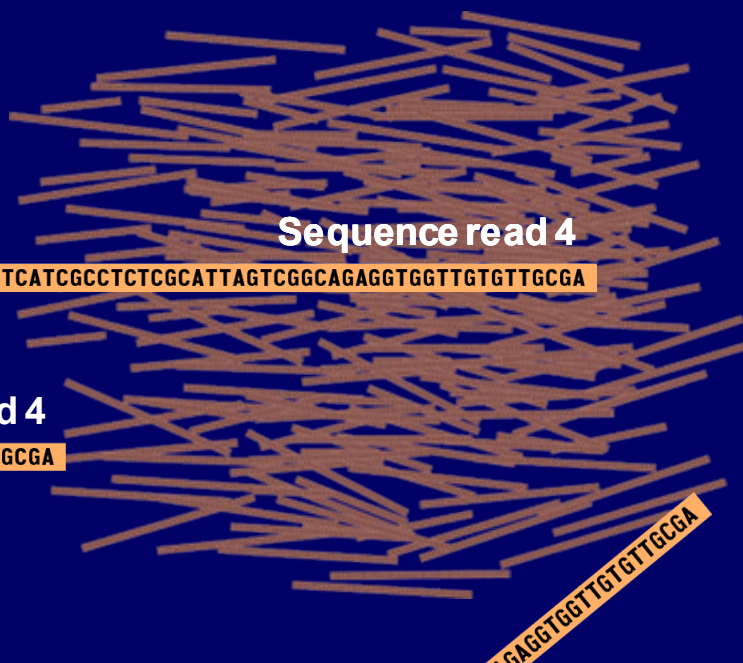
# Reference-Assisted Genome Sequence Generation



## Reference genome sequence

```
GACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAC
TGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGTCATTGGCGGAA
AACTTCGGTTCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTGAG
GAATAGGTTAAATTGAGTGGTTAATAACGGTATGTCTGGGATTAAGTGTAGT
ATAGTGTGCTTATCGGAGACGGTTTAAAGACACGAGTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCTGGTAGTTAGGTGTACAATGCCTGAAGAATAT
TTAAGAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATA
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
TAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGTCAT
CAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTATTAGTACAGTAG
TTAACTTGAGTGGATTACTAACGGTATTTCTGGGATTAATGATTGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAAGGCTGCG
AATTCCACTGTCAAGGCTCGCTTGGGCTTTTTCAGTACCGTCAAGTACCGTAG
TATGCGTTCTTCTCCAGGGGATGTGGCTGCGTGGTCAAATGTGCGGCATAC
GTATTTGCTCGACGTGTTTGTCTCACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAACGGAGTGGCCAGCAGGAAAGCGACTTTGGGTTCTGTCTGTTGT
CATTGGCGGAAAACCTCCGTTCAAGGAGGCGGACACTGATTGACACGGTTTAGCA
GAAGGTTTGAAGGAAAGGTTAAATTGAGTGGTTAATAACGGTATGTCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTAAAGACAGGTTCCCAA
AATCAAGCGGGGTCATTACAACGGTTATTCTGGTAGTTAGGTGTACAATGTC
CTGAAGAAATTTAAGAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGG
TTGTGTTGCGATAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAGGTTATACCCTTGTAGTTAACATGTAGCCCGGCCAT
```

## Generated sequence reads



Sequence read 4

GCACCCCTCATCGCCTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA

Sequence read 4

GCACCCCTCATCGCCTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA

Sequence read 4

No Match

## Routine genome sequence

```
CTGAAGAATAT
TTAAGAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATA
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA

TTAACATGTAGCCGGCCCTATTAGTACAGTAG
TTAACTTGGTGGATTACAAACGGTATTTCTGGGATTAAGGTTAGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAAG

GCTCTACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAA
```

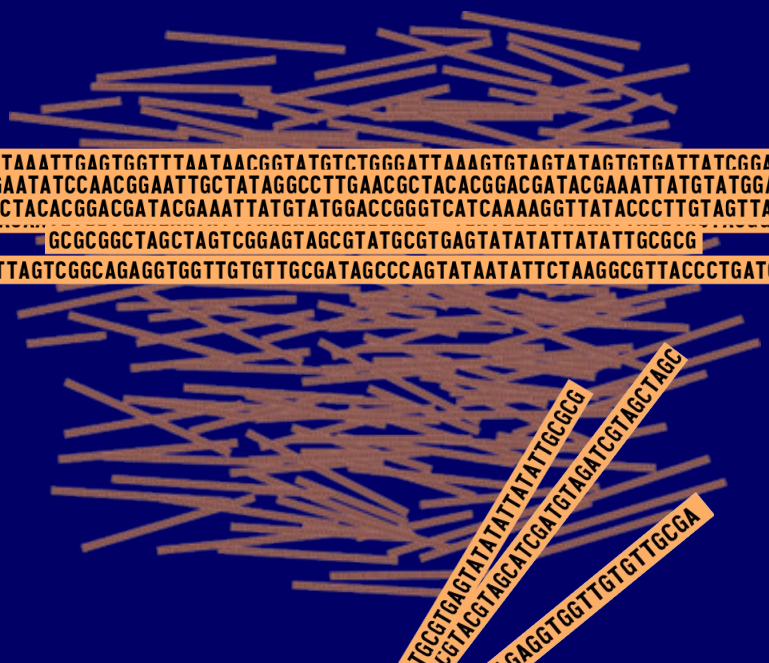
# Reference-Assisted Genome Sequence Generation



Reference genome sequence

```
ACTGGACAGCGCTTCAACGGAAACGGATCTACGTTACAGCCTGCATAATGAAAAC
GGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGCATTGGCGGAA
AACTTCCGTTCCAGGAGCGGACACTGATTGACACGGTTAGCAGAAAGGTTTGAG
GAATAGGTTAAATTGAGTGGTTAATAACGGTATGCTGGGATTAAGTGTAGT
ATAGTGTGATTATCGGAGACGGTTTAAAGACAGGATCCCAAAATCAAGCGGG
GTCATTACAACGGTTATTCCTGGTAGTTAGGTGTACAATGTTCTGAAGAATAT
TTAAGAAAAAAGCACCCTCATAGCCAGTATAATTTCTAAGGCGTTACCCCTGATGAAT
CTTCGATTATCGCGGCCACTGCTCGCCAAATTCGCAGGTTAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
TAGCCAGTATAATTTCTAAGGCGTTACCCCTGCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTGTTGCGATAGCCAGTATAATTTCTAAGGCGTTACCCCTGATGAATATCCAACGGAATTGCTAT
TATAGGCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
CAAAAGGTTATACCCCTTGTAGTTAACATGTAGCCCGGCCCTATTAGTACAGTAG
TTAACTTGAGTGGATTACTAACGGTATTTCTGGGATTAATGATTGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAAGCTCGCC
AATTCGCAGTGTGAGGGTCCGTTGGGTGTTTGCAGTACGCTCAGGTACGCTAG
TATGCGTTCTTCCCTCCAGGGTATGTGGCTGCGTGGTCAAATGTGCGGCATAC
GTATTGCTCGACGTTTGTCTCAGCAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAACGGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGT
CATTGGCGGAAAACCTCCGTTCCAGGAGCGGACACTGATTGACACGGTTTAGCA
GAAAGTTTGGAGGAATAGGTTAAATTGAGTGGTTAATAACGGTATGCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTAAAGACAGGATCCCAA
AATCAAGCGGGGTCATTACAACGGTTATTCCTGGTAGTTAGGTGTACAATGTC
CTGAAGAATATTTAAGAAAAAAGCACCCTCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTGCATTATCGCGGCCACTCTCGCATTAGTGGCAGAGGTGG
TTGTGTTGGGATAGCCAGTATAATTTCTAAGGCGTTACCCCTGATGAATATCC
AACGGAATTGCTATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAAGGTTATACCCTTGTAGTTAACATGTAGCCCGGCCCTAT
```

Generated sequence reads



Routine genome sequence

```
CTGAAGAATAT
TTAAGAAAAAAGCACCCTCATAGCCAGTATAATTTCTAAGGCGTTACCCCTGATGAATATCCAACGGAATTGCTAT
CTTCGATTATCGCGGCCACTGCTCGCCAAATTCGCAGGTTAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCATCAAAAGGTTATACCCTTGTAGTAATTACCTACTACGGTCGACCATAC
GCGCGGCTAGCTAGTCGGAGTAGCGTATGCGTGAGTATATATTATTGCGCG
TAAACATGTAGCCCGGCCCTATTAGTACAGTAG
TTAACTTGGTGGATTACAAACGGTATTTCTGGGATTAAGGATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAAG
GCTCTCAGCAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAA
```



# Reference-Assisted Genome Sequence Generation



## Reference genome sequence

```
GACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAAC
TGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGCATTGGCGGAA
AACTTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTGAG
GAATAGGTTAAATTGAGTGGTTTAAATAACGGTATGTCTGGGATTAAGTGTAGT
ATAGTGTGCTTATCGGAGACGGTTTTAAGACACGAGTTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCTGGTAGTTTAGGTGTACAATGTCCTGAAGAATAT
TTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATA
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
TAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
CAAAAGGTTATACCCTTGTAGTTAACATGTAGCCCGGCCCTATTAGTACAGTAG
TTAACTTGAGTGGATTACTAACGGTATTCTGGGATTAATGATTGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGGCTCGCC
AATTCGCACTGTCAAGGTCGCTTGGGTGTTTTGCACTAGCGTCAGGTACGCTAG
TATGCGTTCTTCCAGGGGTATGTGGCTGCGTGGTCAAATGTGCGGCATAC
GTATTTGCTCGACGTGTTGCTCTCACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAAACGGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGT
CATTGGCGGAAAACTTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCA
GAAGGTTTGAAGTAAGGTTAAATTGAGTGGTTTAAATAACGGTATGTCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTTAAGACACGAGTTCCCAA
AATCAAGCGGGGTCATTACAACGGTTATTCTGGTAGTTTAGGTGTACAATGTC
CTGAAGAATATTTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGG
TTGTGTTGCGATAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAAGGTTATACCCTTGTAGTTAACATGTAGCCCGGCCCTAT
```

Each line of a routine genome sequence is read and localized many times to ensure its accuracy – often more than 30 times!

## Routine genome sequence

```
ACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAAC
GGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGCATTGGCGGAA
AACTTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTGAG
GAATAGGTTAAATTGAGTGGTTTAAATAACGGTATGTCTAGGATTAAGTGTAGT
ATAGTGTGATTATCGGAGACGGTTTTAAGACACGAGTTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCTGGTAGTTTAGGTGTACAATGTCCTGAAGAATAT
TTAAGAAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATA
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
TAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
CAAAAGGTTATACCCTTGTAGTTAACATGTAGCCCGGCCCTATTAGTACAGTAG
TTAACTTGCGTGGATTACAAACGGTATTCTGGGATTA...GTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGGCTCGCC
AATTCGCACTGTCAAGGTCGCTTGGGTGTTTTGCACTAGCGTCAGGTACGCTAG
TATGCGTTCTTCCAGGGGTATGTGGCTGCGTGGTCAAATGTGCGGCATAC
GTATTTGCTCGACGTGTTGCTCTCACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAAACGGAGTTGCCGACGACGAAAGCGAC-TTGGGTTCTCTGTTGT
CATTGGCGGAAAACTTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCA
GAAGGTTTGAAGTAAGGTTAAATTGAGTGGTTTAAATAACGGTATGTCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTTAAGA
AGTTTAGGTGTACAATGTC
CTGAAGAATATTTAAGAGAAAAGCACC--TCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGG
TTGTGTTGCGATAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAAGGTTATACCCTTGTAGTTAACATGTAGCCCGGCCCTAT
```

# Comparing the Two Genome Sequences



## Reference genome sequence

```
GACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAAC
TGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGCATTGGCGGAA
AACTTCCGTTTCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTGAG
GAATAGGTTAAATTGAGTGGTTTAATAACGGTATGTCTGGGATTAAGTGTAGT
ATAGTGTGCTTATCGGAGACGGTTTTAAGACACGAGTTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCTGGTAGTTAGGTGTACAATGCTCTGAAGAATAT
TTAAGAAAAAGCACCCTCATCGCCTAGAATTACCTACTACGGTCGACCATAC
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
TAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
CAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTATTAGTACAGTAG
TTAACTTGAGTGGATTACTAACGGTATTTCTGGGATTAATGATTGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGGCTCGCC
AATTCGCACTGTCAAGGTCGCTTGGGTGTTTTGCACTAGCGTCAGGTACGCTAG
TATGCGTTCTTCCCTCCAGGGGTATGTGGCTGCGTGGTCAAATGTGCGGCATAC
GTATTTGCTCGACGTGTTTGTCTCACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAAACGGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGT
CATTGGCGGAAAACCTCCGTTTCAGGAGGCGGACACTGATTGACACGGTTTAGCA
GAAGGTTTGAGGAATAGGTTAAATTGAGTGGTTTAATAACGGTATGTCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTTAAGACACGAGTTCCCAA
AATCAAGCGGGGTCATTACAACGGTTATTCTGGTAGTTAGGTGTACAATGTC
CTGAAGAATATTTAAGAAAAAGCACCCTCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGG
TTGTGTTGCGATAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTAT
```

## Compare sequences to find differences

## Routine genome sequence

```
ACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAAC
GGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGCATTGGCGGAA
AACTTCCGTTTCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTGAG
GAATAGGTTAAATTGAGTGGTTTAATAACGGTATGTCTAGGATTAAGTGTAGT
ATAGTGTGATTATCGGAGACGGTTTTAAGACACGAGTTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCTGGTAGTTAGGTGTACAATGCTCTGAAGAATAT
TTAAGAAAAAGCACCCTCATCGCCTAGAATTACCTACTACGGTCGACCATAC
CTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
TAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
CAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTATTAGTACAGTAG
TTAACTTGCGTGGATTACAAACGGTATTTCTGGGATTA...GTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGGCTCGCC
AATTCGCACTGTCAAGGTCGCTTGGGTGTTTTGCACTAGCGTCAGGTACGCTAG
TATGCGTTCTTCCCTCCAGGGGTATGTGGCTGCGTGGTCAAATGTGCGGCATAC
GTATTTGCTCGACGTGTTTGTCTCACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAAACGGAGTTGCCGACGACGAAAGCGAC-TTGGGTTCTCTGTGTTGT
CATTGGCGGAAAACCTCCGTTTCAGGAGGCGGACACTGATTGACACGGTTTAGCA
GAAGGTTTGAGGAATAGGTTAAATTGAGTGGTTTAATAACGGTATGTCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTTAAGA
AGTTTAGGTGTACAATGTC
CTGAAGAATATTTAAGAGAAAAGCACC--TCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTCGATTATCGCGGCCACTCTCGCATTAGTCGGCAGAGGTGG
TTGTGTTGCGATAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTAT
```



# Detecting Human Genomic Variants



Reference genome sequence

```
GACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAAC
TGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGCATTGGCGGAA
AACTTCGGTTCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTGAG
GAATAGGTTAAATTGAGTGGTTTAAATAACGGTATGCTGGGATTAAGTGTAGT
ATAGTGTGCTTATCGGAGACGGTTTAAAGACACGAGTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCTGGTAGTTAGGTGTACAATGCCTGAAGAATAT
TTAAGAAAAAGCACCCTCATCGCCTAGAATTACCTACTACGGTCGACCATA
CTTCGATTATCGCGCCACTCTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
TAGCCAGTATAATTTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
CAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTATTAGTACAGTAG
TTAACTTGAAGTGGATTACTAACGGTATTCTGGGATTAATGATTGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGGCTCGCC
AATTGCGACTGTCAGGGTCGCTTGGGTGTTTTGCACTAGCGTCAGGTACGCTAG
TATGCGTCTTCCCTCCAGGGTATGTGGCTGCGTGGTCAAAATGTCGGCATA
GTATTTGCTCGACGTGTTGCTCTCAGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGC
ATAATGAAAACGGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGCTGTTGT
CATTGGCGGAAAACCTCCGTTCAGGAGGCGGACACTGATTGACACGGTTTAGCA
GAAGGTTTGAAGTAAGGTTAAATTGAGTGGTTTAAATAACGGTATGCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTAAAGACAGAGTCCCAA
AATCAAGCGGGTCATTACAACGGTTATTCTGGTAGTTAGGTGTACAATGTC
CTGAAGAATATTAAGAAAAGCACCCTCATCGCCTAGAATTACCTACTACG
GTCGACCATACTTCGATTATCGCGCCACTCTCGCATTAGTCGGCAGAGGTGG
TTGTGTTGCGATAGCCAGTATAATTTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTAT
```

Genomic variants

Yellow vs. Red



Routine genome sequence

```
ACTGGACAGCGCTTCAACGGAACGGATCTACGTTACAGCCTGCATAATGAAAAC
GGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGTCTGTTGCATTGGCGGAA
AACTTCGGTTCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTGAG
GAATAGGTTAAATTGAGTGGTTTAAATAACGGTATGCTGGGATTAAGTGTAGT
ATAGTGTGCTTATCGGAGACGGTTTAAAGACACGAGTCCCAAATCAAGCGGG
GTCATTACAACGGTTATTCTGGTAGTTAGGTGTACAATGCCTGAAGAATAT
TTAAGAAAAAGCACCCTCATCGCCTAGAATTACCTACTACGGTCGACCATA
CCATCGCCTAGAATTACCTACTACTCGCATTAGTCGGCAGAGGTGGTTGTTGCGA
TAGCCAGTATAATTTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
CAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTATTAGTACAGTAG
TTAACTTGAAGTGGATTACTAACGGTATTCTGGGATTAATGATTGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGGCTCGCC
AATTGCGACTGTCAGGGTCGCTTGGGTGTTTTGCACTAGCGTCAGGTACGCTAG
TATGCGTCTTCCCTCCAGGGTATGTGGCTGCGTGGTCAAAATGTCGGCATA
GCGTGGTCAAAATGTCGGCATAACGTTATTTGCTCGACGTGTTGCTCTCACGAAC
TTGACCTGGAGATCAAGGAGATGTTTCTTGTGCAACTGGACAGCGCTTCAACGG
AACGGTCTACGTTACAGCCTGCATAATGAAAACGGAGTTGCCGACGACGAAAG
CGACTGGGTTCTGTTGTTGTCATTGGCGGAAAACCTCCGTTCAAGGAGCGG
ACACTGATTGACACGGTTTAGCAGAAGGTTTGAAGTAAGTTAAATTGAGTGG
TTAATAACGGTATGCTGGGATTAAGTGTAGTATAGTGTGATTATCGGAGAC
GGTTTTAAGAAGTTTAGGTGTACAATGTCTGAAAGAATATTTAAGAAAAGCACCCTC
ATCGCCTAGAATTACCTACTACGGTCGACCATACTTCGATTATCGCGGCCACT
CTCGCATTAGTCGGCAGCAGCAGAGGTGGTTGTTGCGATAGCCAG
TATAATTTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGCTATAGGCC
TTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCATCAAAGGT
```



# Human Diversity in Reference Genome Sequences



## Reference genome sequence

genome

```
GACAGCGCTTCAACGGAAACGGATCTACGTTACAGCCTGCATAATGAAAC
TGCCGACGACGAAAGCGACTTTGGGTTCTGCTGTTGTCATTGGCGGAA
AACTTCGTTCCAGGAGGCGGACACTGATTGACACGGTTTAGCAGAAGGTTGAG
GAATAGGTTAAATTGAGTGGTTTAATAACGGTATGCTGGGATTAAGTGTAGT
ATAGTGTGCTTATCGGAGACGGTTTTAAGACACGAGTCCCAAATCAAGCGGG
GTCATTACAACGGTTATCTCGGTAGTTAGGTGACAATGCTCTGAAGAATAT
TTAAGAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACGGTCGACCATA
CTTCGATTATCGGGCCACTCTCGATTAGTCCGAGAGGTGGTTGTTGCGA
TAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCCAACGGAATTGC
TATAGGCCTTGAACGCTACACGGACGATACGAAATTATGTATGGACCGGGTCAT
CAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTATTAGTACAGTAG
TTAACTTGAGTGGATTACTAACGGTATTTCTGGGATTAATGATTGTATAGTGT
CAAGCCAGTCTCGCAGCAACGCTCGTCAGCAAACGAAAGAGCTTAAGGCTCGCC
AATTCGCACTGTCAGGGTCGCTTGGGTGTTTTGCACTAGCGTACGGTACGCTAG
TATGCGTTCTTCTCCAGGGGATGTTGGCTGCGTGGTCAAATGTCGGCATA
GTATTTGCTCGACGGTGGTCTCACGAACTTGACCTGGAGATCAAGGAGATG
TTTCTTGTGCAACTGGACAGCGCTCAACGGAAACGGATCTACGTTACAGCCTGC
ATAATGAAACGGAGTTGCCGACGACGAAAGCGACTTTGGGTTCTGCTGTTGT
CATTGGCGGAAAACCTCCGTTCCAGGAGGCGGACACTGATTGACACGGTTAGCA
GAAGGTTTGGGAAATAGGTTAAATTGAGTGGTTTAAACGGTATGCTGGGAT
TAAAGTGTAGTATAGTGTGATTATCGGAGACGGTTTTAAGACACGAGTCCCAA
AATCAAGCGGGGTCATTACAACGGTTATCTCGGTAGTTAGGTGACAATGTC
CTGAAGAATTTAAGAAAAAGCACCCCTCATCGCCTAGAATTACCTACTACG
GTCGACCATACCTTCGATTATCGGGCCACTCTCGATTAGTCCGAGAGGTGG
TTGTTGCGATAGCCAGTATAATATTCTAAGGCGTTACCCTGATGAATATCC
AACGGAATTGCTATAGGCTTGAACGCTACACGGACGATACGAAATTATGTATG
GACCGGGTCATCAAAAGGTTATACCCTTGTAGTTAACATGTAGCCGGCCCTAT
```



Reference genome sequence 3

Reference genome sequence 4

Reference genome sequence 5

Reference genome sequence 5

Reference genome sequence 6

Reference genome sequence 6

Reference genome sequence 7

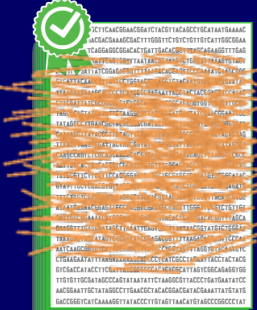
Reference genome sequence 7

Reference genome sequence 8

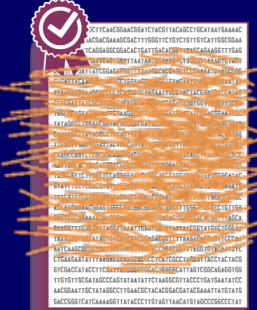
Reference genome sequence 8

# No Single Reference Genome Sequence is Ideal

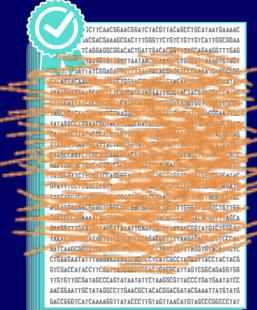
Reference genome sequence 1



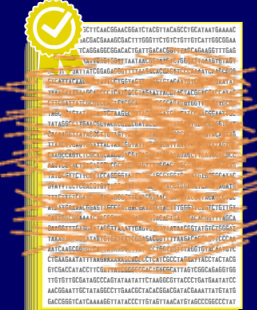
Reference genome sequence 2



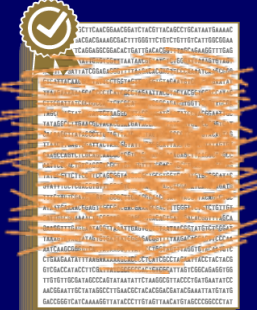
Reference genome sequence 3



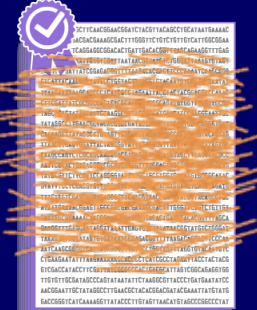
Reference genome sequence 4



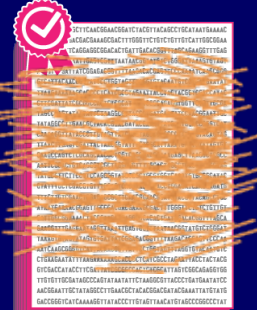
Reference genome sequence 5



Reference genome sequence 6



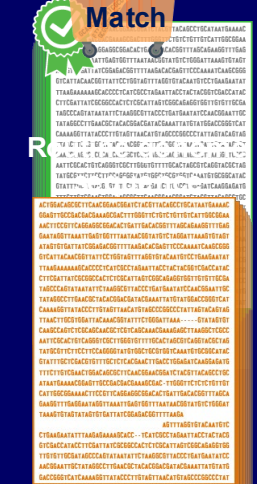
Reference genome sequence 7



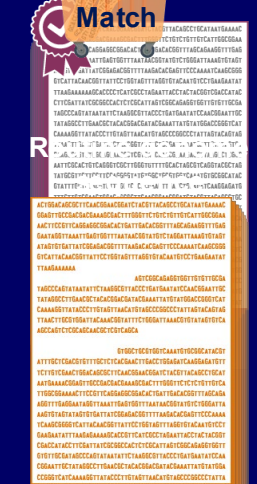
Reference genome sequence 8



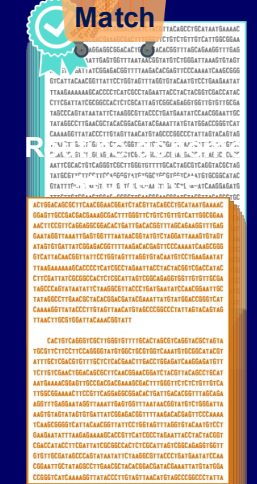
Reference genome sequence 1



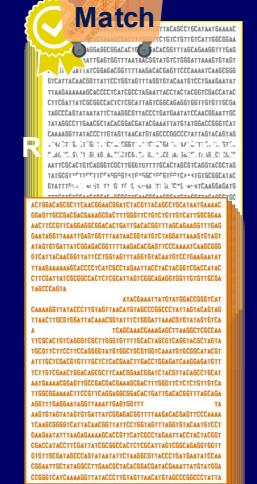
Reference genome sequence 2



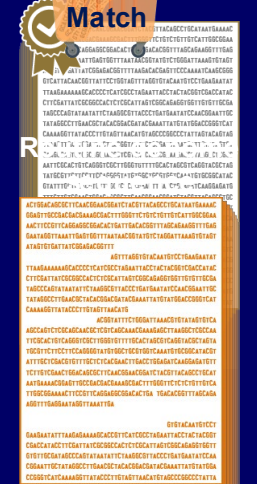
Reference genome sequence 3



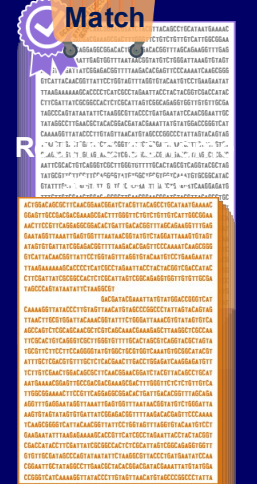
Reference genome sequence 4



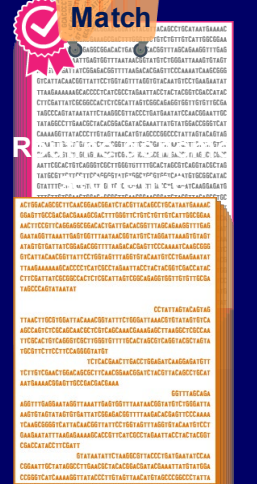
Reference genome sequence 5



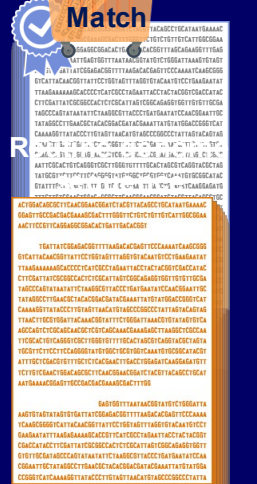
Reference genome sequence 6



Reference genome sequence 7



Reference genome sequence 8





# Human Pangenome Reference



- **Composite of multiple human reference genome sequences**
- **Captures the breadth of human genomic variation much better than any one human reference genome sequence**
- **Enables more accurate and complete detection of genomic variants across diverse human populations**



# Using a Human Pangenome Reference



Human pangenome reference

Routine genome sequence

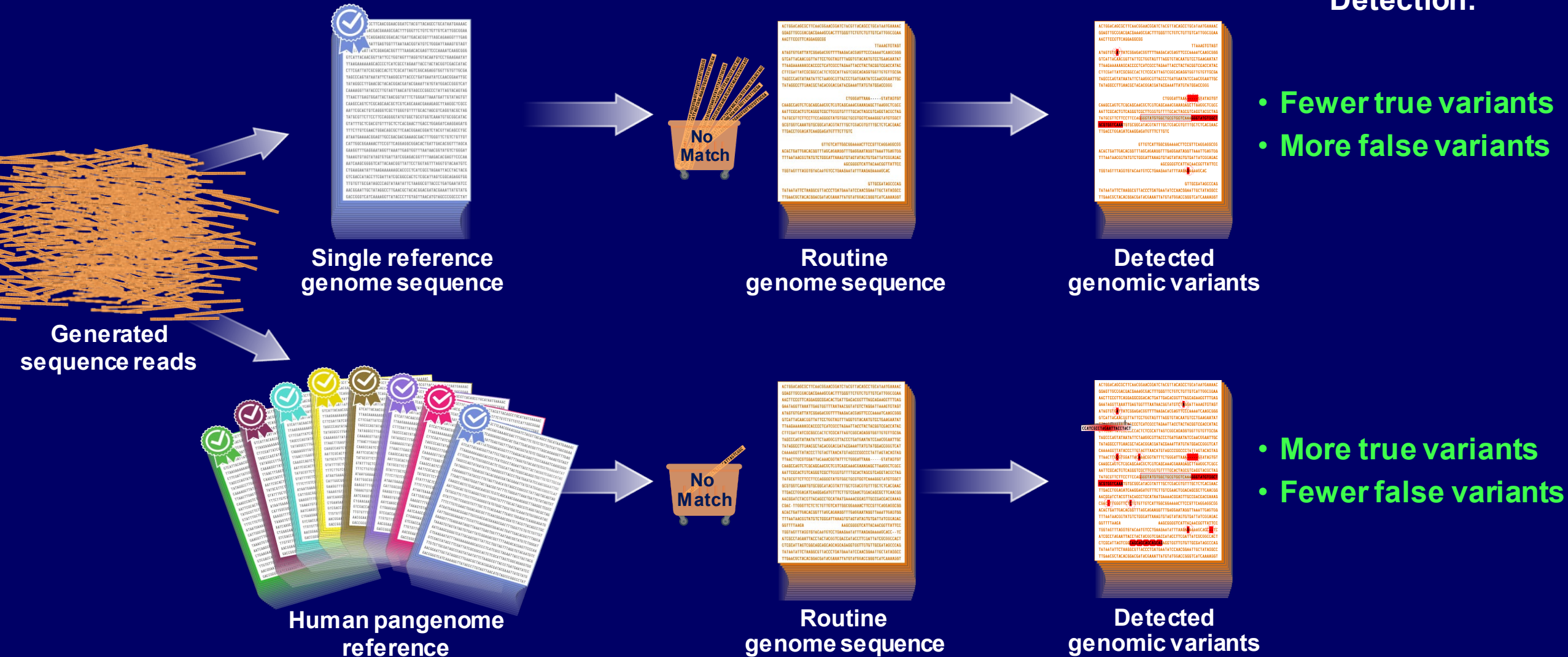


```
ACTGCACAGCGCTTCAACGACGATCTACOTTACACCTCCATTAATAAAC
GGAGTTCGCGACGACGACAGGCACTTTGGTCTCTCTCTTTGATTCAGCGGGA
AACTCCGTTCCAGAGGGGACACTGATGACAGGTTTACGACAAGGTTTGA
GAATAGGTTAAATGAGTGGTTTAATACGGTATGCTAGGATAAAGTGTAG
ATAGTGGATTATCGGAGAGCGGTTTAAAGCAGAGTCCCAAAATCAAGCGG
GTCATTACAGCGGTTATCTGGTGGATTAGGTGACAACTGCTGAAGATAT
TTAAAGAAAAGACCCCTCAGCCCTAGAAATACCTACTACGGTCACCAT
CTCGATTATCGGCGACTCTGCCATTAGTCGGCAGAGGTTGGTGTGGCGA
TACCGCAGTATAAATATCTAAGCGCTACCTGTAGTGAATATCCACGGAAATGC
TATACGCTTACAGCTCACACGACATACAAATATATATGAGCCGGTACAT
CAAAAGTTATACCTCTGTAGTAACTATGACCTGACCGCCCTAATATACAGT
TTAATCTGGTGGATTACAAAGGATTTCTGGGATTAACGTTATATGTCAC
AGCCACTGCGACGCAAGCTGCTGACGAAAGGAGAGCTTAAGGTCGCCAA
TTCGCACTGTCAGGTCCTGGTGGTTTGCAGTACGGTCAGGTACCGTAGTA
TGGCTTCTTCTCCAGGGATATGCGCTGGTGGTCAATGTCGGCATACGT
ATTGCTCGAGGTTTGGCTCACGAACTTACCTGGAGATCAAGAGATGTT
TCTGTGCACTGACGCGCTTCAAGGGACGATCATGTTACAGCCCTGCAT
AAATGAACGGAGTTCCGCGACGACAAAGGCACTTGGGTTCTCTCTGTTCA
TTGGCG GCGACACTTATACAGGTTTACGAG
AGGTTGAGGAAATAGTTAAATGAGTGGTTTAAAGCGTATGCTGGGATTA
AAGTGTAGTATAGTGTGATTCGAGACGGTTTTAAGCAGAGTCTCCAAA
TCAAGCGGGTCTACAGCGGTTATCTGGTGGATTAGGTGACATGCTCT
GAAGAAATTTAAGAGAAAGACCGCTTACCGCTAGAAATACCTACTACGGT
CGACCATACCTTCAATATCGCGGCACTCTCGCATAGTCGGCAGAGGTTGTT
GTTTGGATACCGCAGTATAAATTTAAGGGTACCTTACCTGAATATCAAC
CGGATGCTATAGCCCTTGAAGCTCACGGACGATACGAAATATGATGGA
CCGGTCAACAAAGTATACCTTGTAGTTAAGATGAGCCCGCCCTATTA
```

More complete routine genome sequence

# Improved Genomic Variant Detection Using a Human Pangenome Reference

Genomic Variant Detection:





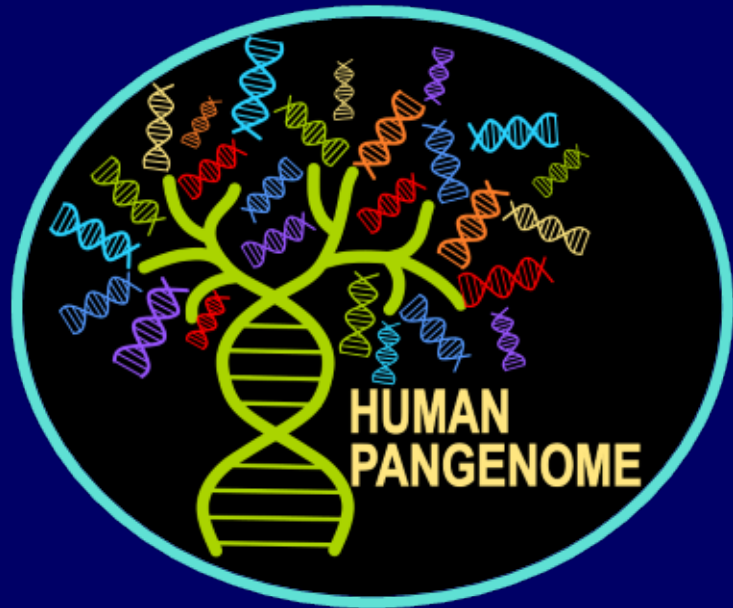
# NHGRI's Human Genome Reference Program

**NIH** National Human Genome Research Institute

Home / Research Funding / Funded Programs and Projects / Human Genome Reference Program

## Human Genome Reference Program

The human genome reference is used by essentially all researchers who need to align and assemble experimental or patient genome sequence data. It also serves as a consensus coordinate system for reporting results.



The international journal of science / 11 May 2023

# nature

The cover of the journal Nature, dated 11 May 2023. It features a central graphic of a globe with multiple, overlapping, colorful DNA strands (in shades of red, orange, yellow, green, and blue) wrapped around it, representing genetic diversity. The title "nature" is prominently displayed at the top in a large, black, serif font. Below the globe, the words "HUMAN PANGENOME" are written in a large, bold, black, sans-serif font. Underneath that, a subtitle reads: "Data from 47 individuals combine to create reference resource that reflects human diversity".

### Article

## A draft human pangenome reference

<https://doi.org/10.1038/s41586-023-05896-x>

Received: 9 July 2022  
Accepted: 28 February 2023  
Published online: 10 May 2023

Open access  
[Check for updates](#)

Wen-Wei Liao<sup>1,3,8,9</sup>, Mobin Aari<sup>4,10</sup>, Jana Ebler<sup>1,6,9</sup>, Daniel Doerflinger<sup>6,9</sup>, Marina Haukness<sup>4</sup>, Glenn Hickey<sup>4</sup>, Shuangjie Lu<sup>12</sup>, Julian K. Lucas<sup>4</sup>, Jean Montong<sup>4</sup>, Haley J. Abel<sup>4</sup>, Silvia Buoniuti<sup>4</sup>, Xian H. Chang<sup>4</sup>, Haoyu Cheng<sup>6,10</sup>, Justin Chu<sup>4</sup>, Vincenza Colonna<sup>4,11</sup>, Jordan M. Eizenga<sup>4</sup>, Xiaowen Feng<sup>10</sup>, Christian Fischer<sup>13</sup>, Robert S. Fulton<sup>12,13</sup>, Shilpa Garg<sup>14</sup>, Cristian Groza<sup>4</sup>, Andrea Ouarracino<sup>6,9</sup>, William T. Harvey<sup>4</sup>, Simon Heumos<sup>15,16</sup>, Kerstin Howe<sup>16</sup>, Miten Jain<sup>17</sup>, Tsung-Yu Lu<sup>18</sup>, Charles Markello<sup>4</sup>, Fergal J. Martin<sup>19</sup>, Matthew W. Mitchell<sup>14</sup>, Katherine M. Munson<sup>17</sup>, Moses Njagi Mwaniki<sup>20</sup>, Adam M. Novak<sup>4</sup>, Hugh E. Olsen<sup>4</sup>, Trevor Pasout<sup>4</sup>, David Porubsky<sup>21</sup>, Pjotr Prins<sup>4</sup>, Jonas A. Sibbesen<sup>22</sup>, Jouni Sirén<sup>4</sup>, Chad Tomlinson<sup>23</sup>, Flavia Villani<sup>24</sup>, Mitchell R. Vollger<sup>25</sup>, Lucinda L. Antonacci-Fulton<sup>26</sup>, Gunjan Baksi<sup>26</sup>, Carl A. Baker<sup>27</sup>, Anastasiya Belyaeva<sup>28</sup>, Konstantinos Billis<sup>29</sup>, Andrew Carroll<sup>30</sup>, Pi-Chuan Chang<sup>31</sup>, Sarah Cody<sup>32</sup>, Daniel E. Cook<sup>33</sup>, Robert M. Cook-Deegan<sup>34</sup>, Omar E. Cornejo<sup>35</sup>, Mark DiKhans<sup>36</sup>, Peter Eberhard<sup>37,38</sup>, Susan Fairley<sup>39</sup>, Olivier Fedrigo<sup>40</sup>, Adam L. Felsenfeld<sup>41</sup>, Giulio Formenti<sup>42</sup>, Adam Frankish<sup>43</sup>, Yan Gao<sup>44</sup>, Nanibaa' A. Garrison<sup>45,46,47</sup>, Carlos Garcia Giron<sup>48</sup>, Richard E. Green<sup>49,50</sup>, Leanne Haggerty<sup>51</sup>, Kendra Hoekzema<sup>52</sup>, Thibaut Hourlier<sup>53</sup>, Hanlee P. Ji<sup>54</sup>, Eimear E. Kenny<sup>55</sup>, Barbara A. Koenig<sup>56</sup>, Alexey Kolesnikov<sup>57</sup>, Jan O. Korbe<sup>58,59</sup>, Jennifer Kordosky<sup>60</sup>, Sergey Koren<sup>61</sup>, HoJoon Lee<sup>62</sup>, Alexandra P. Lewis<sup>63</sup>, Hugo Magalhães<sup>64</sup>, Santiago Marco-Sola<sup>65</sup>, Pierre Marjion<sup>66</sup>, Ann McCartney<sup>67</sup>, Jennifer McDaniel<sup>68</sup>, Jacquelyn Mountcastle<sup>69</sup>, Maria Nattestad<sup>70</sup>, Sergey Nurk<sup>71</sup>, Nathan D. Olson<sup>72</sup>, Alice B. Popejoy<sup>73</sup>, Daniela Pulu<sup>74</sup>, Mikko Rautiainen<sup>75</sup>, Allison A. Regier<sup>76</sup>, Arang Rhie<sup>44</sup>, Samuel Sacco<sup>77</sup>, Ashley D. Sanders<sup>78</sup>, Valerie A. Schneider<sup>79</sup>, Bergerson I. Schultz<sup>80</sup>, Kishwar Shaifin<sup>81</sup>, Michael W. Smith<sup>82</sup>, Heidi J. Sofia<sup>83</sup>, Ahmad N. Abou Tayoun<sup>84,85</sup>, Françoise Thibaud-Nissen<sup>86</sup>, Francesca Floriana Tricomini<sup>87</sup>, Justin Wagner<sup>88</sup>, Brian Walenz<sup>89</sup>, Jonathan M. D. Wood<sup>90</sup>, Aleksey V. Zimin<sup>91,92</sup>, Guillaume Bourque<sup>93,94</sup>, Mark J. P. Chaisson<sup>95</sup>, Paul Flicek<sup>96</sup>, Adam M. Phillippy<sup>97</sup>, Justin M. Zook<sup>98</sup>, Evan E. Eichler<sup>108</sup>, David Haussler<sup>106</sup>, Ting Wang<sup>109</sup>, Erich D. Jarvis<sup>104,109</sup>, Karen H. Miga<sup>110</sup>, Erik Garrison<sup>112</sup>, Tobias Marschall<sup>14,12</sup>, Ira M. Hall<sup>121</sup>, Heng Li<sup>122</sup> & Benedict Paten<sup>123</sup>



# Analyzing a Person's Genome Sequence

## 1. Detecting all genomic variants in the generated genome sequence

Required: Reference Genome Sequence (or Pangenome Reference)

Examples: GRCh38.p13 (Build 38) or T2T-CHM13

## 2. Filtering & Prioritizing detected genomic variants to identify those most likely to be clinically relevant (e.g., pathogenic variants in the case of rare genetic diseases)

Required: Reference Population Databases (Aggregated Genomic Variants)

Example: gnomAD

# **Frequencies of Genomic Variants: Rare Disease Diagnostics as Prototype**

- **Vast majority of pathogenic genomic variants are rare**
- **But being rare does not mean a genomic variant is pathogenic**
- **However, being common means a genomic variant is unlikely to be pathogenic**
- **Therefore, genomic variants are initially FILTERED into groups that are common (removed) and rare (prioritized)**
- **Following filtering, prioritized variants are further assessed for possible pathogenicity**

# Analyzing a Person's Genomic Variants

Number of Variants in  
Person's Genome Sequence

~3-5M

10,000's

1000's

100's

10-20

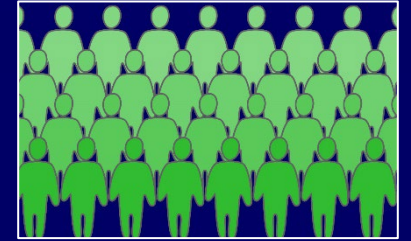
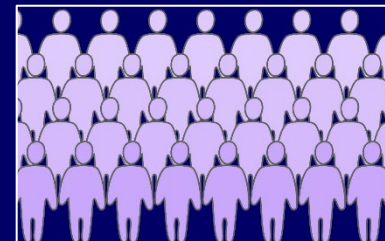
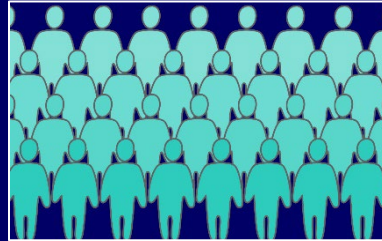
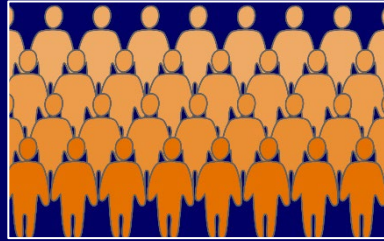
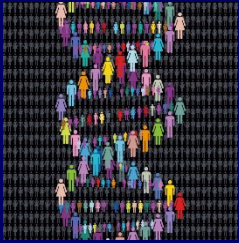
Filtering  
&  
Prioritizing





# Frequencies of Genomic Variants Vary Among Ancestral Populations

## Different Ancestral Populations



**Variant 1**

**30%**

**35%**

**20%**

**18%**

**Variant 2**

**0.001%**

**0%**

**0.001%**

**0.0001%**

**Variant 3**

**0.001%**

**30%**

**25%**

**20%**

**Variant 4**

**0.01%**

**0.0001%**

**0.001%**

**30%**

**Frequency of Each Variant in Each Ancestral Population**

# Overcoming Inequities in Genomic Diagnoses

## Decreasing Incorrect Genomic Variant Classifications

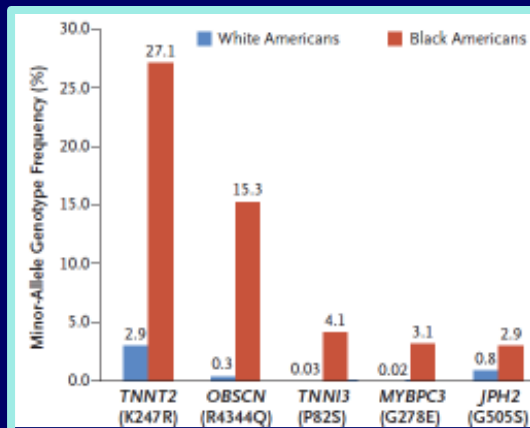
The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL ARTICLE

### Genetic Misdiagnoses and the Potential for Health Disparities

Arjun K. Manrai, Ph.D., Birgit H. Funke, Ph.D., Heidi L. Rehm, Ph.D., Morten S. Olesen, Ph.D., Bradley A. Maron, M.D., Peter Szolovits, Ph.D., David M. Margulies, M.D., Joseph Loscalzo, M.D., Ph.D., and Isaac S. Kohane, M.D., Ph.D.

*Manrai et al., N Engl J Med (2016)*



Misclassification of 5 genomic variants for cardiomyopathy

## Increasing Diagnoses via Equitable Study Recruitment and Clinical Testing

Petrovski and Goldstein *Genome Biology* (2016) 17:157  
DOI 10.1186/s13059-016-1016-y

Genome Biology

OPEN LETTER Open Access

### Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine

Slavé Petrovski<sup>1,2\*</sup> and David B. Goldstein<sup>1\*</sup>

*Petrovski & Goldstein Genome Biol (2016)*

communications  
biology

ARTICLE Check for updates

<https://doi.org/10.1038/s42003-023-05708-y> OPEN

### The frequency of pathogenic variation in the All of Us cohort reveals ancestry-driven disparities

Eric Venner<sup>1,2</sup>, Karynne Patterson<sup>2</sup>, Divya Kalra<sup>1</sup>, Marsha M. Wheeler<sup>2</sup>, Yi-Ju Chen<sup>1</sup>, Sara E. Kalla<sup>1</sup>, Bo Yuan<sup>1</sup>, Jason H. Karnes<sup>3,4</sup>, Kimberly Walker<sup>1</sup>, Joshua D. Smith<sup>2</sup>, Sean McGee<sup>2</sup>, Aparna Radhakrishnan<sup>2</sup>, Andrew Haddad<sup>5</sup>, Philip E. Empey<sup>6</sup>, Qiaoyan Wang<sup>1</sup>, Lee Lichtenstein<sup>7</sup>, Diana Toledo<sup>7</sup>, Gail Jarvik<sup>8,9</sup>, Anjane Musick<sup>10</sup> & Richard A. Gibbs<sup>1</sup> on behalf of the All of Us Research Program Investigators\*

*Venner et al., Commun Biol (2024)*

# Analyzing a Person's Genome Sequence

## 1. Detecting all genomic variants in the generated genome sequence

Required: Reference Genome Sequence (or Pangenome Reference)

Examples: GRCh38.p13 (Build 38) or T2T-CHM13

## 2. Filtering & Prioritizing detected genomic variants to identify those most likely to be clinically relevant (e.g., pathogenic variants in the case of rare genetic diseases)

Required: Reference Population Databases (Aggregated Genomic Variants)

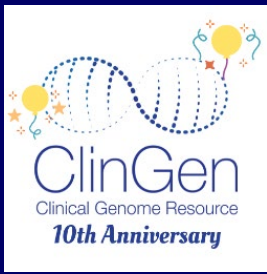
Example: gnomAD

## 3. Establishing the clinical relevance of prioritized genomic variants

Required: Knowledgebase with Information about Pathogenicity of Genomic Variants

Example: ClinGen





# Clinical Genome Resource (ClinGen)

**Mission: Build and support an authoritative central resource that defines the clinical relevance of genes and variants for use in precision medicine and research**

**Global Network of Contributors:  
>2,700 Experts from 69 Countries**



**2,682 Gene-Disease Validity Curations**



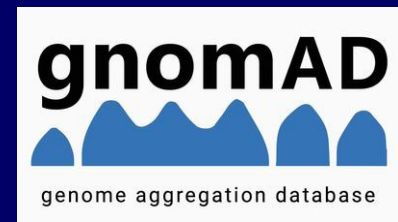
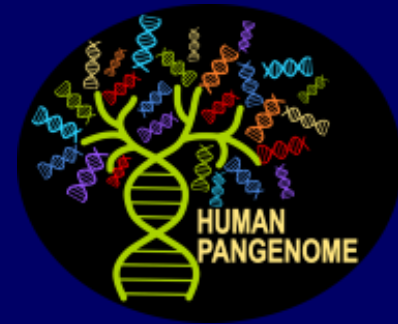
**7,130 Variant Pathogenicity Curations**



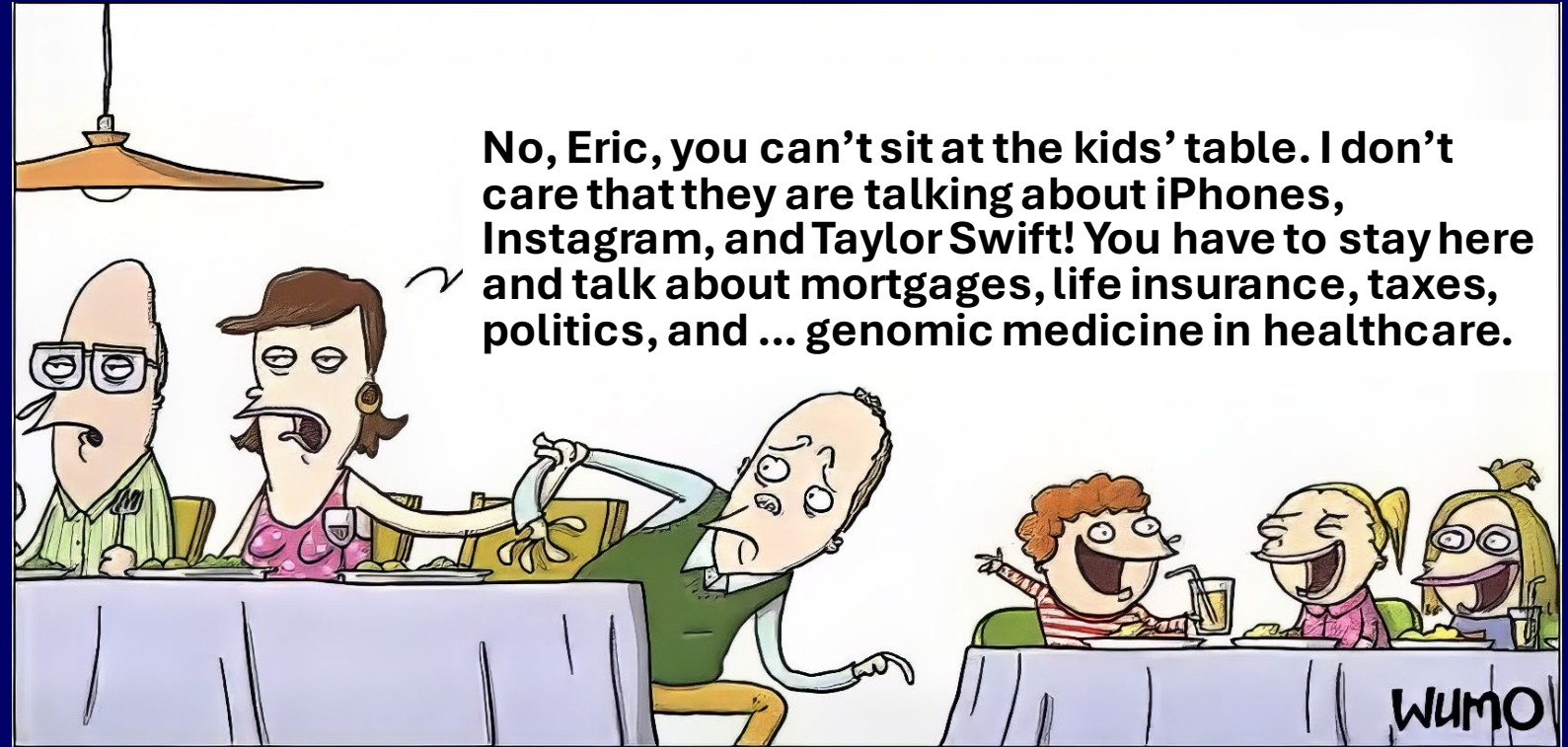
**248 Clinical Actionability Reports**

# Required for Accurate and Equitable Analyses of a Person's Genome Sequences

1. **Appropriately matched human genome reference sequence – or a human pangenome reference**
2. **Reference population database (with aggregated genomic variant information) for appropriately matched ancestral population(s)**
3. **Robust knowledgebase of curated information about the likely pathogenicity of genomic variants (developed by expert panels)**



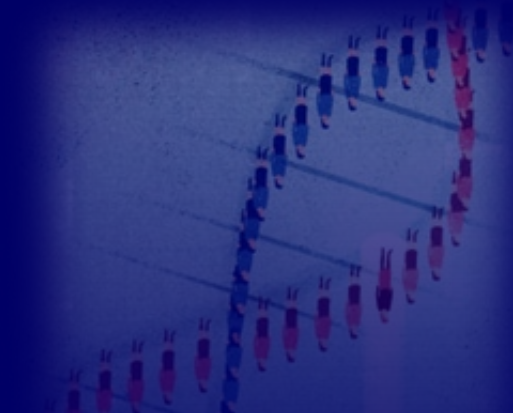
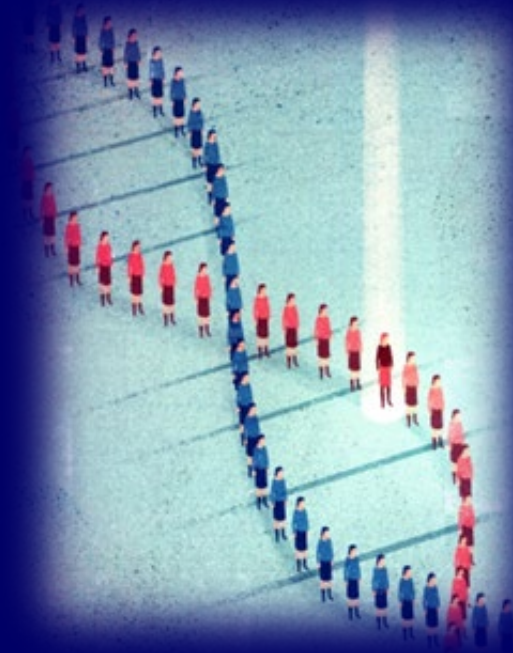
# Genomics Arrives at the 'Adult Table'





# Genomics and Society

G G G T T A T G G C G A G  
G T T C G G G A A T C C  
C T T A C G G A A A T C T  
C T T T G A C G A C T C C  
C T T A G A G G A C T C C  
G A G C T G G T C T A G A  
G A A C T G G T A T A G G  
C T A G G C G T T A C A  
C T T G G C G T G A C A  
A G C T T G G C C G A A C  
G G C T A G C C G A A C  
C A G T A C A T G A A C G  
C G C T A C A T G T A C C  
C V G I V C V I G V V C G  
G C C I I V C C C G Y V C  
Y C C I I G C C C G Y V C  
C I I G C C G I G V C V  
C I V C C C C I I V C V  
G V V C I G G I V I V G  
G V G C I G C I C I V C Y



# Societal Challenges with Genomic Medicine

## *Genomic Literacy*





The **Forefront**  
of **Genomics**<sup>®</sup>





# 2020 NHGRI Strategic Vision



## Perspective

# Strategic vision for improving human health at The Forefront of Genomics

<https://doi.org/10.1038/s41586-020-2817-4>

Received: 30 June 2020

Accepted: 4 September 2020

Published online: 28 October 2020

[Check for updates](#)

Eric D. Green<sup>1</sup>, Chris Gunter<sup>2</sup>, Leslie C. Biesecker<sup>3</sup>, Valentina Di Francesco<sup>4</sup>, Carla L. Easton<sup>5</sup>, Elise A. Feingold<sup>6</sup>, Adam L. Felsenfeld<sup>7</sup>, David J. Kaufman<sup>8</sup>, Elaine A. Ostrander<sup>9</sup>, William J. Pavan<sup>10</sup>, Adam M. Philippou<sup>11</sup>, Anastasia L. Wiley<sup>12</sup>, Jyoti Gupta Dayal<sup>13</sup>, Britny J. Kish<sup>14</sup>, Allison Mandich<sup>15</sup>, Christopher R. Wellington<sup>16</sup>, Kris A. Wetterstrand<sup>17</sup>, Sarah A. Batey<sup>18</sup>, Darryl Leja<sup>19</sup>, Susan Vazquez<sup>20</sup>, William A. Gahl<sup>21</sup>, Bettie J. Graham<sup>22</sup>, Daniel L. Kastner<sup>23</sup>, Paul Liu<sup>24</sup>, Laura Lyman Rodriguez<sup>25</sup>, Benjamin D. Solomon<sup>26</sup>, Vance L. Bonham<sup>27</sup>, Lawrence C. Brody<sup>28</sup>, Carolyn M. Hutter<sup>29</sup> & Tari A. Manolio<sup>30</sup>

Starting with the launch of the Human Genome Project three decades ago, and continuing after its completion in 2003, genomics has progressively come to have a central and catalytic role in basic and translational research. In addition, studies increasingly demonstrate how genomic information can be effectively used in clinical care. In the future, the anticipated advances in technology development, biological insights, and clinical applications (among others) will lead to more widespread integration of genomics into almost all areas of biomedical research, the adoption of genomics into mainstream medical and public health practices, and an increasing relevance of genomics for everyday life. On behalf of the research community, the National Human Genome Research Institute recently completed a multi-year process of strategic engagement to identify future research priorities and opportunities in human genomics, with an emphasis on health applications. Here we describe the highest-priority elements envisioned for the cutting-edge of human genomics going forward—that is, at 'The Forefront of Genomics'.

Beginning in October 1990, a pioneering group of international researchers began an audacious journey to generate the first map and sequence of the human genome, marking the start of a 13-year odyssey called the Human Genome Project<sup>1</sup>. The successful and early completion of the Project in 2003, which included parallel studies of a set of model organism genomes, catalysed enormous progress in genomics research. Leading the signature advances has been a greater than one million-fold reduction in the cost of DNA sequencing<sup>2</sup>. This decrease has allowed the generation of innumerable genome sequences, including hundreds of thousands of human genome sequences (both in research and clinical settings), and the continuous development of assays to identify and characterize functional genomic elements<sup>3,4</sup>. These new tools, together with increasingly sophisticated statistical and computational methods, have enabled researchers to create rich catalogues of human genomic variants<sup>5,6</sup>, to gain an ever-deepening understanding of the functional complexities of the human genome<sup>7</sup>, and to determine the genomic bases of thousands of human diseases<sup>8,9</sup>. In turn, the past decade has brought the initial realization of genomic medicine<sup>10</sup>, as research successes have been converted into powerful tools for use in healthcare. Including somatic genome analysis for cancer (enabling development of targeted therapeutic agents)<sup>11</sup>, non-invasive prenatal genetic screening<sup>12</sup>, and genomics-based tests for a growing set of paediatric conditions and rare disorders<sup>13</sup>, among others.

In essence, with growing insights about the structure and function of the human genome and ever-improving laboratory and computational technologies, genomics has become increasingly woven into the fabric

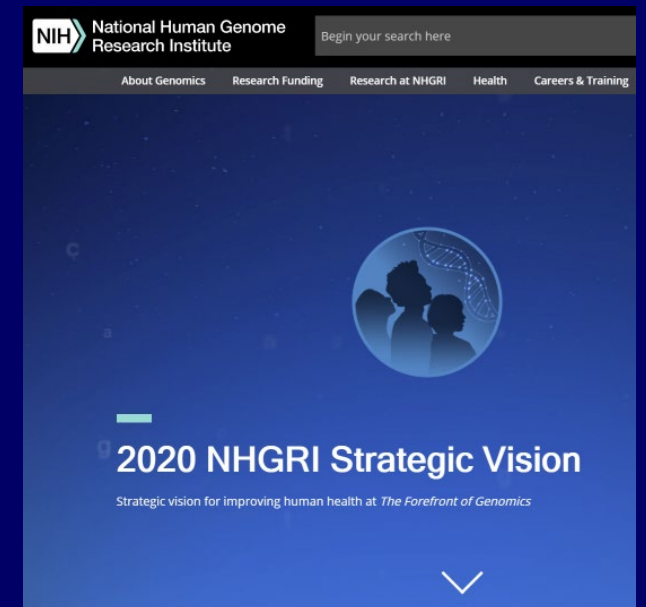
of biomedical research, medical practice, and society. The scope, scale, and pace of genomic advances so far were nearly unimaginable when the Human Genome Project began; even today, such advances are yielding scientific and clinical opportunities beyond our initial expectations, with many more anticipated in the next decade.

Embracing its leadership role in genomics, the National Human Genome Research Institute (NHGRI) has developed strategic visions for the field at key inflection points, in particular at the end of the Human Genome Project in 2003<sup>14</sup> and then again at the beginning of the last decade in 2011<sup>15</sup>. These visions outlined the most compelling opportunities for human genomics research, in each case informed by a multi-year engagement process. NHGRI endeavoured to start the new decade with an updated strategic vision for human genomics research. Through a planning process that involved more than 50 events (such as dedicated workshops, conference sessions, and webinars) over the past two years (see <http://genome.gov/genomics2020>), the Institute collected input from a large number of stakeholders, with the resulting input catalogued and synthesized using the framework depicted in Fig. 1.

Unlike the past, this round of strategic planning was greatly influenced by the now widely disseminated nature of genomics across biomedicine. A representative glimpse into this historic phenomenon is illustrated in Fig. 2. During the Human Genome Project, NHGRI was the primary funder of human genomics research at the US National Institutes of Health (NIH), but the past two decades have brought a greater than tenfold increase in the relative fraction of funding coming from other parts of the NIH.

National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. ✉: [egreen@nhgri.nih.gov](mailto:egreen@nhgri.nih.gov)

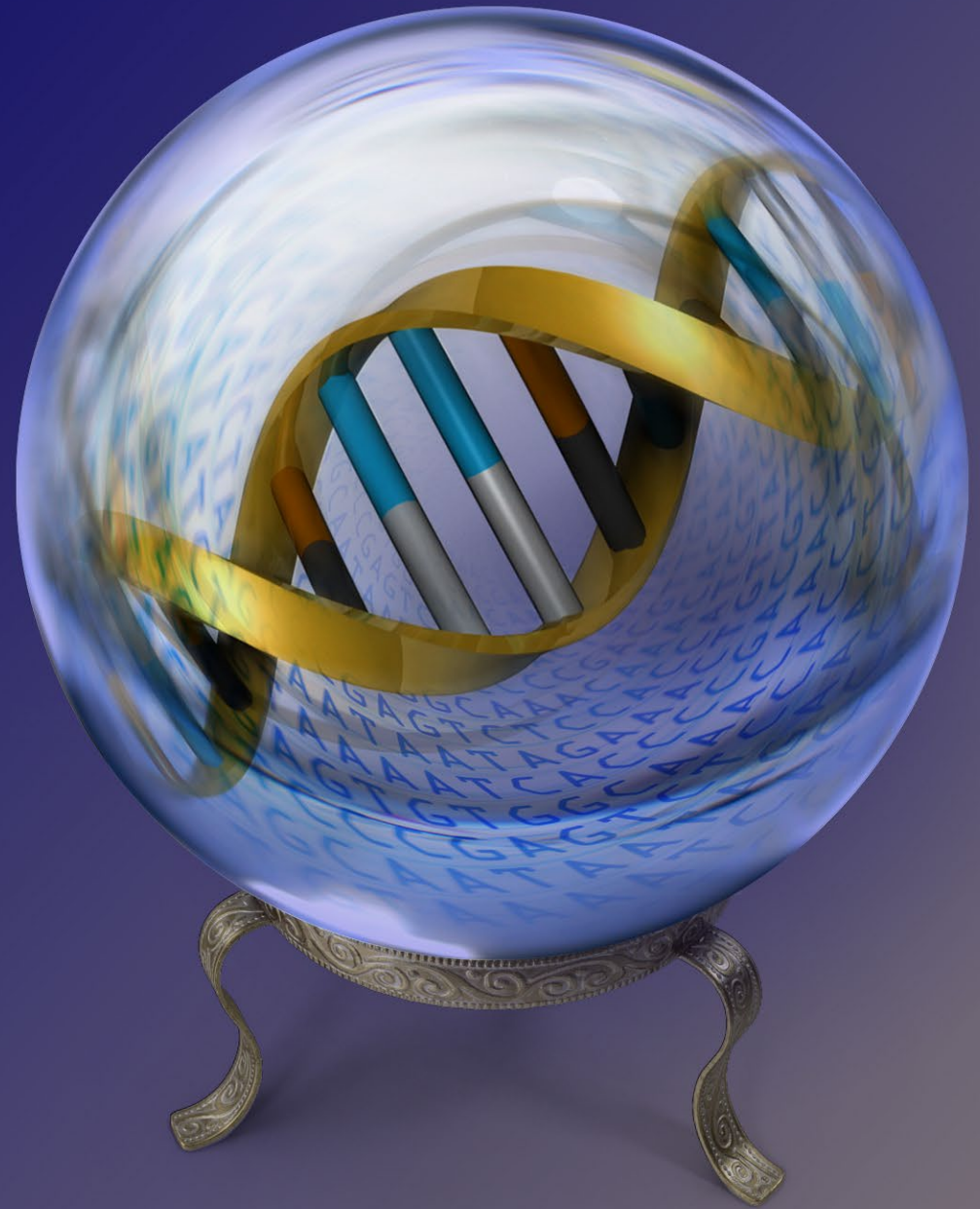
Nature | Vol 586 | 29 October 2020 | 683



Nature (2020)

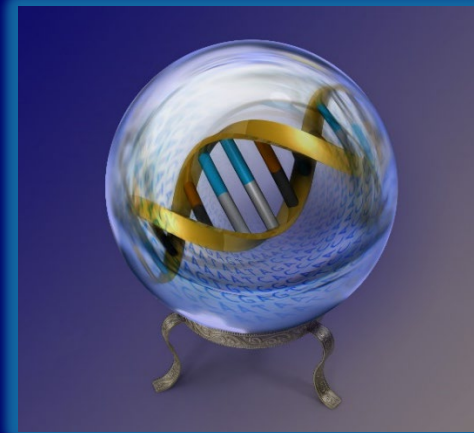
—

# Bold Predictions for Human Genomics by 2030





# Bold Predictions for Human Genomics by 2030



## Perspective

### Box 5

#### Bold predictions for human genomics by 2030

Some of the most impressive genomics achievements, when viewed in retrospect, could hardly have been imagined ten years earlier. Here are ten bold predictions for human genomics that might come true by 2030. Although most are unlikely to be fully attained, achieving one or more of these would require individuals to strive for something that currently seems out of reach. These predictions were crafted to be both inspirational and aspirational in nature, provoking discussions about what might be possible at the Forefront of Genomics in the coming decade.

1. Generating and analysing a complete human genome sequence will be routine for any research laboratory, becoming as straightforward as carrying out a DNA purification.
2. The biological function(s) of every human gene will be known; for non-coding elements in the human genome, such knowledge will be the rule rather than the exception.
3. The general features of the epigenetic landscape and transcriptional output will be routinely incorporated into predictive models of the effect of genotype on phenotype.
4. Research in human genomics will have moved beyond population descriptors based on historic social constructs such as race.
5. Studies that involve genome sequences and associated phenotypic information for millions of human participants will be regularly featured at school science fairs.
6. The regular use of genomic information will have transitioned from boutique to mainstream in all clinical settings, making genomic testing as routine as complete blood counts.
7. The clinical relevance of all encountered genomic variants will be readily predictable, rendering the diagnostic designation 'variant of uncertain significance (VUS)' obsolete.
8. An individual's complete genome sequence along with informative annotations will, if desired, be securely and readily accessible on their smartphone.
9. Individuals from ancestrally diverse backgrounds will benefit equitably from advances in human genomics.
10. Breakthrough discoveries will lead to curative therapies involving genomic modifications for dozens of genetic diseases.

the use of genomics in medicine from diagnosing and treating disease to maintaining health.

Sharp barriers between research and clinical care obstruct the virtuous cycle of moving scientific discoveries rapidly into clinical care and bringing clinical observations back to the research setting<sup>10</sup> (Fig. 3). Learning healthcare systems—in which real-time data on outcomes of healthcare delivery are accessed and used to enhance clinical practice—can lead to continuous care improvement, but only if the barriers between research and clinical care are reduced<sup>11</sup>. For example, offering genome sequencing to all members of a healthcare system, performed in conjunction with research and patient engagement and provided in real time<sup>12</sup>, could help to assess the clinical utility of genomic information and may allow providers to improve diagnosis and management. System-wide implementation of such an experiment requires not only extensive patient and provider education, sophisticated informatics capabilities, and genomics-based clinical decision support, but also the development and evaluation of data security and privacy protections to ensure patient confidentiality<sup>13</sup>. Patients should be engaged in the design of such systems and informed at entry to them (and periodically thereafter), so as to be fully aware of the nature of the ongoing research

### Box 5

#### Bold predictions for human genomics by 2030

Some of the most impressive genomics achievements, when viewed in retrospect, could hardly have been imagined ten years earlier. Here are ten bold predictions for human genomics that might come true by 2030. Although most are unlikely to be fully attained, achieving one or more of these would require individuals to strive for something that currently seems out of reach. These predictions were crafted to be both inspirational and aspirational in nature, provoking discussions about what might be possible at the Forefront of Genomics in the coming decade.

1. Generating and analysing a complete human genome sequence will be routine for any research laboratory, becoming as straightforward as carrying out a DNA purification.
2. The biological function(s) of every human gene will be known; for non-coding elements in the human genome, such knowledge will be the rule rather than the exception.
3. The general features of the epigenetic landscape and transcriptional output will be routinely incorporated into predictive models of the effect of genotype on phenotype.
4. Research in human genomics will have moved beyond population descriptors based on historic social constructs such as race.
5. Studies that involve analyses of genome sequences and associated phenotypic information for millions of human participants will be regularly featured at school science fairs.
6. The regular use of genomic information will have transitioned from boutique to mainstream in all clinical settings, making genomic testing as routine as complete blood counts.
7. The clinical relevance of all encountered genomic variants will be readily predictable, rendering the diagnostic designation 'variant of uncertain significance (VUS)' obsolete.
8. An individual's complete genome sequence along with informative annotations will, if desired, be securely and readily accessible on their smartphone.
9. Individuals from ancestrally diverse backgrounds will benefit equitably from advances in human genomics.
10. Breakthrough discoveries will lead to curative therapies involving genomic modifications for dozens of genetic diseases.

with their clinical data and the goals and potential risks of their participation<sup>14</sup>. Extending such studies across many healthcare systems should reveal common challenges and solutions<sup>15,16</sup>, thereby enhancing the learning healthcare model for genomic medicine more broadly (Fig. 3).

#### Concluding thoughts

The dawn of genomics featured the launch of the Human Genome Project in October 1990<sup>17</sup>. Three decades later, the field has seen stunning technological advances and high-profile programmatic successes, which in turn have led to the widespread infusion of genomic methods and approaches across the life sciences and, increasingly, into medicine and society.

NHGRI has for the third time<sup>18</sup> since the Human Genome Project undergone an extensive horizon-scanning process to capture, synthesize, and articulate the most compelling strategic opportunities for the next phase of genomics—with particular attention to elements that are most relevant to human health. The now near-ubiquitous nature of genomics (including in the complex healthcare ecosystem) presented practical challenges for attaining a holistic assessment of the field. Another reality was that the NHGRI investment in genomics has now been multiplied many-fold by the seeding of human genomics throughout the broader research community. These changes reflect a continued maturation of both the field (in general) and NHGRI (more specifically), nicely aligning with the institute's evolving leader-

Embracing that role, NHGRI formulated the strategic vision described here, which provides an optimistic outlook that the successes in human genomics over the past three decades will be amplified in the coming decade. Many of the details about what is needed to fulfill the promise of genomics have now come into focus. Major unsolved problems remain—among them determining the role for the vast majority of functional elements in the human genome (especially those outside of protein-coding regions), understanding the full spectrum of genomic variation (especially that implicated in human disease), developing data-science capabilities (especially those that keep pace with data generation), and improving healthcare through the implementation of genomic medicine (especially in the areas of prevention, diagnosis, and therapeutic development). The new decade also brings research questions related to the societal implications of genomics, including those related to social inequities, pointing to the continued importance of investigating the ethical, legal, and social issues related to genomics. But now more than ever, solutions to these problems seem to be within striking distance. Towards that end (and with the characteristic spirit of genomics audacity), we offer ten bold predictions of what might be realized in human genomics by 2030 (Box 5).

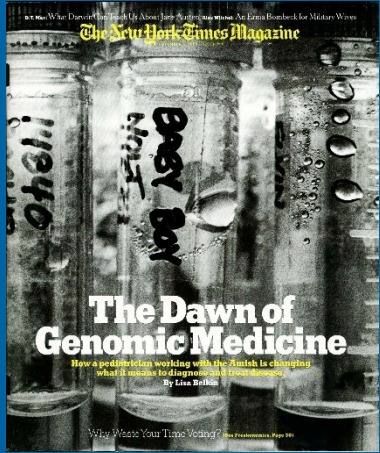
of human genomics and emphasizes broad strategic goals as opposed to implementation tactics. The realization of these goals will require further training in conjunction with the collective creativity, energies, and resources of the global community of scientists, funders, and research participants. NHGRI has taken some initial steps to implement this vision, although these will inevitably need to be adapted as advances occur and circumstances change. Indeed, the final words of this strategic vision were formulated as the world moved urgently to deal with the coronavirus disease 2019 (COVID-19) pandemic (see below), providing a vivid reminder of the need for nimbleness and the importance of nurturing all parts of the research continuum—from basic to translational to clinical—for protecting public health and advancing medical science. Despite the seismic changes seen in genomics since the inception of the field, the fundamental sense of curiosity, marvel, and purpose associated with genome science seems to be timeless. In concluding NHGRI's previous strategic vision<sup>18</sup>—published just under a decade ago—the then-envisioned opportunities and challenges were provided with "... a continuing sense of wonder, a continuing need for urgency, a continuing desire to balance ambition with reality, and a continuing responsibility

Embracing that role, NHGRI formulated the strategic vision described here, which provides an optimistic outlook that the successes in human genomics over the past three decades will be amplified in the coming decade. Many of the details about what is needed to fulfill the promise of genomics have now come into focus. Major unsolved problems remain—among them determining the role for the vast majority of functional elements in the human genome (especially those outside of protein-coding regions), understanding the full spectrum of genomic variation (especially that implicated in human disease), developing data-science capabilities (especially those that keep pace with data generation), and improving healthcare through the implementation of genomic medicine (especially in the areas of prevention, diagnosis, and therapeutic development). The new decade also brings research questions related to the societal implications of genomics, including those related to social inequities, pointing to the continued importance of investigating the ethical, legal, and social issues related to genomics. **But now more than ever, solutions to these problems seem to be within striking distance. Towards that end (and with the characteristic spirit of genomics audacity), we offer ten bold predictions of what might be realized in human genomics by 2030 (Box 5).**



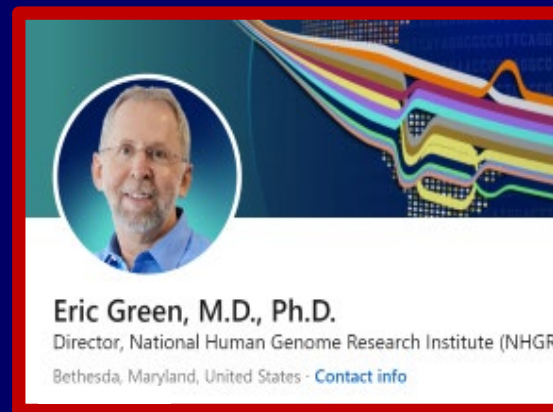


# A ~21-year Pivot: Bringing Genomic Medicine Into Focus



# Let's Stay Connected!

[genome.gov/stayconnected](https://genome.gov/stayconnected)



Subscribe:



[genome.gov/email](https://genome.gov/email)

Follow:



[@NHGRI\\_Director](https://twitter.com/NHGRI_Director)

Follow:



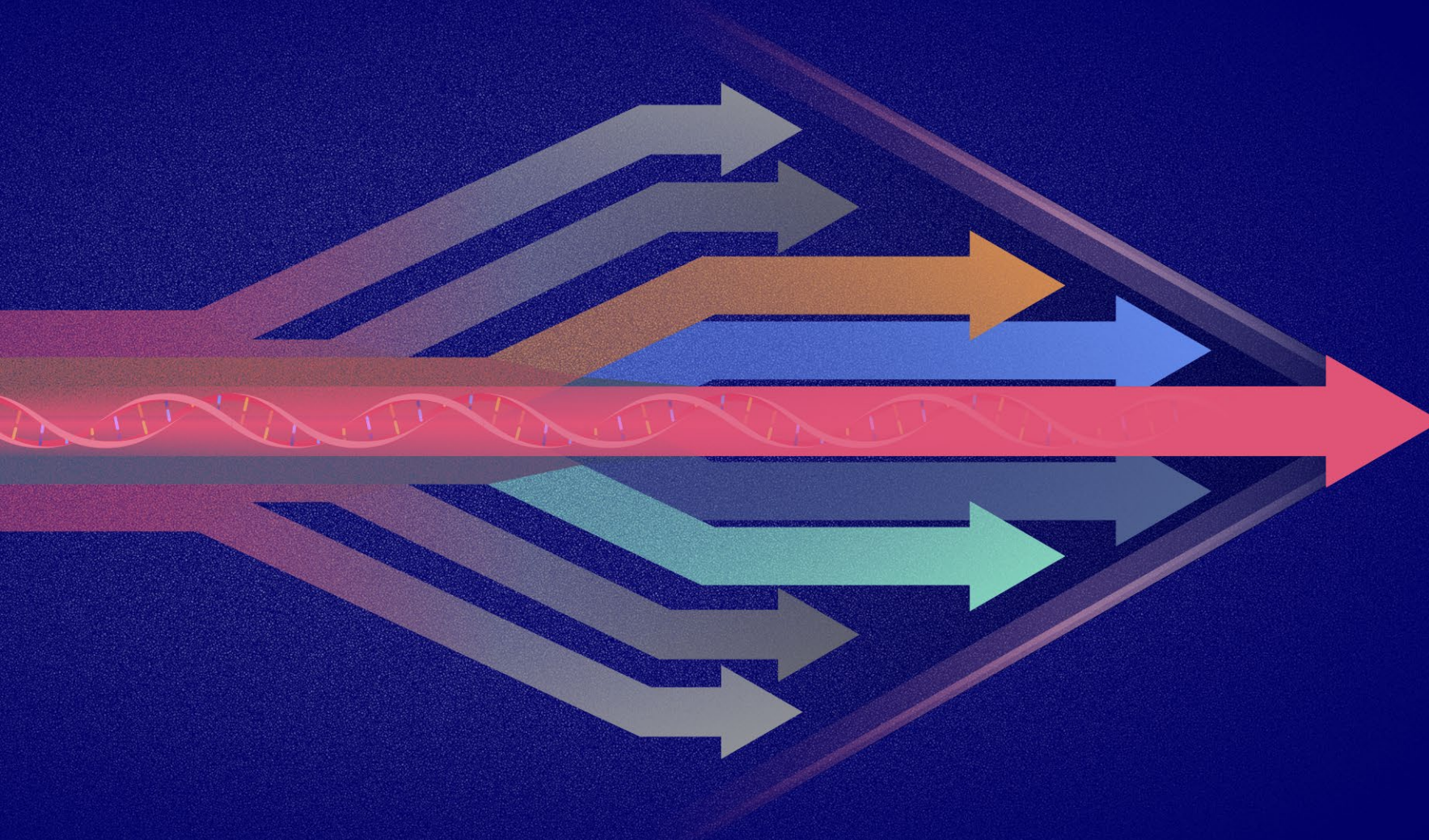
[linkedin.com/in/  
eric-green-md-phd](https://linkedin.com/in/eric-green-md-phd)

Follow:



[@NHGRI\\_Director](https://instagram.com/NHGRI_Director)





The **Forefront**  
of **Genomics**<sup>®</sup>



# How to Represent a Human Pangenome Reference



# Human Pangenome Reference Graph

## Visualizing a Human Pangenome Reference

