

# THE HUMAN STORY BEHIND HUMAN GENOME SEQUENCING

In 2008, **Professor James Lupski**, from **Baylor College of Medicine**, USA, contributed to the scientific work behind the first human whole genome sequencing. Two years later, he was one of the first people to have his own genome sequenced. He is considered a pioneer of clinical genomics, a field that has revolutionised the diagnosis, understanding, clinical management and treatment of genetic conditions and diseases, including James's own.



**PROFESSOR  
JAMES LUPSKI**

*Baylor College of Medicine & Texas Children's Hospital,  
Houston, Texas, USA*

## Field of research

Clinical Genomics

## Research highlights

James is a pioneer in the field of clinical genomics – he contributed to the first personal genome sequencing efforts and was then one of the first people to have their genome sequenced

## Funders

US National Institutes of Health (NIH): National Institute of Neurological Disorders and Stroke (R35 NS105078), National Institute of Human Genome Research (U01 HD011758)

**H**idden in the nuclei of your cells, DNA molecules lie tightly coiled, harbouring the genetic secrets that make you who you are. Thanks to advancements in clinical genomics, it is now possible to unravel your DNA's code, allowing scientists to better understand and treat genetic conditions.

## What is a genome?

Every living thing has a genome, and every genome is unique. "A genome is the complete set of DNA in each cell of an organism," explains Professor James Lupski, a clinical genomicist at Baylor College of Medicine. This DNA is composed of two complementary strands that twist around each other, forming a double-helix structure, linked like steps in a ladder by pairs of molecules called bases. The size of a genome can be measured by the total number of base pairs (bp) present in an organism's DNA.

## TALK LIKE A ...

## CLINICAL GENOMICIST

**Autosomal recessive inheritance** — when an individual receives two mutated versions of a gene, one from each biological parent

**Chromosome** — the structure containing DNA in a cell's nucleus

**Chronic disease** — a long-term health condition

**Diploid** — the presence of two sets of chromosomes, one supplied by each biological parent

**Haploid** — the presence of only one set of chromosomes in sex cells

**Recombination** — the process by which pieces of DNA are broken and recombined, creating new combinations of genetic information

The size of a genome varies dramatically between species and is not proportional to the size of the organism. For example, horses have 2.7 billion bp in their genome, whereas tiny desert locusts have around 8.8 billion bp. One of the largest recorded genomes belongs to *Polychaos dubium*, a single-celled amoeba, which boasts an impressive 670 billion bp. In comparison, the human genome seems somewhat meagre, containing a modest 3.2 billion bp. But this genome is what sets us apart – it is what makes us human.

Our 3.2 billion bp are split into 23 chromosomes, and two complete chromosomal sets are housed within the nucleus of almost every cell in our body. The exceptions are egg and sperm cells. These sex cells are haploid, meaning they each contain only one set of chromosomes. "During sexual reproduction, both biological parents contribute one set of chromosomes to their child, providing them with a full diploid

genome of 46 chromosomes and roughly 6 billion bp of DNA," explains James.

## What causes changes to the genome?

While a child inherits genetic information from both biological parents, the process of recombination during sexual reproduction causes this DNA to get mixed up. This is the main cause of our individuality and results in children having traits (e.g., hair and eye colour) that are a mixture of those of their biological parents.

Alterations to DNA are also caused by mutations. There are three main types of genetic mutation: single nucleotide variation (SNV), in which just one base pair changes; indels, in which small sections of DNA (up to 50 bp) are inserted or deleted; and copy number variants (CNV), which occur when large sections of DNA are lost or gained in the maternally or paternally inherited genome. "CNVs can involve millions of base pairs," says James. When this happens, it causes genomic



© RFBSIP/stock.adobe.com

rearrangements, mutations that occur when large sections of DNA in the genome are repeated (as they are in CNVs), rearranged or deleted.

### What are the impacts of genetic mutations?

Genetic mutations are the main driver of evolution, allowing individuals to develop new traits that, if useful, are passed on through the generations. However, not all mutations are beneficial. A mutation can change a nucleotide at a given position of the human genome and involve just a single gene, or it can affect large sections of DNA involving multiple genes, such as in genomic rearrangements. Such variations can cause genetic diseases that can severely impact people's lives.

Thanks to the process of genome sequencing, scientists can now study the genomes of individual people in detail. This technique gave rise to the field of clinical genomics, where scientists and medical professionals investigate the genetic causes of diseases, to understand their biological mechanisms at the molecular level.

### What is whole genome sequencing?

Whole genome sequencing (WGS) is the technique to read an individual's entire genome. Scientists start by taking a cell sample, such as blood or skin cells, from the person. DNA is extracted from these cells and cut into short fragments roughly 100 bp long. These fragments are put into a sequencer which determines the order of their base pairings. This process produces millions of sequenced fragments that are then stitched together by a powerful computer program and compared to the haploid reference human genome (HRHG).

The HRHG is a composite genome created by scientists from the genomes of multiple individuals. It is used as the standard reference against which researchers compare genomes of people being sequenced to study their unique individual variation. This comparison allows geneticists to determine which genetic mutations are present in the person and helps them

understand how mutations might affect gene function and contribute to disease. WGS allows individuals to understand how their genetics might impact their health.

### What are the challenges of personal genome sequencing?

"Ethical considerations are always at the forefront of research involving human subjects," says James. "Potential benefits must outweigh the risks, and the research participants must provide informed consent before their genome is sampled and sequenced."

From a technical perspective, sequencing all 6 billion bp in a human genome is an incredibly complex task. Moreover, analysing the data produced and comparing them to the HRHG requires monumental amounts of computing power. The databases produced by some genomic research projects contain thousands of trillions of base pairs. Analysing these 'big data' sets and drawing conclusions about the potential health consequences of an individual's genome is a daunting task. Without recent advancements in computational science and big data analyses, personal genome sequencing would not be possible.

Financially, both the experimental and computational sequencing methods are expensive. At first, this was a significant barrier to genome sequencing. However, as the technique has progressed, the cost of sequencing an individual genome has decreased, and next-generation sequencing is now possible. Additionally, the value of information produced by genome sequencing increases as we learn more about human biology and disease.

### What did the first personal genome sequence show?

In 2008, James contributed to sequencing and analysing the very first personal genome using next-generation technologies, that of Dr James Watson, co-discoverer of the Watson-Crick double-helix structure of DNA. This WGS was a major milestone in the emerging field of clinical genomics, as it established that personal genome sequencing had become a technically and financially viable process.

“

**ETHICAL CONSIDERATIONS ARE ALWAYS AT THE FOREFRONT OF RESEARCH INVOLVING HUMAN SUBJECTS.**

”

Sequencing Watson's genome and comparing it to the HRHG provided important insights into genomics. One significant discovery was the massive amount of variation between Watson's genome and the HRHG. This shows that, despite humans sharing a large portion of our DNA, we are all truly unique as individuals. This initial sequencing also indicated the HRHG is a robust reference tool for mapping variations in individual genomes, proving it can be relied upon for these kinds of studies.

These original findings, along with the hundreds of thousands of personal genomes sequenced since 2008, have paved the way for personal genomics to be used in clinical settings. By sequencing an individual's genome, medical professionals can diagnose genetic disorders. WGS can be important for medical management in families and can help physicians prescribe the most appropriate treatments.

### Why is clinical genomics important?

"Clinical genomics can tell individuals whether any rare diseases run in their family," explains James. It can also highlight whether a person might be more prone to common conditions such as heart disease or cancer. Understanding the gene variants that cause genetic diseases allows medical professionals to provide personalised treatments to their patients. This means that, in the future, people with genetic conditions will hopefully be able to live healthier lives, thanks to the work of clinical genomicists such as James.



# JAMES'S GENOMIC STORY

**J**ames played a key part in the first personal whole genome sequencing. As a clinical genomicist, his role was to interpret the variations in Watson's genome and decipher the potential medical implications. After this scientific milestone, researchers wanted to repeat the process for someone with a genetic disease, so James volunteered.

"When I saw what could be done with the Watson personal genome, the opportunity to provide insight into my own disease and a 'real life' clinical situation was readily apparent," says James, who has Charcot-Marie-Tooth disease (CMT), a disorder involving the function of the peripheral nerves. "I volunteered out of pure scientific curiosity, and because I hoped we might discover something to help families suffering with CMT."

## What is CMT?

CMT is a chronic disease that causes peripheral neuropathy, the degradation of nerve cells in the body's extremities, mainly in the hands and feet. CMT can cause foot deformities that can make

walking very difficult, as well as muscle wasting, weakness, numbness and tingling in the hands and feet. "From a young age, I had difficulty in finding a pair of shoes that would fit and had trouble walking," says James.

## What did James's genome sequencing uncover?

Before sequencing his genome, James had studied CMT in his lab at Baylor College of Medicine for over twenty years. During this time, he discovered the copy number variant (CNV), known as the CMT1A duplication, responsible for the major form of the disease, as well as several other gene variations that can cause CMT. "But I had never been able to determine the cause of my disease," he says, as none of these variations could have been the cause of CMT in his family (three of his seven siblings also have CMT).

"In my family, CMT appeared to segregate as an autosomal recessive trait," explains James. This means two copies of a mutated gene must be present in an individual for the disease to develop.

The causes of CMT that James and others had already discovered did not match with the characteristics of autosomal recessive inheritance.

So, James's genome was sequenced, then analysed by Claudia Gonzaga-Jauregui, one of the graduate students in his lab. She discovered he had a variation in a gene that is important for maintaining peripheral nerve cells, which was consistent with the expected characteristics of recessive inheritance. "I cannot tell you how exciting that moment was!" exclaims James. For the first time, he understood the cause of his family's disease, and his team had shown that personal genome sequencing can pinpoint the genetic causes of a disease.

This breakthrough was reported around the world and ushered in a new era in genomics. As well as being a huge personal triumph, James's team's discovery helped pave the way for personalised medicine to become a reality, potentially helping millions of people and families around the globe living with genetic diseases.

## Pathway from school to clinical genomics

- At school and college, biology, chemistry, maths and computer science will give you the foundations of the skills and knowledge needed by clinical genomicists.
- Some universities offer degrees in genomics. Degrees in genetics, medicine, molecular biology and biochemistry could also lead to a career in clinical genomics.
- Volunteering in research labs and finding internships is a great way to get practical, hands-on experience.

## Explore careers in clinical genomics

- Discover genetics with the American Society of Human Genetics ([www.ashg.org/discover-genetics](http://www.ashg.org/discover-genetics)) and find out what careers are available in the field ([www.ashg.org/careers-learning/career-flowchart](http://www.ashg.org/careers-learning/career-flowchart))
- The American College of Medical Genetics and Genomics provides information about careers in medical genetics ([www.acmg.net/ACMG/Education/Student/Careers\\_in\\_Medical\\_Genetics.aspx](http://www.acmg.net/ACMG/Education/Student/Careers_in_Medical_Genetics.aspx))



# Q&A

## Meet James

### What were your interests when you were younger?

As a child, I always liked to take things apart, then put them back together, particularly the neighbour's lawnmower. As a teenager, I enjoyed building and riding small motorcycles, and making fireworks. I taught myself how to make gunpowder and became terribly excited with the explosive results! I was also fond of playing chess, watching football and listening to rock-and-roll music.

### How did CMT impact your teenage years?

I had ten operations during high school and the first few years of college. After each, I spent considerable time in a wheelchair or on crutches while I recovered, so I spent most of high school being home-schooled. I only had tutoring for a couple of hours a day, four days a week, leaving me with plenty of free time to explore my curiosity.

### To what extent did your personal experiences influence your interest in genetics and medicine?

A lot! Living with CMT made me fascinated by genetics and medicine. I vividly recall that the home tutor who taught me biology exposed me to DNA for the first time. I had been trying to understand genetics by thinking about the transmission of CMT with respect to my siblings and me. Now, I could explore it by studying how a real molecule influences the chemistry and biology of a living organism (in this case, me).

**“HUMAN GENETICS IS THE STUDY OF HUMANS, BY HUMANS. IT HAS BEEN AN INCREDIBLE PRIVILEGE TO BE BOTH THE SUBJECT AND INVESTIGATOR IN CLINICAL GENOMICS RESEARCH EFFORTS.”**



*James discusses the results of a western blotting experiment with trainees*

I wanted to go to medical school, so studied chemistry and biology at university. I had always been fascinated with chemistry and learnt to love laboratory studies and the practical work of benchtop empirical science.

### While studying at university, how did you benefit from helping in research laboratories?

In my opinion, this was key; the best way to learn science is by doing science. During my time at New York University, I volunteered in a lab in the medical school. The scientists there were happy to have me clean the dishware. They also taught me how to do real experiments, and I gradually took on more responsibility and fulfilled a technician's role. By my third year of university, I was working in an organic chemistry lab, which was both fascinating and fun. During the summers after my third and fourth years, I participated in an Undergraduate Research Program at Cold Spring Harbor Laboratories on Long Island, New York, where I learnt to do experiments in molecular biology and recombinant DNA research.

### What are your most memorable career moments?

A major highlight was when my lab identified the mutation responsible for CMT – the CMT1A duplication. Also, identifying the gene that causes CMT in my family, through sequencing my own genome, was a huge moment. Human genetics is the study of humans, by humans. It has been an incredible privilege to be both the subject and investigator in clinical genomics research efforts.

### James's top tips

1. Follow your interests, aspirations and dreams!
2. Always keep learning.
3. Don't be afraid to explore, and enjoy the process.