

DNA repeats—the genome's dark matter

November 22 2019



The sequencing device has tiny nanopores that can determine both the DNA

sequence and the epigenetic signature. Credit: MPI f. Molecular Genetics/ Pay Gießelmann

Expansions of DNA repeats are very hard to analyze. A method developed by researchers at the Max Planck Institute for Molecular Genetics in Berlin allows for a detailed look at these previously inaccessible regions of the genome. It combines nanopore sequencing, stem cell, and CRISPR-Cas technologies. The method could improve the diagnosis of various congenital diseases and cancers in the future.

Large parts of the genome consist of monotonous regions where short sections of the genome repeat hundreds or thousands of times. But expansions of these "DNA repeats" in the wrong places can have dramatic consequences, like in patients with Fragile X syndrome, one of the most commonly identifiable hereditary causes of cognitive disability in humans. However, these repetitive regions are still regarded as an unknown territory that cannot be examined appropriately, even with modern methods.

A research team led by Franz-Josef Müller at the Max Planck Institute for Molecular Genetics in Berlin and the University Hospital of Schleswig-Holstein in Kiel recently shed light on this inaccessible region of the genome. Müller's team was the first to successfully determine the length of genomic tandem repeats in patient-derived stem cell cultures. The researchers additionally obtained data on the epigenetic state of the repeats by scanning individual DNA molecules. The method, which is based on nanopore sequencing and CRISPR-Cas technologies, opens the door for research into repetitive genomic regions, and the rapid and accurate diagnosis of a range of diseases.

A gene defect on the X chromosome

In Fragile X syndrome, a repeat sequence has expanded in a gene called FMR1 on the X chromosome. "The cell recognizes the repetitive region and switches it off by attaching methyl groups to the DNA," says Müller. These small chemical changes have an epigenetic effect because they leave the underlying genetic information intact. "Unfortunately, the epigenetic marks spread over to the entire gene, which is then completely shut down," explains Müller. The gene is known to be essential for normal brain development. He states: "Without the FMR1 gene, we see severe delays in development leading to varying degrees of intellectual disability or autism."

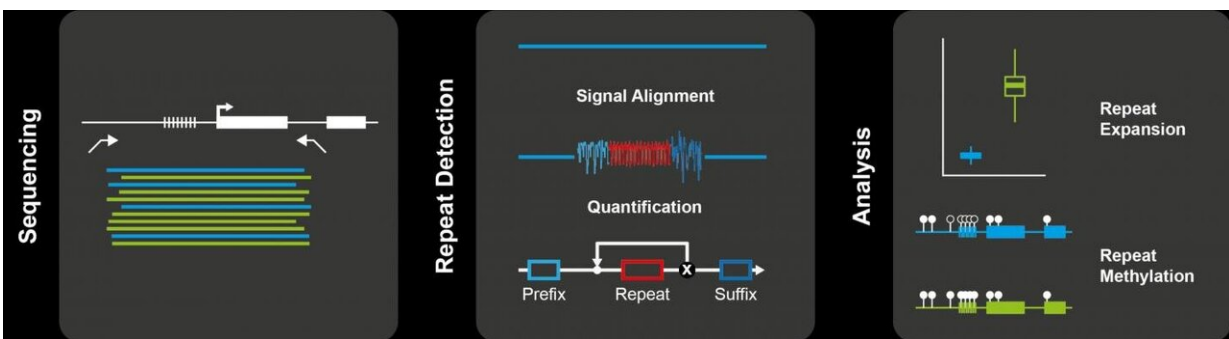
Female individuals are, in most cases, less affected by the disease, since the repeat region is usually located on only one of the two X chromosomes. Since the unchanged second copy of the gene is not epigenetically altered, it is able to compensate for the genetic defect. In contrast, males have only one X chromosome and one copy of the affected gene and display the full range of clinical symptoms. The syndrome is one of about 30 diseases that are caused by expanding short tandem repeats.

First precise mapping of short tandem repeats

In this study, Müller and his team investigated the genome of [stem cells](#) that were derived from patient tissue. They were able to determine the length of the repeat regions and their epigenetic signature, a feat that had not been possible with conventional sequencing methods. The researchers also discovered that the length of the repetitive region could vary to a large degree, even among the cells of a single patient.

The researchers also tested their process with cells derived from patients that contained an expanded repeat in one of the two copies of the C9orf72 gene. This mutation leads to one of the most common

monogenic causes of frontotemporal dementia and amyotrophic lateral sclerosis. "We were the first to map the entire epigenetics of extended and unchanged repeat regions in a single experiment," says Müller. Furthermore, the region of interest on the DNA molecule remained physically wholly unaltered. "We developed a unique method for the analysis of single molecules and for the darkest regions of our genome—that's what makes this so exciting for me."



The researchers proceeded step by step and first selectively enriched molecules with the repeat, analyzed the electrical signal, and then determined the length of the repeats and their epigenetic signature. Credit: MPI f. Molecular Genetics/ Pay Gießelmann

Tiny pores scan single molecules

"Conventional methods are limited when it comes to highly repetitive DNA sequences. Not to mention the inability to simultaneously detect the epigenetic properties of repeats," says Björn Brändl, one of the first authors of the publication. That's why the scientists used nanopore sequencing technology, which is capable of analyzing these regions. The DNA is fragmented, and each strand is threaded through one of a hundred tiny holes ("nanopores") on a silicon chip. At the same time,

electrically charged particles flow through the pores and generate a current. When a DNA molecule moves through one of these pores, the current varies depending on the chemical properties of the DNA. These fluctuations of the electrical signal are enough for the computer to reconstruct the genetic sequence and the epigenetic chemical labels. This process takes place at each pore and, thus, each strand of DNA.

Genome editing tools and bioinformatics illuminate "dark matter"

Conventional sequencing methods analyze the entire genome of a patient. Now, the scientists designed a process to look at specific regions selectively. Brändl used the CRISPR-Cas system to cut DNA segments from the genome that contained the repeat region. These segments went through a few intermediate processing steps and were then funneled into the pores on the sequencing chip.

"If we had not pre-sorted the molecules in this way, their signal would have been drowned in the noise of the rest of the genome," says bioinformatician Pay Giesselmann. He had to develop an algorithm specifically for the interpretation of the electrical signals generated by the repeats: "Most algorithms fail because they do not expect the regular patterns of repetitive sequences." While Giesselmann's program "STRique" does not determine the genetic sequence itself, it counts the number of sequence repetitions with high precision. The program is freely available on the internet.

Numerous potential applications in research and the clinic

"With the CRISPR-Cas system and our algorithms, we can scrutinize any section of the genome—especially those regions that are particularly

difficult to examine using conventional methods," says Müller, who is heading the project. "We created the tools that enable every researcher to explore the dark matter of the [genome](#)," says Müller. He sees great potential for basic research. "There is evidence that the repeats grow during the development of the nervous system, and we would like to take a closer look at this."

The physician also envisions numerous applications in clinical diagnostics. After all, repetitive regions are involved in the development of cancer, and the new method is relatively inexpensive and fast. Müller is determined to take the procedure to the next level: "We are very close to clinical application."

More information: Pay Giesselmann et al, Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing, *Nature Biotechnology* (2019). [DOI: 10.1038/s41587-019-0293-x](https://doi.org/10.1038/s41587-019-0293-x)

Provided by Max Planck Society

Citation: DNA repeats—the genome's dark matter (2019, November 22) retrieved 28 April 2024 from <https://phys.org/news/2019-11-dna-repeatsthe-genome-dark.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.