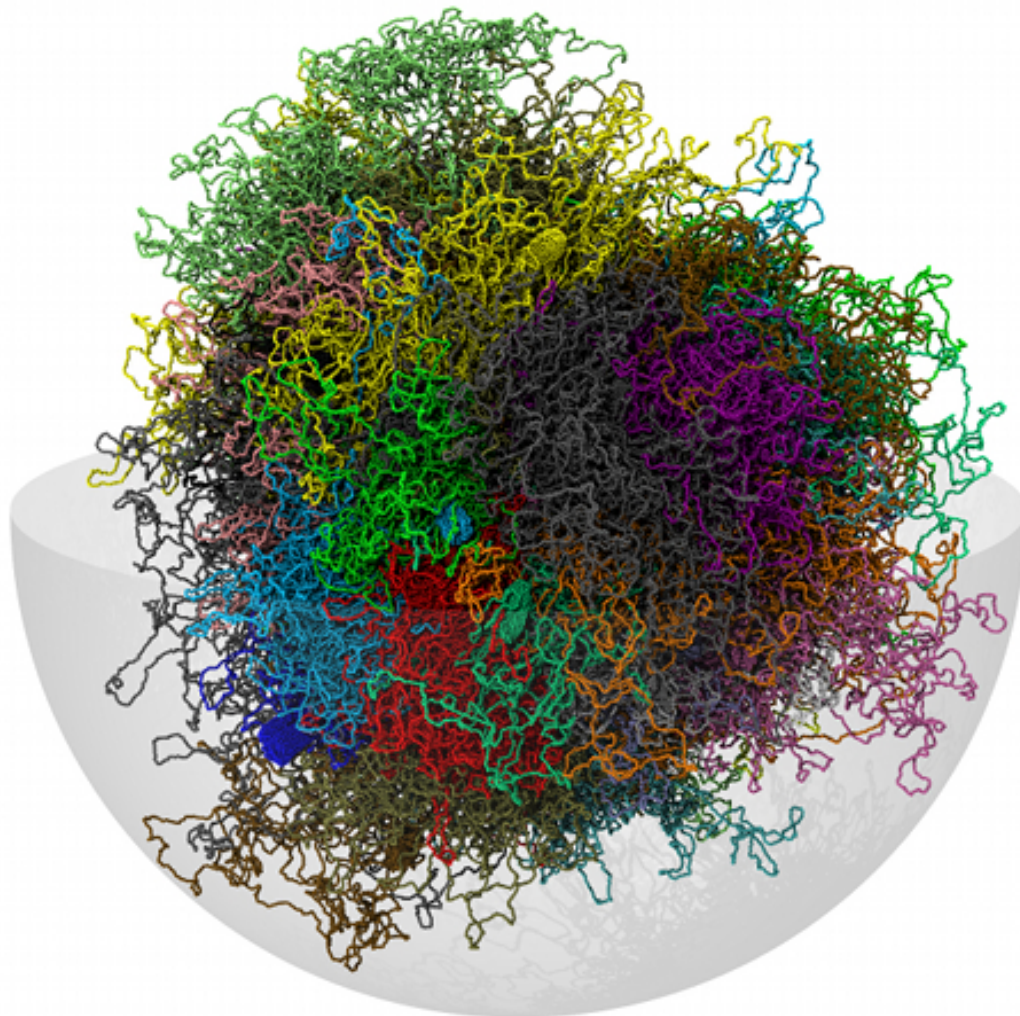# "Neighbour maps" reveal the genome's 3D shape

A new method provides a three-dimensional identikit of chromosomes

October 27, 2016

A group coordinated by the International School for Advanced Studies (SISSA) in Trieste has built a three-dimensional computer model of the human genome. The shape of DNA (as well as its sequence) significantly affects biological processes and is therefore crucial for understanding its function. This new study has provided a first three-dimensional, approximate but realistic, identikit of the human genome. Thanks to the characteristics of the new method, the structural

reconstruction based on both experimental information and statistical methods will be refined as new experimental data become available. The study, carried out in collaboration with the University of Oslo, has just been published in *Scientific Reports* (a journal of the Nature group).

Genome sequencing is a milestone in modern biology as it allows access to the entire "list of instructions" (the chemical sequence of genetic makeup) for the development and function of organisms. Sequencing the genome is a bit like writing down the exact order of the colour of beads in a necklace: knowing how they are arranged along the thread gives us no indication as to the shape of the necklace. The shape of the DNA strand can be highly complex, given that the chromosomes are loosely arranged in an apparently chaotic tangle in the cell nucleus. Since the shape of chromosomes may have a decisive effect on their function, it is important that it should be characterised, in part because scientists think the DNA tangle in the nucleus is only apparently chaotic and that it has instead a specific "geography" for each tissue and stage of cell life.

"Arriving at a precise description of the shape of the DNA tangle is unfortunately incredibly complicated", explains Cristian Micheletti, SISSA professor and coordinator of the new study. "In our case, we used experimental data on 'proximity pairs'".

"Imagine having to create a map of a city", he explains, "based only on information like 'the post office is opposite the station', 'the chemist is close to the gym', 'the fruit and vegetable market is near the football field' and so on. If you have only a small number of such statements to go by, your map will be approximate and in some cases indeterminate. But if you have hundreds, thousands or even more, then your map will become increasingly precise and accurate. This is the logic we followed".

"Proximity pairs" therefore refers to information on the closeness of two points on the map. In the case of nuclear DNA, this information was provided by a technique (which Micheletti defines as "brilliant") known as Hi-C, developed by North American research groups in 2010. In this chemical-physical technique, bits of genome located close to each another in the nucleus are tied together and then identified by their sequence. By collecting large numbers of these proximity pairs scientists discovered which points of the chromosomes lie close to each other in the nucleus. While this is today the most powerful technique for investigating DNA organisation in the nucleus, it is still inadequate for inferring its overall shape. "For this reason, we thought we would try to go 'further'", comments Micheletti.

Mapping "neighbours"

"We used a public database of proximity pairs initially derived from a single Hi-C experiment. The database contained information about hundreds of thousands of proximity pairs", explains Marco Di Stefano, a researcher who completed his PhD at SISSA in 2014 (with this project) and first author of the paper. Di Stefano is currently a post-doc at the National Centre for Genomic Analysis in Barcelona. The researchers created a coarse-grained virtual model of all the chromosomes in a "basic" three-dimensional conformation. They then identified the position of the two bits of DNA of each proximity pair, to draw them closer to each other by appropriately bending the strand.

"By repeating this operation for all the proximity pairs known experimentally we obtained a tangled, though not random, structure that revealed the shape of all the human genome chromosomes, which lay concealed within the data", explains Di Stefano. "It goes without saying that the greater the number of pairs are used, the more precise the resulting 3D model".

In actual fact, after this first phase Micheletti and colleagues added a new series of experimental data to the model. "Just as we were working on the project, a new, more detailed set of Hi-C data was published, so that we used those as well", says Micheletti. "To tell the truth, we had some concerns that our new method might not be sufficiently robust and that the new dataset might clash with and 'spoil' the previously obtained 3D model. But, almost to our surprise, we saw that the conformation remained quite similar. Not only, the new data simply refined the model and almost by magic the various areas of the chromosomes went to position themselves in the correct locations of the nucleus. This we find even more convincing than having succeeded in describing the real data to a good degree of approximation, and we hope that in the future the data collected will allow us to reveal with increasing detail the shape of the DNA enclosed in our cells".

USEFUL LINKS:

• **Link to the original paper in Scientific Reports: www.nature.com/articles/srep35985**

IMAGES:

• **Credits: SISSA**

VIDEO:

• **Watch animated chromosomes on youtube: https://youtu.be/bS2cX1of35c**

Contact: