B·DEBATE International Center for Scientific Debate BARCELONA

biocat | Obra Social "la Caixa"

Synopsis

GTEx: THE POST-GENOME PROJECT

20th and 21st April 2017

CO-ORGANIZED BY:

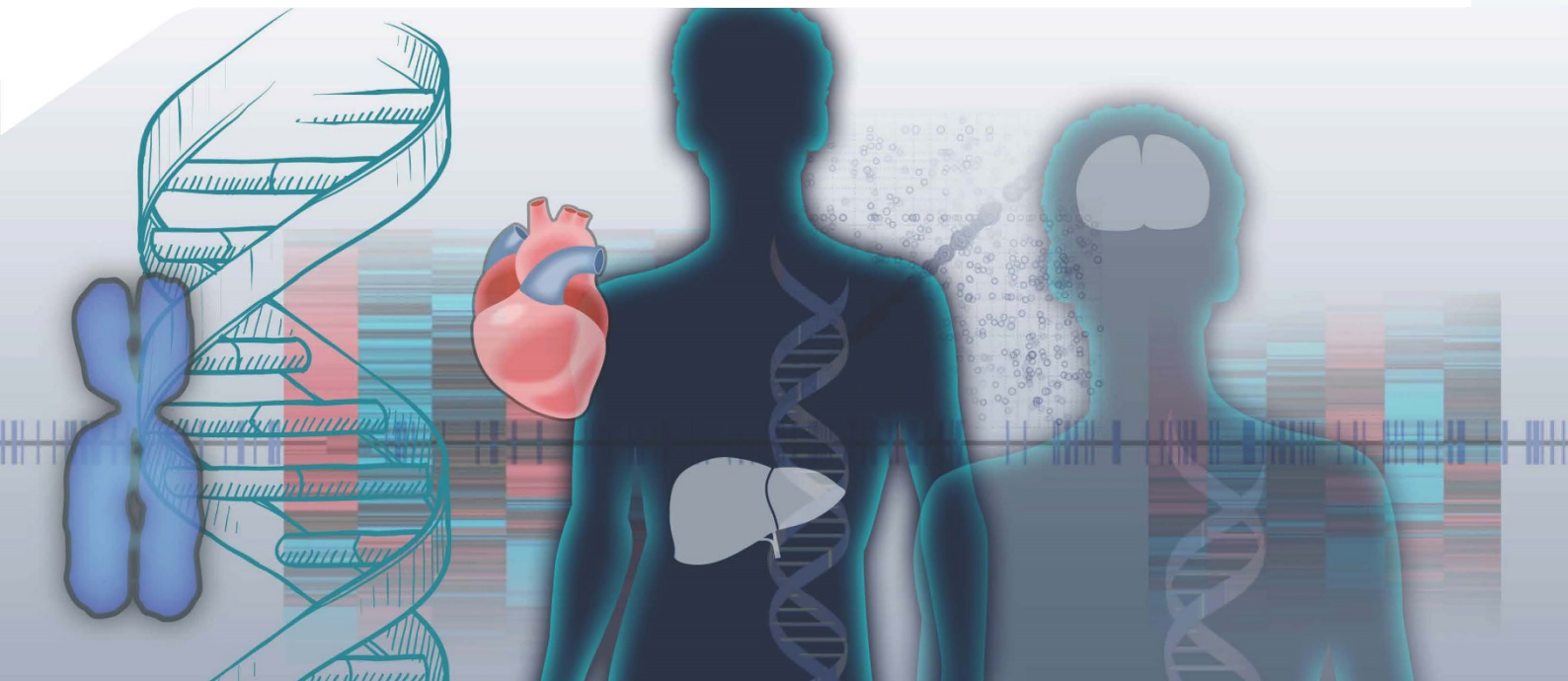
CRG Centre for Genomic Regulation | GTEx

COLLABORATOR:

BROAD INSTITUTE

SPONSORS:

The Company of Biologists | FEDER | PRBB | elixir | INB | QIAGEN | illumina



GTEx: THE POST-GENOME PROJECT

The Human Genome Project, completed in 2003, launched a revolution cut short. It was the first to sequence our DNA and led to the ever-falling price of the techniques that made this large-scale sequencing possible. It came with the promise of helping us understand our biology and the deviations that manifest in illness, but **only scratched the very surface of the darkness**. It showed us that there were many, many more layers underneath.

When looking for associations between DNA and illness, most weren't found in the 2-3% made up of genes but outside of them. They were in what is called the dark genome, which for years was believed to have no purpose. After that, the ENCODE project established, however, that up to 80% of the whole genome is involved in some sort of activity. The majority of dark DNA regulates the activity (reading) of genes. Depending on how it is done in each place, there can be different cells (heart, kidney, brain) even though they all have similar DNA.

The GTEx project (genotype-tissue expression) took it one step further. **Its purpose was to establish the relationship between genetics and expression** (different readings) in the form of RNA and proteins in each type of tissue. These ties should explain the dark associations found in studies. And help understand the basis of how the genome and each disease works. To do so, more than 20,000 tissue samples from nearly one thousand donors are being analyzed and the data is being shared openly.

Members of the project, all international experts, met for the first time in Europe to share the most recent advances. They did so on 20 and 21 April 2017 at the debate 'The Genotype Tissue Expression (GTEx) Project Community Meeting. Enhancing the Usage of Human Genomics for the benefit of all' organized by the Center for Genomic Regulation (CRG) and B·Debate, an initiative of Biocat and the "la Caixa" Foundation to promote scientific debate.

CONCLUSIONS

- The GTEx project has already collected data from more than 20,000 samples from 960 donors. **Studying these will allow this project to go further than the Genome Project**.
- Most associations made between genetics and illnesses found in large studies weren't in genes, so they were hidden. GTEx analyzes these associations and the DNA expression of each sort of tissue, **trying to shed light on the mechanisms involved**.
- Some of the analyses are helping advance fields like **diabetes**, **cardiovascular disease**, **cancer** and **autism**. There is even **pharmacogenetics**, allowing doctors to prescribe drugs based on a patent's genetic profile.
- The information is allowing researchers to **study DNA architecture and how it works**, including the way some regions communicate with others within a certain tissue.

1. A PROJECT-ARCHIVE OF COMPLEXITY

The GTEx project began in 2010. In its founding paper, the consortium that would carry out this project introduced it like this: "**Genome-wide association studies have identified thousands of loci for common diseases, but for the majority of these, the mechanisms underlying disease susceptibility remain unknown**. Most associated variants are not correlated with protein-coding changes, suggesting that polymorphisms in regulatory regions are likely to contribute to many disease phenotypes. The careful examination of gene expression and its relationship to genetic variation has thus become a critical next step in the elucidation of the genetic basis of common disease."

This association is established with eQTLs, a term that hides the relationship between the sequence of letters in the DNA and the greater or lesser chance that the gene it affects will be expressed, first as RNA and later as a protein. **It would be something like the relationship between a book and reading it**: some variants (different letters) make this easier and others complicate it, as if some made a few letters bigger and others translated them into a foreign language. But the variants can act in different ways in different tissues, depending on the cells and their environment, so it's not enough to just

sequence the genome, take a blood sample and measure the RNA or proteins. DNA is shared, but how it is expressed has to be studied individually for each organ, each type of tissue. **The relationship has to be established as if it were a piece of precious metal**. This is the only way the information from the first association studies can be understood.

In 2015, the first results obtained through the project were published. For these, 1,500 tissue samples were analyzed from 175 donors, taken just hours before death. There were already some curious results in that data. For example, **the variation of gene expression was much greater between different organs in a single person than between different individuals**, as explained Roderic Guigó, head of Bioinformatics and Genomics at the Center for Genomic Regulation (CRG) in Barcelona and one of the leaders of this B·Debate. Plus, most of the variation was caused by gender, ethnicity or age. **They found differences associated with being male or female in more than 750 genes**, mostly in mammary tissue. And up to 2,000 genes (10% of the total) changed activity level with age.

Now the project is much further along. As explained Kristin Ardlie, head researcher on the project at the Broad Institute in Massachusetts, they have created "an atlas of gene expression and the eQTLs for **960 donors**." The sample includes "53 tissues, with samples from up to 11 different areas of the brain." **They have collected more than 20,000 samples** and the analysis results are stored in an open archive available to any researcher who wants to use them.

Expanding the samples and data has allowed scientists to prove, for example, that nearly all tissues differ in expression between men and women. As demonstrated by Shmuel Pietrokovski, researcher at the Weizmann Institute of Science in Israel, most of the differences are found in mammary tissue, but also in muscle and adipose tissue (fat). **And even in the anterior cingulate cortex, a region of the brain involved in many cognitive functions**.

Apart from gender differences, the huge amount of data is starting to be analyzed to obtain information on the risks and mechanisms of different diseases, as well as the intimate function of the genome, with its extremely complex architecture and cell messaging.

2. A SHOWCASE AGAINST DISEASE

"According to data from the World Health Organization, the number of people with diabetes has quadrupled since 1980. It now affects more than 400 million people," explained Ana Viñuela, a researcher at the University of Geneva. The attempts to explain their genetic predisposition are a clear example of the obscurity in these relationships. **Up to 90% of genetic variations associated with diabetes are in the dark DNA, beyond the traditional genes**. To find out what role they play, projects like GTEx that combine many layers of information can be highly useful. But with illness like diabetes, which are quite complex with so many mechanisms in play, **which tissues should be analyzed?** Fat, muscle, pancreas? And if we analyze the pancreas, how should we get and study the tiny part of this organ that is responsible for producing insulin?

Some of the conclusions from their studies are, for example, that some regions of the pancreas are the most highly associated with signals shown in genetic studies, but not the only ones. And the adipose tissue under the skin is a fairly reliable mirror for these regions, which could be used to obtain information much more easily.

"Cardiovascular diseases are the leading cause of death in the world, and account for 10% of all spending in the healthcare system in Catalonia," explained Joao Curado, bioinformatics specialists at CRG Barcelona. His team is working with GTEx data to try to identify a **panel of blood-based RNA markers** to diagnose cardiovascular disease easily, non-invasively and, above all, early. This would push treatment forward, making it cheaper and more effective.
To do so, they've divided the project into three phases. In the first, they've identified 100 potential markers of the disease. In the second, currently underway in collaboration with several hospitals in Barcelona, they are trying to reduce and simplify the panel with blood from patients and control subjects (healthy individuals). In the final phase, they will test its real value.

Other useful applications of the project lie in **pharmacogenomics**, the discipline that attempts to personalize drug dosing based on individual genetics. Assaf Gottlieb, a professor at the University of Texas, is working on a panel to predict the necessary dose of warfarin, a very common anti-coagulant. This would prevent doses that are too high, which increase the patient's risk of bleeding, or too low, with the risk of being ineffective. There are genetic panels already but they aren't very reliable. Adding expression data from the GTEx project is allowing them to "improve prediction of the necessary doses."

Furthermore, **"the effects of genetic variants may be different in each person,"** explained Stephanie Castel, a researcher at the New York Genome Center. This is because it depends on their genetic context (other variables that affect them) and environment. Her analyses using GTEx expression data aim to discover why the same variants can lead to **autism** in some individuals but not in others. They are even combining information from GTEx with data from other databases to study whether certain modifications we are born with can foster **cancer**, paving the road for it once the triggering mutations make their appearance.

3. UNDERSTANDING MOLECULAR ARCHITECTURE

The final message of DNA, the information it contains, is a blueprint to manufacture proteins, the bricks and workers of cells. Cecilia Lindskog is a researcher at the University of Uppsala, Sweden, and a member of the Human Protein Atlas, a parallel initiative with ties to GTEx that aims to establish a map, even if temporary, of the proteins present in each of our tissues. Some of their conclusions show that, in addition to the fact that "half of all genes are expressed in proteins in every type of tissue in the body, there are many proteins exclusive to only one tissue." And what could be most important, **"many proteins that could be drug targets are found in all of them."**

On the road to proteins, the GTEx project can help untangle part of the complex world of DNA regulation. For example, part of the dark genome is made up of **"switches"**, regions that activate or block genes from being read. **Many studies have estimated that there are up to 400,000 all over the genome**, but as explained Simon Fishilevich, a student at the Weizmann Institute of Science in Israel, "These tasks have led to many contradictions and redundancies." His team is trying to refine these results by combining data from different archives. And by tying switches to the genes they act on, as their function in many cases isn't yet known. Mutations in some of them, for example, are behind genetic diseases like polydactyly ('extra' fingers or toes) or cases of congenital, untreatable diarrhea.

"The regulatory elements (the switches) are distributed according to a modular architecture, grouped together but separate," concluded Emmanouil Dermitzakis, professor at the University of Geneva and one of the leaders of the GTEx project. **What is curious is that they are associated with target genes that can be near, far or even on different chromosomes, "as if they were speaking."**

But how and how much do they speak? In which tissues? What happens when there are variants? How does the conversation change? This is why GTEx was created. Although, as explained Ayellet Segré, researcher at the Broad Institute in Cambridge (Massachusetts), so far we only know about the visible part of the iceberg.