Researchers discover mechanism linking mutations in the 'dark matter' of the genome to cancer

June 13 2022

Credit: Pixabay/CC0 Public Domain

For many years, the human genome was viewed as a book of life in which sections of great eloquence and economy of expression were

interspersed with vast stretches of gibberish. The legible sections contained the code for making cell proteins; the other regions, representing about 90% of the entire genome, were dismissed as "junk DNA," having no discernable purpose.

Research has taught scientists otherwise. Far from being useless filler, many non-coding sections have been shown to play a key role in regulating gene activity—increasing or decreasing it as needed. For cancer scientists, this has raised questions of their own: If mutations in coding regions cause cells to make flawed proteins, what do mutations in non-coding regions do? How does a mutation in the hinterlands of the genome—in areas devoid of genes—contribute to cancer?

Given that non-coding regions are involved in gene regulation, researchers have hypothesized, naturally, that mutations in these zones play havoc with gene activity in ways conducive to cancer. Study after study, however, has found this generally not to be the case, leaving the biological impact of non-coding mutations something of a mystery.

Thinking locally

In a new paper in the journal *Nature Genetics*, Dana-Farber investigators provided an answer. They did so by the scientific equivalent of thinking locally—narrowing the scope of their investigation to the specific DNA in which non-coding mutations occur. They found that in the overwhelming number of cases examined, such mutations have an epigenetic effect—that is, they change how tightly the DNA at those locations is wrapped. That, in turn, affects how open those locations are to binding to other sections of DNA or certain proteins, all of which can influence the activity of genes involved in cancer.

The discovery reveals, for the first time, a pervasive biological mechanism by which non-coding mutations can influence cancer risk. It

also opens the way to therapies that—by disrupting that mechanism—can reduce at-risk people's likelihood of developing certain cancers.

"Studies have identified an enormous number of mutations across the genome that are potentially involved in cancer," says Alexander Gusev, Ph.D., of Dana-Farber, the Eli and Edythe L. Broad Institute and Brigham and Women's Hospital, who co-authored the paper with Dana-Farber's Dennis Grishin, Ph.D. "The challenge has been understanding the biology by which these variations increase cancer risk. Our study has uncovered an important part of that biology."

Does mutation change expression?

To identify inherited—or germline—mutations that increase a person's risk of developing cancer, investigators conduct what are known as [genome-wide association studies](link), or GWASs. In these, researchers collect blood samples from tens or hundreds of thousands of people and scan their genomes for mutations or other variations that are more common in people with cancer than in those without the disease.

Such tests have yielded thousands of such mutations, but only a small percentage of them are in coding portions of the genome that are relatively easy to link to cancer. Breast cancer is one example. "More than 300 mutations have been identified that are associated with an increased risk of the disease," Gusev states. "Less than 10% of them are actually within genes. The rest are in 'desert' regions, and it hasn't been clear how they influence disease risk."

To try to make that connection, researchers gather two sets of data: one, GWAS data showing mutations in a specific type of cancer; and two, data on another genomic feature of that cancer type—such as an abnormally high or low level of activity in certain genes. By looking for

areas of overlap between these data sets, in a process called colocalization, researchers can determine whether the mutations correspond with a rise or fall in the activity of those genes. If such a relationship exists, it would help explain how non-coding mutations can lead to cancer.

Despite a massive investment in this type of research, however, colocalization studies have turned up very few such correspondences. "The vast number of mutations identified by GWASs have been found to have no colocalizing gene at all," Gusev remarks. "For the most part, non-coding mutations associated with cancer risk don't overlap with the changes in gene expression [activity] documented in public data sets."

Looking closer to home

With that route looking increasingly unenlightening, Gusev and Grishin tried another more fundamental approach. Instead of beginning with the premise that non-coding mutations might influence gene expression, they asked how they alter their home environment—whether they affect the coiling of DNA in their immediate vicinity.

"We hypothesized that if you look at the effect of these mutations on local epigenetics—specifically, whether they caused nearby DNA to be wound more tightly or loosely—we'd be able to detect changes that wouldn't be evident in expression-based studies," Gusev relates.

Their reasoning: "If a mutation has an effect on disease, that effect will probably be too subtle to capture at the level of gene expression but may not be too subtle to capture at the level of local epigenetics—what is happening right around the mutation," Gusev says.

It's as if previous studies sought to understand how a brush fire in California could affect the weather in Colorado, whereas Gusev and

Grishin wanted to see its effect on the hillside where it began.

To do that, they performed a different type of overlay study. They took GWAS data on cancer-related mutations and data on epigenetic changes in seven common types of cancer and examined whether—and where—they intersected.

The results came in stark contrast to those from colocalization studies. "We found that whereas most non-coding mutations don't have an effect on gene expression, most of them do have an impact on local epigenetic regulation," Gusev states. "We now have a basic biological explanation of how the vast majority of cancer-risk mutations are potentially linked to cancer, whereas previously no such mechanism was known."

Using this approach, the researchers created a database of mutations that can now be linked to cancer risk by a known biological mechanism. The database can serve as a starting point for research into drugs that—by targeting that mechanism—can lower an individual's risk of developing certain cancers.

"If we know, for example, that a certain transcription factor [a protein involved in switching genes on and off] binds to one of these cancer-associated mutations, we may be able to develop drugs targeting that factor, potentially reducing the likelihood that people born with that mutation will contract cancer," Gusev says.