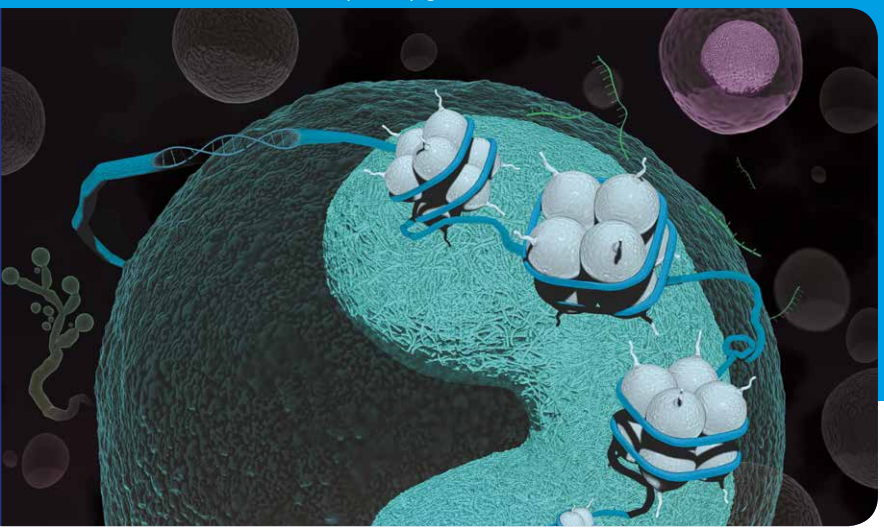


A cut-and-paste operation that defines cell fates

If a person's cells all carry the same genetic information, how can they take on dramatically different forms and functions to fulfill the needs of various tissues and organs? One answer lies in the fact that each type of cell activates a unique subset of its genes to produce RNA and protein molecules. In the past, most efforts to understand the way hematopoeitic (blood) stem cells differentiate into so many different cell types have focused on the processes that activate or silence genes. But the pathway from gene to protein is regulated at many steps, and each offers the cell a possibility to fine-tune its population of molecules.

Recently BLUEPRINT scientists obtained the closest look ever at the role of one of these steps, called alternative splicing, in blood cell development. The project, headed by scientists at the Cambridge University and the Sanger Center of the Wellcome Trust in Cambridge (UK), has revealed that this process produces specialized forms of many proteins that go on to play a crucial role in several steps of blood cell development. The work has yielded new insights into the extent to which cells use alternative splicing, and is helping to explain the reasons that mutations in genes cause a range of blood diseases. The work appeared in the Sept. 26, 2014, edition of Science.

Cells are often nudged along particular pathways of specialization when they begin to produce proteins called transcription factors. These molecules are often unique to specific types of cells; they dock onto DNA and stimulate the transcription of DNA sequences into RNA molecules. Some are messenger RNAs (mRNAs), which bear the recipes for protein molecules. In the past, most studies of cell development have focused on identifying transcription factors and how they regulate genes in specific types of cells. The study now shows that other epigenetic mechanisms may play an equally important role in determining a cell's fate.

Genes have a complex structure in which sequences called exons, which contain protein-encoding information, are separated by long regions of non-coding information (introns). An RNA contains both when it is first made as a copy of a gene, but the introns have to be removed before it can function as an mRNA. This job is carried out by a molecular machine called the spliceosome. It recognizes an intron and cuts the strand of RNA before an intron, then again after it. The intron is discarded and the spliceosome rejoins the loose ends to create a final, strand-like mRNA. Genes can have many introns and exons, and splicing occasionally also removes specific exons. This process is called alternative splicing, and it permits cells to create various proteins based on the information in a single gene. Such molecules can have different "meanings" in cell development, the way the meaning of a sentence can change by removing a word. ("My mother ate the cat food by mistake" has a quite different meaning if you remove the word "food".)

"Alternative splicing is known to play a role in many aspects of the lives of cells," says Augusto Rendon, a scientist at the University of Cambridge who played a leading role in this BLUEPRINT project. "And defects in this process have been linked to blood cancers and a number of other diseases. This can occur when cells fail to remove noncoding information, or are unable to produce a particular version of a protein that is needed for some important function. But it has been extremely difficult to capture a picture of the full scope of alternative splicing and its impact on processes such as cell differentiation."

Recent dramatic reductions in the speed and cost of sequencing and other methods have now made it possible to carry out such an assessment. Augusto, his colleague Mattia Frontini, and their BLUEPRINT partners decided to conduct a global study of the way alternative splicing influences blood cell development.

From umbilical cord blood, the scientists obtained hematopoeitic stem cells (HSCs) and five other types of cells that represented crucial steps along the way to producing fully specialized blood. One was grown in cell cultures and stimulated to develop into two more types. These eight kinds of cells represented major "decision-making" stages in HSC specialization – points at which cells diverge to acquire different fates.

Now the researchers began studying the vast quantities of RNA molecules found in the cells using "next generation sequencing" technologies. These highly sensitive methods are able to pick up molecules produced in low quantities and can also provide quantitative information. This was important because many cells produce different versions – called isoforms – of the same protein. Cells respond not only to the presence or absence of a particular variant, but whether it is produced at higher or lower levels than other isoforms.

A massive computational effort was required to compile and analyze the data, one of the specialties of bioinformatics teams at the Sanger Center and their campus neighbor, the European Bioinformatics Institute (EBI). The EBI hosts enormous, publicly accessible databases of gene, RNA, and protein sequences and many other types of biological data. The center played a leading role in assembling and interpreting data from the Human Genome Project, information on human variation and mutations found in diseases, and the genomes of a huge number of other species and is participating in the International Human Epigenome Project (IHEC).

Computational methods were used to analyze the sequences of RNA molecules captured from the various types of blood cells. Sequences were compared to each other – and to the human genome sequence. This revealed thousands of cases in which particular coding regions were missing from RNA molecules, representing different isoforms created by splicing. (Finding "sentences" like "My mother ate the cat," "My mother ate the cat food," and "My mother ate the cat food by mistake" would point to a single gene that had undergone three different types of alternative splicing.)

During their analysis the scientist made a second important discovery. Very early in hematopoeitic development, cells not only produce a large number of alternative spliced mRNAs but also a high number of noncoding mRNAs. At first glance, noncodingRNAs look like normal mRNAs with joined exons, but their sequences do not code for proteins. They exert their functions by forming a structure that helps regulate whether a protein-coding mRNA is created in the first place or is used to produce proteins. The cells seem to use two methods – noncoding RNAs and alternative splicing – as a means of regulating the production of proteins:

The analysis revealed a huge number of cases – about 7000 – in which alternative splicing had generated molecules specifically used in one of the eight types of hematopoeitic cells. For some of these novel splice isoforms it appeared that they are involved in pushing cells toward a particular developmental pathway.

To validate these findings, the scientists performed experiments to show that a particular isoform was truly essential to a specific stage of blood development, because blocking its production alters the cell's fate.
One example seemed particularly interesting. The study revealed a new splice form of a transcription factor called NFIB. NFIB was known to activate genes in various tissues but its role in blood cells was rather unclear. It was known that the NFIB is silenced in hematopoeitic stem cells (the parent of all other blood cell types) when they take their first step of development.

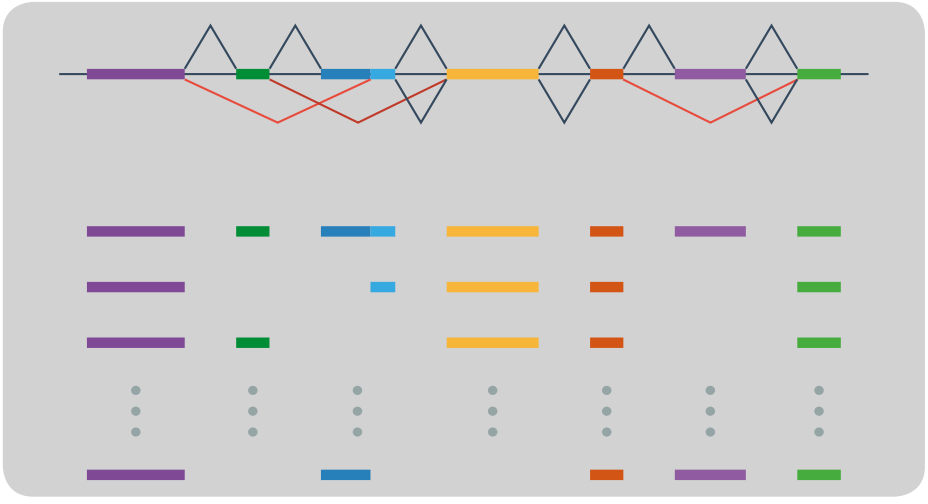
The scientists now found that NFIB is also expressed as a new variant produced by splicing in differentiated blood cells called megakaryocytes (MKs), which are important for blood clotting. MK cells manufacture a short form of NFIB protein lacking the first exon; as a result, the protein lacks a structure needed for its binding to DNA. Without the exon, this form of the NFIB was unable to activate its target genes.

But the protein also seems able to perform a second job. When scientists blocked the production of this isoform in megakaryocyte erythroyte progenitors – which give rise to MKs and red blood cells – they observed a sharp reduction in the cells' differentiation into mature MKs. This meant that cells needed for blood clotting are not produced. Conversely, artificially increasing the amount of the short form of NFIB in the progenitors yielded more MK cells. The two procedures verified that the cells needed enough of the short version to differentiate into red blood cells, and that its role was essential in this process.

"The results of the project," Augusto Rendon says, "represent a huge leap in our understanding of the role that alternative splicing plays in a developmental system. We have opened a new window on a type of epigenetic control whose full extent and influence have been poorly understood. We have gained a wealth of new data about the connection between specific isoforms and stages of blood cell development that will help us understand some of the effects of defective splicing in a number of important diseases. The splicing sites and events BLUEPRINT has discovered will help scientists probe other systems and tissues throughout the body. The information we have gained has been put into public databases so that it can be freely accessed by scientists, physicians, and many others across the world."

Reference:

- *Chen et al. Transcriptional diversity during lineage commitment of human blood progenitors. Science 345, 1251033 (2014).*



Most genes contain several protein-encoding sequences (exons, colored bars). During alternative splicing, some may be removed to create proteins with different contents, which give the molecules different functions in various types of cells and tissues.