Our bodies are composed of trillions of cells, each with its own job. Cells in our stomach help digest our food, while cells in our eyes detect light, and our immune cells kill off bugs. To be able to perform these specific jobs, every cell needs a different set of tools, which are formed by the collection of proteins that a cell produces. The instructions for these proteins are written in the approximately 20,000 genes in our DNA.

Despite all these different functions and the need for different tools, all our cells contain the exact same DNA sequence. But one central question remains unanswered – how does a cell know which combination of the 20,000 genes it should activate to produce its specific toolkit?

The answer to this question may be found in the pieces of DNA that lie between our protein-producing genes. Although our cells contain a lot of DNA, only a small part of this is actually composed of genes. We don't really understand the function of most of this other sequence, but we do know that some of it has a function in regulating the activity of genes. An important class of such regulatory DNA sequences are the enhancers, which act as switches that can turn genes on in the cells where they are required.

However, we still don't understand how these enhancers know which genes should be activated in which cells. It is becoming clear that the way DNA is folded inside the cell is a crucial factor, as enhancers need to be able to interact physically with genes in order to activate them. It is important to realise that our cells contain an enormous amount of DNA – approximately two meters! – which is compacted in a very complex structure to allow it to fit into our tiny cells. The long strings of DNA are folded into domains, which cluster together to form larger domains, creating an intricate hierarchical structure. This domain organisation prevents DNA from tangling together like it would if it were an unwound ball of wool, and allows specific domains to be unwound and used when they are needed.

Researchers have identified key proteins that appear to define and help organise this domain structure. One such protein is called CTCF, which sticks to a specific sequence of DNA that is frequently found at the boundaries of these domains. To explore the function of these CTCF boundaries in more detail and to investigate what role they may play in connecting enhancers to the right genes, our team studied the domain that contains the ?-globin genes, which produce the haemoglobin that our red blood cells use to circulate oxygen in our bodies.

Firstly, as expected from CTCF's role in defining boundaries, we showed that CTCF boundaries help organise the ?-globin genes into a specific domain structure within red blood cells. This allows the enhancers to physically interact with and switch on the ?-globin genes in this specific cell type. We then used the gene editing technology of CRISPR/Cas9 to snip out the DNA sequences that normally bind CTCF, and found that the boundaries in these edited cells become blurred and the domain loses its specific shape. The ?-globin enhancers now not only activate the ?-globin genes, but cross the domain boundaries and switch on genes in the
This study provides new insights into the contribution of CTCF in helping define these domain boundaries to help organise our DNA and restrict the regulation of gene activity within the cells where it is needed. This is an important finding that could explain the misregulation of gene activity that contributes to many diseases. For example in cancer, mutations of these boundary sequences in our DNA could lead to inappropriate activation of the genes that drive tumour growth.

The full study, 'Tissue-specific CTCF–cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo', can be read in the journal Nature Cell Biology.