

A vast field of galaxies, including spiral, elliptical, and irregular shapes, scattered across a dark blue background. The galaxies are in various colors, including yellow, orange, red, and white, with some showing prominent spiral arms. Two bright, multi-colored stars with lens flare effects are visible in the upper left and center-left areas.

The dark side
of the cell



Gazing up at the starry sky of a balmy summer's night is one of life's simple pleasures. As you do so, ponder this: there is a surprising parallel between the vast majesty of the cosmos and the inner workings of your body's cells. The universe and you share a similar scientific mystery, one that Lars Steinmetz and his team at EMBL Heidelberg are helping to solve.

Cosmologists have long sought to explain why the stars and galaxies are scattered across the cosmos in the way that they are. There simply isn't enough normal matter around to account for the gravity to explain why the universe keeps expanding. So they have proposed the existence of invisible "dark" matter, which exists in greater amounts than visible matter, to account for it. No-one has yet seen this mysterious substance, but if it does exist, much about the universe suddenly makes sense.

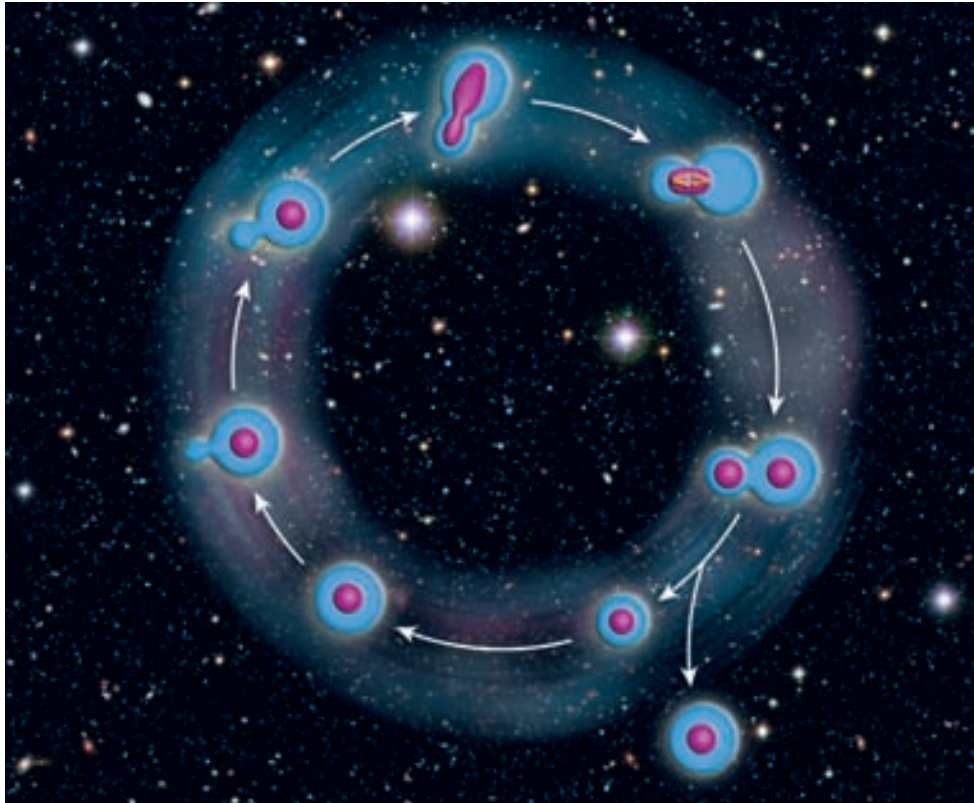
Molecular biologists have their own equivalent of dark matter to puzzle over. In many ways, their problem is the opposite of the one faced by cosmologists: the dark matter of the cell is entirely visible, yet its function is a mystery. This matter takes the form of molecules that, according to theory, really shouldn't be there. Yet they are there, in spades, and finding out what they do could plug a gap in our understanding of how our cells and bodies work.

“I hope our high-resolution ‘satellite imagery’ [...] will stir the imagination of scientists.”

The molecules in question are RNA molecules. RNA is a large molecule that is similar to DNA and is essential to all forms of life. Historically, however, biologists underestimated the extent of the role of RNA in the cell, thinking that it acted mainly as the cellular equivalent of a courier and a bricklayer. This idea had its roots in the discovery of how genes spell out the instructions for making a protein. A gene consists of a sequence of DNA "letters" that encode instructions for stringing together the building blocks, or amino acids, that make up a particular protein. When the cell makes a protein, it begins by reading the gene and making an RNA copy of the sequence of letters. This process, called transcription, produces a kind of RNA called messenger RNA. This makes its way to the cell's protein-making factories, where molecular robots called ribosomes read the RNA and piece together the correct sequence of amino acids.

But as molecular biologists started exploring genomes, they ran into a mystery. There seemed to be long stretches of DNA that didn't seem to code for proteins at all. The sequencing of the human genome confirmed this: less than two per cent of our genome consists of protein-coding genes. The rest was apparently functionless "junk" DNA. A few years later, work from a number of laboratories including EMBL-EBI and EMBL Heidelberg, revealed that almost all of this "junk" was being transcribed into RNAs of differing lengths. Because they weren't involved in producing proteins, these RNAs were named "non-coding RNAs". As they vastly outnumbered the messenger RNAs and yet had no known function, they soon became known as the dark matter of the genome.

During the past few years, scientists have discovered that some of these RNAs – short molecules called microRNAs and small interfering RNAs – are involved in controlling gene activity. But the function of the longer non-coding RNAs remains mysterious. So Lars and his team decided to take a closer look at what happens in yeast cells. Yeast have much of their basic biology in common with our cells, but unlike human cells, they are easy to manipulate with genetics and molecular biology. Biologists can, for example, disable a single yeast gene or RNA with great precision, allowing them to work out its function.



The scientists measured all the RNAs produced during yeast's cell cycle.

Like the human genome, the yeast genome churns out vast quantities of non-coding RNA. Lars' team was the first to profile and publish the total RNA output – or “transcriptome” – of a yeast called *Saccharomyces cerevisiae*, also known as baker's yeast. As well as being used for baking and for brewing beer, this organism is commonly used in molecular biology studies. “We were very surprised because we saw all these RNAs showing up in regions where we didn't expect to see them,” says Lars.

So the team took a closer look at the mystery RNAs. Some came from parts of the genome that lay in large gaps between protein-coding genes, yet showed no signs of coding for proteins themselves. Others were more mysterious still. They were produced from the same genes that code for normal proteins, but were a kind of mirror image of the messenger RNA produced by that gene.

Normally, when a cell makes a messenger RNA, it begins by unzipping the double helix of the DNA that encodes the gene. DNA is made up of two intertwining strings of molecular letters, or bases. When a cell reads a gene to make a messenger RNA for a protein, it only reads one of the two strands. This coding strand is called the “sense” strand. The opposite strand is called the “antisense” strand and until now, biologists believed that it was almost never transcribed. Lars' data, however, suggested that these mirror-image-like antisense transcripts were commonplace.

To try to get a handle on what some of these RNAs might be doing, Lars' group teamed up with bioinformaticians Peer Bork, at EMBL Heidelberg, and Wolfgang Huber, then at EMBL-EBI. With the help of the Genomics Core Facility, they measured how the production of these RNAs varied over the course of a cell's life, as it went through the repetitive cycle of growing and dividing into two new cells. This process, called the cell cycle, is fundamental to the biology of all cells. What's more, abnormalities in the cell cycle are hallmarks of a number of human diseases, particularly cancer.

If the production of certain RNAs varied over the course of the cell cycle, Lars reasoned, this would suggest that they might somehow be involved in either controlling or executing it. So the

team undertook a heroic set of experiments: to sample yeast cells every five minutes over the course of three cell cycles (each lasting an hour or more) and measure all their RNAs.

Halfway through the experiments, disaster struck. The team realised that one of the reagents they were using was creating mistakes in the data: producing “phantom” transcripts that didn’t really exist in the cell. Fortunately, they devised a way to solve the problem, but they still had to repeat all their work. “It was a huge effort,” says Lars. But the hard work paid off – 18 months later, the team had high-quality results. “The data that we published are fresh, clean data,” says Lars.

The data have resulted in a highly detailed “atlas” describing all the transcripts in yeast and how they vary during the cell cycle. Some non-coding transcripts from the gaps in between protein-coding genes were transcribed in a periodic fashion, others were not. Some antisense transcripts were periodic, as were their sense counterparts. But other periodic antisense transcripts were twinned with non-periodic sense transcripts, and vice-versa: some periodic sense transcripts had non-periodic antisense opposites. Just to make things even more complicated, many transcripts seemed to come from overlapping sections of the genome.

The big question is: what are they all doing? “These are all very hotly debated areas,” says Lars. “We cannot be conclusive at the moment in saying how many of these are functional.” One key argument in the debate is whether these non-coding RNAs have any function at all, or whether they are simply produced accidentally by the cell’s transcription machinery being a bit “overenthusiastic”. Lars is sceptical of this idea: “I am of the belief that if something happens in biology it’s not just noise, there must be a reason for it,” he says. He points out that the cell has to invest a lot of energy into producing all that RNA: “Transcription costs something for the cell.”

The team is conducting experiments to test out their hypotheses, and have already uncovered hints that antisense transcripts could help fine-tune gene expression (see page 96). As well as revealing more about what these RNAs are doing, the atlas promises to give new insights into how the cell cycle works. “The high resolution of this dataset allows one to get a much more accurate way to model the cell cycle,” says Lars. At the moment, such models are based on the behaviour of protein-coding genes. Lars’ data will allow scientists to not only include the information on the non-coding transcripts, they will also be able to benefit from the more detailed protein-coding transcript data the study has provided. Another benefit is the innovation Lars and his team developed to counter the problem reagent that creates false results. Molecular biology laboratories will be able to produce more reliable results by following Lars’ modified method. “I hope our high-resolution ‘satellite imagery’ of the global regulation of all transcripts during the yeast cell cycle will stir the imagination of scientists to further explore their favourite regulatory patches in more detail, perhaps conferring an as yet unknown functional role to the new non-coding RNA molecules we have mapped,” says Marina Granovskaya, a postdoc in Lars’ lab who performed the yeast experiments.

So biologists now have a starting point: the entire constellation of yeast RNAs and how they vary through the cell cycle laid out in fine detail, an atlas to help them navigate their way to a better understanding of what this mysterious cellular “dark matter” is doing. The only question that remains is whether they will beat the cosmologists to finding an answer.

Granovskaia M, Jensen L, Ritchie ME, Toedling J, Ning Y, Bork P, Huber W, Steinmetz LM (2010) High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol* **11**: R24