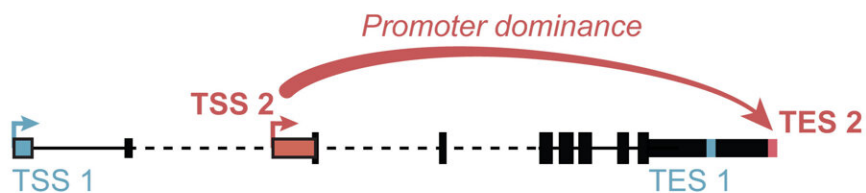


The beginning is the end: How promoters predefine where genes end

May 12 2023

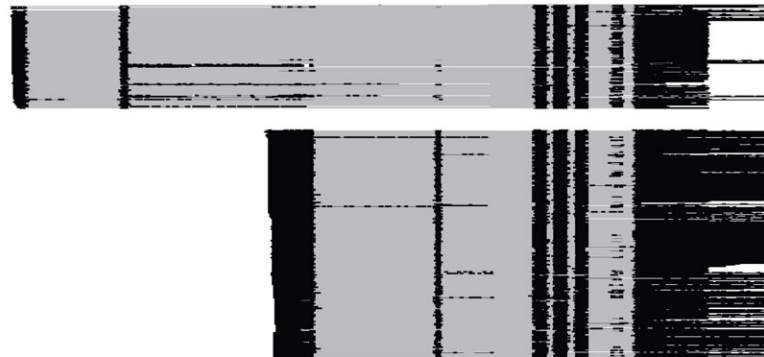
Model of the gene *stathmin*



Conventional (short-read) mRNA sequencing traces



Nanopore long-read-sequencing: individual reads



Example of a gene that contains two possible start sites (TSS), and two possible end sites (TES). Black boxes in the gene model show sequences that will be translated into protein. With conventional short-read mRNA sequencing, in which signal represents the accumulation of reads, it is not possible to distinguish

different molecules. In long-read sequencing, each horizontal line represents one mRNA molecule from beginning to end. For the gene *stathmin*, we can see that TSS1 predetermines ending at TES1, and TSS2 will lead to TES2. Credit: MPI of Immunobiology & Epigenetics, Hilgers

Each gene in our DNA has a beginning and an end. Defining the gene's extremities properly is crucial in producing functional protein. Much research has been done to identify what determines when, where, and at which site on the DNA a gene "starts." But where a gene ends is a different story—selection of transcription termination sites has been assumed to depend on downstream elements and extrinsic factors.

In their most recent study published in the journal *Cell*, researchers from the Max Planck Institute of Immunobiology and Epigenetics made the surprising finding that for most of our genes, the site of transcription start determines the site of transcription end. This phenomenon is well-conserved across species and pre-determines mRNA end sites at the very beginning of transcription, and plays a crucial role in [cell identity](#) and functionality.

All cells in an organism contain an identical DNA sequence. What determines the identity and function of individual cells and tissues is the set of genes that will be active in a given place, at a given time. These active genes are transcribed from the DNA template into distinct messenger RNA (mRNA) molecules and will encode the proteins the cell needs to function.

At specific places called promoters, a complex molecular machinery starts transcribing DNA sequences into mRNA. Interestingly, most genes contain multiple possible sites where transcription can start or end. This means that for each gene, depending on the start or termination site, the

mRNAs can be different. Expressing one gene in different variants expands the diversity and functionality of the genome many times over. At the same time, it adds another layer of complexity to the study of the genome.

RNA snapshots from beginning to end

Scientists at the Max Planck Institute of Immunobiology and Epigenetics in Freiburg wanted to know how many different start and end sites each gene uses, in which combination, and whether the combinations were different in different conditions. "The technical problem to answer this question is that we have to read each and every mRNA molecule from all genes from the very beginning to the very end. This a humongous task that has not been undertaken before," says Valérie Hilgers, a research group leader at the MPI-IE.

The scientists used a tweaked next-generation sequencing technology to read out the individual mRNAs. For conventional short-read sequencing, each mRNA is broken into shorter fragments that are amplified and then sequenced to produce the read. Bioinformatic techniques are then used to piece together the reads like a jigsaw, into a continuous sequence.

For full-length mRNA information of the entire genome in several *Drosophila* tissues, including the brain, the Hilgers teamed up with the Deep Sequencing Facility of the MPI to optimize specific long-read-sequencing technologies. "Long-read sequencing allows for the retrieval of much longer sequencing reads than widely used standard sequencing. However, we even had to optimize this technology and increase the typical read length by several fold to obtain full-length mRNA information in our different model systems," says Carlos Alfonso-Gonzalez, the first author of the publication.

In addition to *Drosophila*, the Hilgers Lab also included a human model

of the nervous system into their study: [cerebral organoids](#)—"mini-brains" cultured in a dish from induced pluripotent stem cells. Transcription end sites were pre-determined at transcription start.

The gathered data representing each mRNA at the full-molecule scale give unprecedented insight into the transcription of individual [genes](#) "We realized that far from start sites (TSSs) and end sites (TESs) being randomly combined one to another, we found that often, sites of transcription start are specifically linked to distinct sites of transcription end," says Hilgers.

This linkage is actually causal: in ovaries, for example, the artificial activation of a TSS that is normally only used in the brain overrides the normal TES and artificially induced the use of the brain TES. This shows the critical role of TSS in shaping the RNA landscape unique to each tissue, and thereby influencing tissue identity.

Promoter dominance drives RNA diversity, gene function and tissue identity

However, one phenomenon stood out. "Certain TSSs show unexpected dominance behavior. They overrule conventional signals to end [transcription](#), outcompete other TSSs, and cause the selection of distinct TESs. Accordingly, we named them dominant promoters," says Alfonso-Gonzalez.

Furthermore, the team found that interactions between these dominant promoters and their associated gene ends was guided by distinct epigenetic signatures. Importantly, the results in *Drosophila* brain cells could be replicated in the human brain organoids, showing that promoter dominance is a conserved, perhaps universal, mechanism for regulating the production of functional proteins and the cells' functionality.

What could be the physiological relevance of this novel mechanism? Through an in-depth sequence conservation analysis, the Freiburg researchers discovered that TSSs and TESs exhibit co-evolution: over millions of years of evolution between species, individual nucleotide changes in the gene start at dominant promoters were accompanied by changes at the corresponding gene end.

"We interpret this observation as a push through evolution, to sustain the interaction between both extremities of the gene, which implies significant importance of these couplings for animal fitness," says Valérie Hilgers.

More information: Valérie Hilgers, Sites of transcription initiation drive mRNA isoform selection, *Cell* (2023). [DOI: 10.1016/j.cell.2023.04.012](https://doi.org/10.1016/j.cell.2023.04.012).
[www.cell.com/cell/fulltext/S0092-8674\(23\)00408-7](https://www.cell.com/cell/fulltext/S0092-8674(23)00408-7)

Provided by Max Planck Institute of Immunobiology and Epigenetics

Citation: The beginning is the end: How promoters predefine where genes end (2023, May 12) retrieved 26 June 2024 from <https://phys.org/news/2023-05-predefine-genes.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.