

VIEWPOINT

Length matters: Disease implications for long genes

BY MARK ZYLKA

22 OCTOBER 2013





Gene class: Autism genes are, on average, four times larger than all other genes expressed in neurons.

In December 2011, my colleagues and I reported that topoisomerase inhibitors can unsilence **UBE3A**, a gene associated with autism and a related disorder called **Angelman syndrome**¹.

Topoisomerases are enzymes that untangle DNA during cell division and when DNA is transcribed into RNA. Precisely how topoisomerase inhibitors affect gene expression in neurons was unknown, however.

While studying this question, we made **two unexpected observations** that have implications for autism².

First, we found that topoisomerase inhibitors lower the expression of nearly all genes that are long (greater than 200 kilobases). This effect is not specific to neurons, as topoisomerase inhibitors similarly decrease the expression of long genes in cancer cell lines^{2,3}.

Second, we found that a large number of these long genes whose expression is lowered by the drugs are associated with **synapses**, the junctions between neurons, and neuronal development; 46 of them are also associated with autism.

For example, neurexin 1 (**NRXN1**) and contactin associated protein-like 2 (**CNTNAP2**), two genes

that are strongly linked to autism, are 1,114 kilobases and 2,305 kilobases in length, respectively. This makes them 18- to 38-fold longer than the average expressed gene, which is about 60 kilobases.

What's more, we noticed that autism genes are approximately four times longer (on average) than all other genes expressed in neurons, making autism genes longer as a class. To our knowledge, this gene-length relationship was never previously described for autism or any other neurodevelopmental disorder.

These gene-length-dependent findings raise an intriguing question. Namely, what are the biological and disease implications for being a long gene? Here, we outline a few of our ideas about long genes, and welcome **your thoughts on this topic**.

Size matters:

As we found, topoisomerase inhibitors impair transcription elongation — the process by which the transcription machinery travels along DNA — in long genes, but not short genes.

Researchers previously assumed that transcriptional mechanisms are indifferent to gene size, and so mechanisms that apply to average genes could be extended to longer genes. Our studies suggest that this assumption is flawed and that gene size has a significant influence on transcription.

How the transcriptional machinery senses gene size is an intriguing topic for further study. One clue could come from a recent study showing that large (102 kilobase) regions of the genome form structures that can be dynamically remodeled by RNA polymerase, an enzyme that produces RNA from DNA, and topoisomerases⁴. Genomic architecture may affect transcription of long genes and short genes differently, depending on where genes are located within these structures.

Random mutations are more likely to hit longer genes, simply because longer genes are bigger targets.

Bigger genes are also bigger targets. If you covered your eyes and threw darts in random directions, you would have a greater chance of hitting a really big dartboard than an average-sized one.

Our genomes are regularly subject to random mutations (deletions, duplications and spontaneous events). For example, Kári Stefánsson's group at **deCODE Genetics** in Iceland found that males accrue about two spontaneous, or *de novo*, mutations per year in reproductive, or germ, cells and that these mutations are passed on to the next generation⁵.

In keeping with the dartboard analogy, these random mutations are more likely to hit longer genes, simply because longer genes are bigger targets. Though not all genomic sites are equal in their rate of mutation⁶, it remains likely that longer genes are more likely to be hit than shorter ones.

What's more, longer genes may be more sensitive to deficits in the mechanisms that splice together the parts of a gene that code for protein. Topoisomerase inhibitors are known to affect splicing, and these effects are more pronounced in long genes³.

Studies in the past two years have found that reductions in the levels of two genes (TDP-43 and FUS/TLS) mutated in patients with amyotrophic lateral sclerosis lead to splicing errors in transcripts with exceptionally long introns, the non-protein-coding regions within genes^{7,8}.

Gene transcription also takes time, and quite a lot of time if the gene is extremely long.

Several labs have measured the transcription rate, which turns out to be around 3.8 kilobases per minute. This drops to 1.0 kilobase per minute in the presence of a topoisomerase inhibitor^{9,10}. For an average gene, it takes about 12 minutes to make one transcript.

But for a really long gene such as CNTNAP2, it takes about 10.4 hours to make one transcript. That's nearly half a day. And that's in a healthy cell. If topoisomerases are impaired in some way, it can take as long as 38 hours to make one transcript. A lot can happen to a neuron in 10 hours, and even more can happen in 38 hours.

If something impairs transcription, even transiently, it could disproportionately affect proteins that originate from long transcripts. It isn't hard to imagine how a drug that transiently affects topoisomerases might affect multiple long genes associated with synaptic function and affect critical periods of brain development.

Avoiding collisions:

DNA polymerases copy a cell's genome during cell division, whereas RNA polymerases make gene transcripts when a gene is expressed. But both of these enzymes travel along the same 'road.' And this has consequences — particularly if a gene is long.

A dividing cell needs to replicate its entire genome, a process that requires DNA polymerase and a lot of time. For example, cell division typically takes about 24 hours. For long genes, the transcribing RNA polymerase has a high probability of colliding with the DNA polymerase over this time frame, particularly if the long gene is actively being transcribed.

Two years ago, researchers studied these collision events in several extremely long genes, including CNTNAP2¹¹. The title of that article — 'Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes' — is revealing.

Shorter genes are much less sensitive to these collision events than long ones are.

Like a head-on car crash, these collisions can be catastrophic. They cause DNA damage in dividing cells, generate mutations and can lead to cell death. Germ cells, such as those in the testes, also divide. And the testes express more genes in common with brain than with any other tissue¹².

It is not hard to imagine how long brain-related transcripts, when expressed in dividing germ cells, might be more susceptible to such collision events. These events could bias destabilizing mutations to these longer genes. In turn, the mutations could be transmitted to children and have the potential to affect the developing brain.

Intriguingly, genes involved in brain function, autism and other diseases have a higher probability of harboring mutations when compared with other genes¹³. Notably, many of these mutation hotspots are located within extremely long genes. (For example, **AUTS2** is 1,193 kilobases; **CSMD1** is 2,059 kilobases; **GABRB3** is 230 kilobases; **GPR98** is 605 kilobases; **KCNMA1** is 753 kilobases; **KIRREL3** is 577 kilobases; **LOC10061653** is 541 kilobases, **NRXN1** is 1,114 kilobases and **SHANK2** is 622 kilobases¹⁴.) It will be interesting to determine whether these collision events contribute to mutations in autism genes, which, as we found, are longer as a class.

By contrast, neurons do not divide — so these collisions are unlikely to occur in neurons. As a result, long genes can be transcribed in neurons with fewer negative consequences. This makes it possible for neurons to use long genes for important things, such as making synapses and wiring the brain.

Topoisomerases, and in particular **TOP1**, help prevent these collisions from occurring, and in doing so, help prevent mutations caused by these collisions¹⁵. By contrast, topoisomerase inhibitors increase genomic instability.

And as we found, topoisomerase inhibitors, even at low doses, profoundly lower the expression of numerous long synaptic genes¹⁶. Thus, drugs that act like topoisomerase inhibitors, and mutations in topoisomerases, could profoundly affect synaptic function and contribute to neurodevelopmental disorders.

In fact, there is growing evidence that mutations in topoisomerases, including **TOP1** and **TOP3beta**, **are associated with autism** and intellectual disability¹⁷⁻²⁰.

In conclusion, the biology of extremely long genes appears to be different from that of average-sized genes, and this can have consequences for diseases such as autism, in which numerous long genes are implicated.

Mark Zylka is associate professor of cell biology and physiology at the University of North Carolina,

Chapel Hill.

References:

1. Huang H.S. *et al. Nature* **481**, 185-189 (2011) [PubMed](#)
2. King I.F. *et al. Nature* **501**, 58-62 (2013) [PubMed](#)
3. Solier S. *et al. Cancer Res.* **73**, 4830-4839 (2013) [PubMed](#)
4. Naughton C. *et al. Nat. Struct. Mol. Biol.* **20**, 387-395 (2013) [PubMed](#)
5. Kong A. *et al. Nature* **488**, 471-475 (2012) [PubMed](#)
6. Michaelson J.J. *et al. Cell* **151**, 1431-1442 (2012) [PubMed](#)
7. Lagier-Tourenne C. *et al. Nat. Neurosci.* **15**, 1488-1497 (2012) [PubMed](#)
8. Polymenidou M. *et al. Nat. Neurosci.* **14**, 459-468 (2011) [PubMed](#)
9. Singh J. and R.A. Padgett *Nat. Struct. Mol. Biol.* **16**, 1128-1133 (2009) [PubMed](#)
10. Darzacq X. *et al. Nat. Struct. Mol. Biol.* **14**, 796-806 (2007) [PubMed](#)
11. Helmrich A. *et al. Mol. Cell* **44**, 966-977 (2011) [PubMed](#)
12. Guo J.H. *et al. Cytogenet. Genome Res.* **111**, 107-109 (2005) [PubMed](#)
13. Michaelson J.J. *et al. Cell* **151**, 1431-1442 (2012) [PubMed](#)
14. Zylka M. Unpublished observations.
15. Tuduri S. *et al. Nat. Cell Biol.* **11**, 1315-1324 (2009) [PubMed](#)
16. King I.F. *et al. Nature* **501**, 58-62 (2013) [PubMed](#)
17. Neale B.M. *et al. Nature* **485**, 242-245 (2012) [PubMed](#)
18. Iossifov I. *et al. Neuron* **74**, 285-299 (2012) [PubMed](#)
19. Stoll G. *et al. Nat. Neurosci.* **16**, 1228-12237 (2013) [PubMed](#)

20. Xu D. *et al. Nat. Neurosci.* **16**, 1238-1247 (2013) [PubMed](#)