HHS Public Access

Author manuscript

Neuron. Author manuscript; available in PMC 2016 April 22.

Published in final edited form as:

Neuron. 2015 April 22; 86(2): 353–355. doi:10.1016/j.neuron.2015.03.059.

Gene length matters in neurons

Mark J. Zylka[*], **Jeremy M. Simon**, and **Benjamin D. Philpot**

Department of Cell Biology and Physiology, UNC Neuroscience Center, The Carolina Institute for Developmental Disabilities, The University of North Carolina at Chapel Hill

Abstract

A recent study by Gabel et al. (2015) found that *Mecp2*, the gene mutated in Rett syndrome, represses long (>100 kb) genes associated with neuronal physiology and connectivity by binding to methylated CA sites in DNA. This study adds to a growing body of literature implicating gene length and transcriptional mechanisms in neurodevelopmental and neurodegenerative disorders.

Mutations in *Methyl-CpG-binding protein 2* (*MECP2*) cause Rett syndrome, an X-linked neurodevelopmental disorder associated with synaptic dysfunction, intellectual disability, and autism symptoms (Chahrour and Zoghbi, 2007). While MECP2 is known to function as a transcriptional repressor, precisely how the loss of MECP2 impairs synaptic function at the level of gene expression has remained unclear. This question has taken on greater relevance now that *de novo* mutations in numerous transcriptional regulators were identified in individuals with autism (De Rubeis et al., 2014; Iossifov et al., 2014), a disorder that is also associated with dysfunctional synapses (Chen et al., 2015; Delorme et al., 2013).

New insights into this question were inspired by recent findings on topoisomerases—a class of enzymes that relieve supercoiling during DNA replication and transcription, and that are implicated in autism, intellectual disability, schizophrenia, and neurodegeneration (Katyal et al., 2014; King et al., 2013; Neale et al., 2012; Stoll et al., 2013). In particular, topoisomerase inhibitors were recently shown to reduce the expression of extremely long (>100 kb) genes in cortical neurons by impairing transcription elongation (King et al., 2013). Many of these long genes were associated with neuronal development and synapses, including several autism candidate genes like Neurexin-1, Neurexin-3, and Neuroligin-1 (King et al., 2013). These findings suggested that a length-dependent impairment of gene transcription, particularly in neural tissues and neurons where longer genes are overrepresented (Figure 1, 2) (Gabel et al., 2015), might impair synaptic function and contribute to neurodevelopmental disorders. In support of this possibility, topoisomerase inhibition reversibly depletes Neurexin-1 and Neuroligin-1 at the protein level and reversibly impairs synaptic function (Mabb et al., 2014).

With these new perspectives on how transcriptional deficits can affect long genes in neurons, the labs of Michael Greenberg and Sacha Nelson, working independently, examined gene expression data from *Mecp2* null mice (Gabel et al., 2015; Sugino et al., 2014). Remarkably, they found that long genes were modestly but reproducibly upregulated,

[*]Corresponding author (zylka@med.unc.edu).

with an inflection point around 100 kb. Gabel and colleagues performed a number of genome-wide studies and re-analyzed numerous existing data sets to convincingly demonstrate a direct relationship between MECP2 function, gene length, and disease severity. This included examining gene expression data from *Mecp2* knockout mice, two mouse lines with Rett syndrome nonsense mutations (R270X and G273X), a mouse line with a Rett syndrome missense mutation (R306C), cultured human neurons derived from embryonic stem cells lacking *Mecp2*, and brain tissue from individuals with Rett syndrome. They also found that long gene transcription was repressed in a mouse model of *Mecp2*-duplication syndrome.

These studies provide strong evidence that the transcriptional repressor function of MECP2 is biased towards longer transcripts. Gabel and colleagues also found that *Fmrp*, the gene mutated in Fragile X Syndrome, targets genes that are longer than average (Gabel et al., 2015). Collectively, these studies extend the concept that transcriptional deficits can affect gene expression in a length-dependent manner, and may represent a common mechanism associated with neurodevelopmental disorders.

Both groups also found that many of the upregulated long genes in *Mecp2*-deficient neurons and brain tissues are involved in cell adhesion, axon guidance, and synapse formation (Gabel et al., 2015; Sugino et al., 2014). It is now well-recognized that changes in synapse number can affect excitatory and inhibitory balance in the brain (Gogolla et al., 2009). As such, a small increase in expression across a large number of long synaptic genes could upset excitatory/inhibitory balance and impair synapse function, as is seen in Rett model mice (Chahrour and Zoghbi, 2007). These studies thus provide new insights into how *Mecp2* loss could transcriptionally impair synaptic function.

Gabel and colleagues also found that mouse and human brains express a greater proportion of long genes relative to non-neural tissues. We found this holds true when a larger number of brain regions, tissues, and cell types are examined (Figure 1). The transcriptomes of frontal cortex and amygdala, two brain regions implicated in autism and other neurodevelopmental disorders (Chen et al., 2015), are particularly biased for longer transcripts relative to other brain regions.

To better understand which brain cell types contribute to this length bias, we re-analyzed the transcriptomes of the nine major cell types that make up the mouse cerebral cortex and CA1 region of the hippocampus, as defined in a recent single cell RNA-seq study (Zeisel et al., 2015). Intriguingly, we found that the transcriptome of each major neuron subtype is biased for longer transcripts relative to non-neuronal cell types (Figure 2). Moreover, the transcriptome of hippocampal pyramidal neurons is significantly longer than that of somatosensory cortex pyramidal neurons (Figure 2). This mirrors the length bias seen at the whole tissue level (Figure 1). Thus, transcriptome length biases can be used to distinguish brain regions and individual neuron subtypes from one another as well as from non-neuronal cells. These findings raise the possibility that brain regions and neuron subtypes with longer transcriptomes may be more sensitive to transcriptional deficits, such as those caused by *Mecp2* deficiency or overexpression.

Given how these transcriptional deficits preferentially affect genes that are 100 kb or larger, this raises the question of how the transcriptional machinery “senses” gene size. Or more accurately, what molecularly distinguishes an extremely long gene from an average length gene? Gabel and colleagues provide important new insight into this question by showing that the density of methylated CA (mCA) sites is higher over long gene bodies when compared to short gene bodies (Gabel et al., 2015). Their data provide compelling evidence that MECP2 binds these mCA sites, effecting greater repression as the density of mCA goes up with gene length. One way they supported this model was by examining gene expression in mice deficient in *Dmnt3a*, the enzyme that generates the methylation mark. *Dmnt3a* loss led to upregulation of long genes, thus phenocopying *Mecp2* deletion. Their data therefore provide compelling evidence that epigenetic marks on the gene body are recognized by MECP2 and promote gene-length-dependent transcriptional repression by MECP2.

Lastly, Gabel and colleagues ended their study with a therapeutic twist. Knowing that the topoisomerase inhibitor topotecan dose-dependently down-regulates long genes (King et al., 2013), they showed it was possible to normalize expression of eight MECP2-repressed genes with low concentrations of topotecan. Ultimately, the success of such a strategy for rescuing Rett syndrome phenotypes will depend on how many long genes are shared between the two conditions, and whether these shared long genes drive the phenotypes under study. Regardless, this proof-of-concept demonstrates that a pharmacological normalization of multiple MECP2-relevant genes is possible.

Taken together, these new studies provide key insights into how mutations in transcriptional regulators could impair synaptic physiology and brain development by altering the expression of numerous long genes. Loss of genes linked to amyotrophic lateral sclerosis (TDP-43 and FUS/TLS) preferentially affects splicing of transcripts with long (>100 kb) first introns (Lagier-Tourenne et al., 2012), suggesting that gene length is a factor in neurodegenerative disorders as well. Thus, it is becoming clear that several transcriptional mechanisms associated with gene length (transcription elongation, transcriptional repression, splicing, and epigenetic mark deposition) represent major molecular points of vulnerability for neurons as well as targets for therapeutic intervention, particularly in the context of neurodevelopmental and neurodegenerative disorders.

Acknowledgments

This work was supported by the National Institute of Mental Health (R01MH093372; M.J.Z., B.D.P.), NINDS and NICHD (P30NS045892, P30HD03110, J.M.S.), and a Pioneer Award from The National Institutes of Health (DP1ES024088; M.J.Z.).

References

Chahrour M, Zoghbi HY. The story of Rett syndrome: from clinic to neurobiology. Neuron. 2007; 56:422–437. [PubMed: 17988628]

Chen JA, Penagarikano O, Belgard TG, Swarup V, Geschwind DH. The emerging picture of autism spectrum disorder: genetics and pathology. Annual review of pathology. 2015; 10:111–144.

De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014; 515:209–215. [PubMed: 25363760]

Delorme R, Ey E, Toro R, Leboyer M, Gillberg C, Bourgeron T. Progress toward treatments for synaptic defects in autism. Nat Med. 2013; 19:685–694. [PubMed: 23744158]

Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, Hemberg M, Ebert DH, Greenberg ME. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature. 2015

Gogolla N, Leblanc JJ, Quast KB, Sudhof TC, Fagiolini M, Hensch TK. Common circuit defect of excitatory-inhibitory balance in mouse models of autism. Journal of neurodevelopmental disorders. 2009; 1:172–181. [PubMed: 20664807]

Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014; 515:216–221. [PubMed: 25363768]

Katyal S, Lee Y, Nitiss KC, Downing SM, Li Y, Shimada M, Zhao J, Russell HR, Petrini JH, Nitiss JL, et al. Aberrant topoisomerase-1 DNA lesions are pathogenic in neurodegenerative genome instability syndromes. Nat Neurosci. 2014; 17:813–821. [PubMed: 24793032]

King IF, Yandava CN, Mabb AM, Hsiao JS, Huang HS, Pearson BL, Calabrese JM, Starmer J, Parker JS, Magnuson T, et al. Topoisomerases facilitate transcription of long genes linked to autism. Nature. 2013; 501:58–62. [PubMed: 23995680]

Lagier-Tourenne C, Polymenidou M, Hutt KR, Vu AQ, Baughn M, Huelga SC, Clutario KM, Ling SC, Liang TY, Mazur C, et al. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. Nat Neurosci. 2012; 15:1488–1497. [PubMed: 23023293]

Mabb AM, Kullmann PH, Twomey MA, Miriyala J, Philpot BD, Zylka MJ. Topoisomerase 1 inhibition reversibly impairs synaptic function. Proc Natl Acad Sci U S A. 2014; 111:17290–17295. [PubMed: 25404338]

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012; 485:242–245. [PubMed: 22495311]

Stoll G, Pietilainen OP, Linder B, Suvisaari J, Brosi C, Hennah W, Leppa V, Torniainen M, Ripatti S, Ala-Mello S, et al. Deletion of TOP3beta, a component of FMRP-containing mRNPs, contributes to neurodevelopmental disorders. Nat Neurosci. 2013; 16:1228–1237. [PubMed: 23912948]

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A. 2004; 101:6062–6067. [PubMed: 15075390]

Sugino K, Hempel CM, Okaty BW, Arnson HA, Kato S, Dani VS, Nelson SB. Cell-type-specific repression by methyl-CpG-binding protein 2 is biased toward long genes. J Neurosci. 2014; 34:12877–12883. [PubMed: 25232122]

Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A, Marques S, Munguba H, He L, Betsholtz C, et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science. 2015; 347:1138–1142. [PubMed: 25700174]

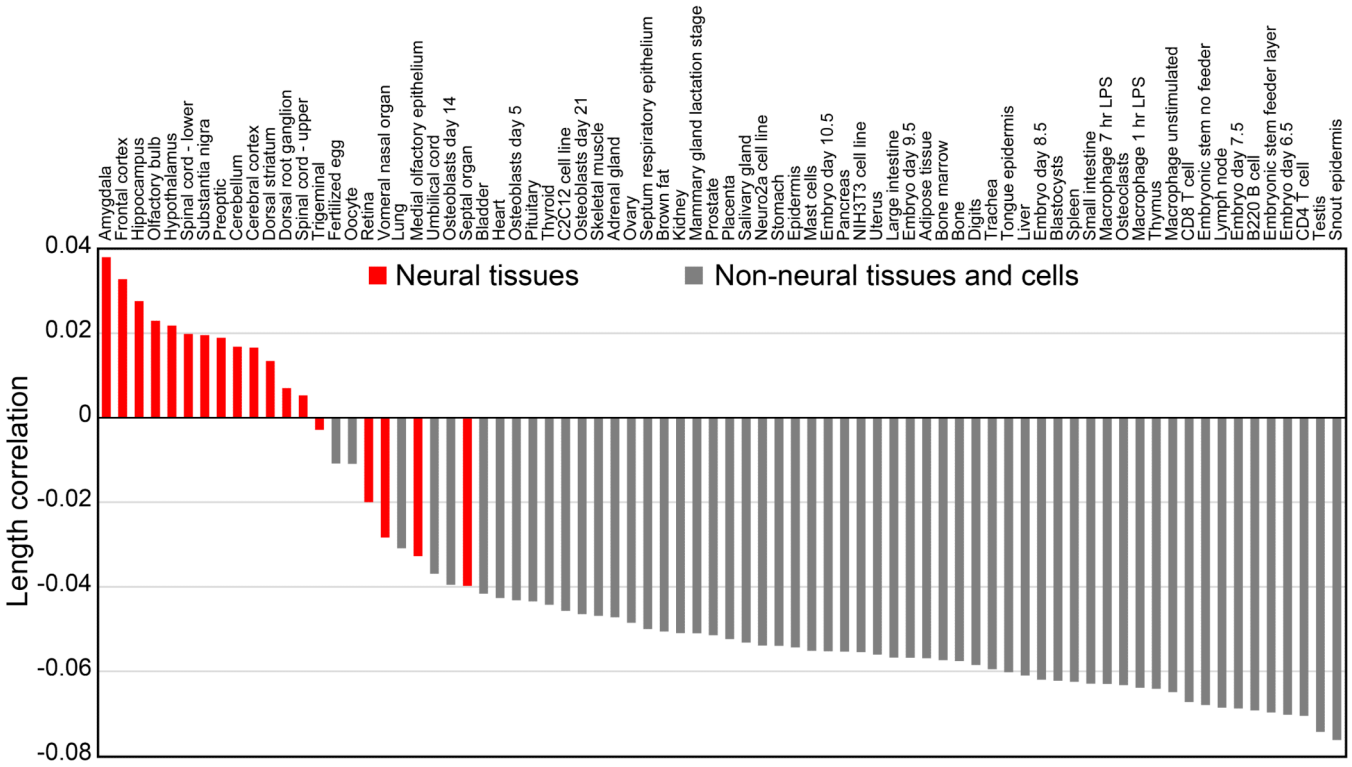

Figure 1. The transcriptome of neural tissues is biased for longer transcripts relative to non-neural tissues and cell lines

Mouse expression data were obtained from BioGPS (Su et al., 2004). Expression intensity for each gene in a given tissue was compared to gene length and Pearson correlation coefficients were computed. Positive values indicate a positive correlation between gene expression and gene length.

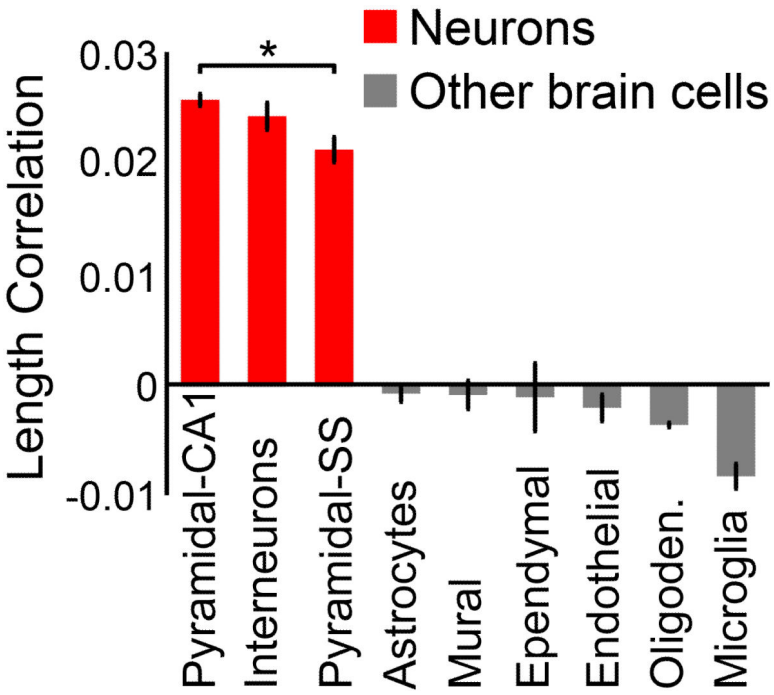

Figure 2. The transcriptome of neurons is biased for longer transcripts relative to other brain cell types

Oligoden. = oligodendrocytes. Hippocampal CA1 versus somatosensory (SS) pyramidal neurons, *p < 0.005 by two-sample t-test. Error bars are standard error of the mean. Gene expression data were obtained from http://linnarssonlab.org/blobs/cortex/expression_mRNA_17-Aug-2014.txt. Total read counts were computed for each cell by taking the sum of all counts for all RefSeq genes, and individual gene values were normalized by this sum. Genes shorter than 500 bp in length were excluded from the analysis. Gene length Pearson correlations for each of the 3,005 individual cells were then computed. Values were averaged and plotted for each of the nine cell classes as defined in Zeisel et al. (2015).