(Macaulay et al., 2017) may now be applied to embryos and could lead to the detection of MdnCNVs in the genome sequences of blastomeres prior to embryonic transcriptional activation. In parallel, the RNA-seq data of those same blastomeres can be scrutinized for the smoking gun, e.g., mutant maternal transcripts or aberrant levels of transcripts, perhaps corresponding to genes affecting DNA damage and repair pathways. It is also possible that mature oocytes or even individual blastomeres will show variations in their transcript profiles, causing some cells to be more prone to CIN than others.

Liu et al. (2017) have identified a MdnCNV mutator phenotype. Characterizing the mechanisms that underpin this mutator phenotype—and CIN, in general—will likely advance our understanding of the origin of genomic disorders, as well as the contribution of similar processes to diseases such as cancer. Further study may also provide insight into genome evolution, as MdnCNV and CIN could provide a means for rapid adaptation of species to environmental changes.

ACKNOWLEDGMENTS

T.V. and J.R.V. are supported by KU Leuven (SymBioSys, PFV/10/016), FWO (G.0924.15) and EU (Horizon 2020 WIDENLIFE: 692065). T.V. is also supported by the Wellcome Trust and MRC. T.V. and J.R.V. are co-inventors on patent applications PCT/EP2011/060211-WO/2011/157846, "Methods for haplotyping single cells;" PCT/EP2013/070858-WO/2014/053664, "High-throughput genotyping by sequencing low amounts of genetic material;" and PCT/EP2014/068315-WO/2015/028576, "Haplotyping and copy number typing using polymorphic variant allelic frequencies."

REFERENCES

Carvalho, C.M., and Lupski, J.R. (2016). Nat. Rev. Genet. *17*, 224–238.

Gawad, C., Koh, W., and Quake, S.R. (2016). Nat. Rev. Genet. *17*, 175–188.

Greco, E., Minasi, M.G., and Fiorentino, F. (2015). N. Engl. J. Med. *373*, 2089–2090.

Itsara, A., Wu, H., Smith, J.D., Nickerson, D.A., Romieu, I., London, S.J., and Eichler, E.E. (2010). Genome Res. *20*, 1469–1481.

Liu, P., Yuan, B., Carvalho, C.M.B., Wuster, A., Walter, K., Zhang, L., Gambin, T., Chong, Z., Campbell, I.M., Coban, Z., et al. (2017). Cell *168*, this issue, 830–842.

Macaulay, I.C., Ponting, C.P., and Voet, T. (2017). Trends Genet. *33*, 155–168.

Robberecht, C., Voet, T., Zamani Esteki, M., Nowakowska, B.A., and Vermeesch, J.R. (2013). Genome Res. *23*, 411–418.

Vanneste, E., Voet, T., Le Caignec, C., Ampe, M., Konings, P., Melotte, C., Debrock, S., Amyere, M., Vikkula, M., Schuit, F., et al. (2009). Nat. Med. *15*, 577–583.

Vermeesch, J.R., Voet, T., and Devriendt, K. (2016). Nat. Rev. Genet. *17*, 643–656.

Zarrei, M., MacDonald, J.R., Merico, D., and Scherer, S.W. (2015). Nat. Rev. Genet. *16*, 172–183.

Linking Protein and RNA Function within the Same Gene

Anthony Szempruch[1] and Mitchell Guttman[1,*]
[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA
*Correspondence: mguttman@caltech.edu
http://dx.doi.org/10.1016/j.cell.2017.02.014

Exposure to ultraviolet light leads to a cell-wide DNA damage response that includes a global reduction in transcription. Williamson et al., identify a protein involved in this process as well as a noncoding RNA produced by alternative processing of RNA transcribed from the same gene that promotes recovery from the repressed state.

Over the past decade it has become clear that mammalian genomes contain thousands of genes that do not code for proteins, but instead produce functional long non-coding RNAs (lncRNAs) (Quinn and Chang, 2016). Recently, there has been great excitement about the discovery of several genes that were initially thought to encode lncRNAs, but are actually translated and produce short peptides that are required for the function of these genes (Kondo et al., 2010; Pauli et al., 2014). In this issue of *Cell*, Williamson et al. demonstrate that this can also work the other way—where canonical protein-coding genes that encode functional protein products can also encode functional lncRNAs (Williamson et al., 2017).

By studying transcriptional recovery following UV damage, the authors find that there is a global switch in pre-mRNA processing resulting in a preference for production of transcripts containing alternative last exons that are not normally included in the dominant mRNA isoform (Figure 1). Of the 84 genes that were alternatively processed in this way, the ASCC3 gene stood out—it was the strongest factor identified as a regulator of transcription following UV damage in a previous genome-wide functional screen (Boeing et al., 2016). The protein encoded by ASCC3 acts to repress transcription following UV damage. Intriguingly, disruption of the UV-specific (or "short") isoform RNA that contains the alternative last

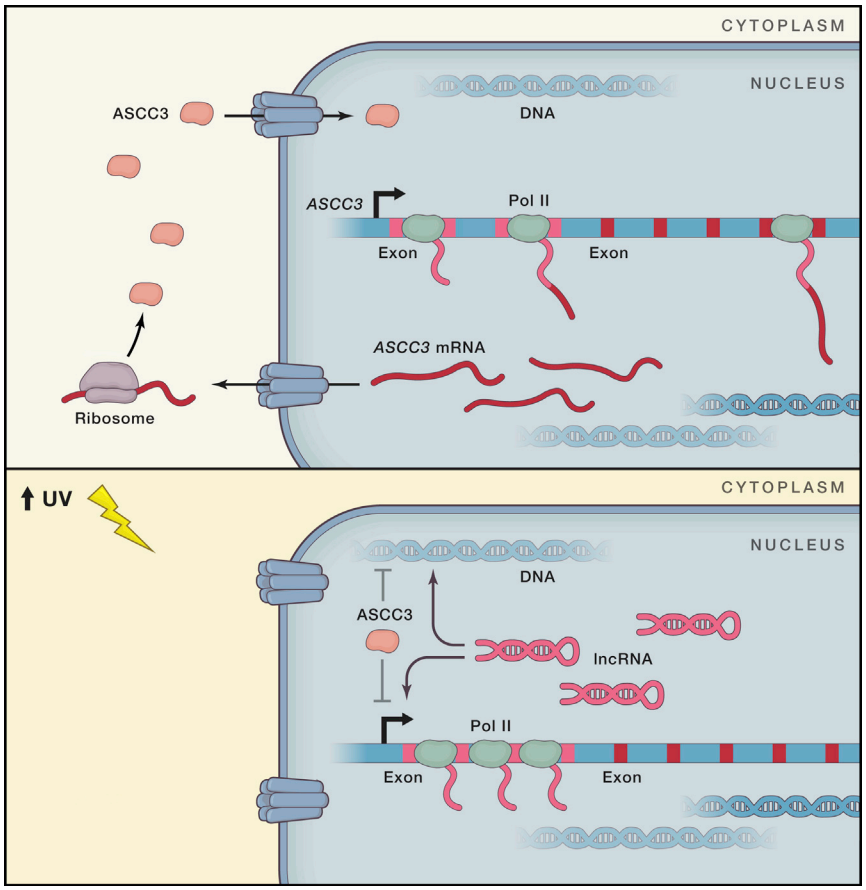

Figure 1. UV Damage Induces Alternative Last Exon Inclusion Resulting in a Functional lncRNA with Antagonistic Properties to the Protein Produced from the Same Gene

The full length ASCC3 transcript functions to repress transcription after UV damage. However, UV treatment results in the upregulation of RNA transcripts containing alternative last exons. In the case of ASCC3, its short RNA isoform is retained within the nucleus, where it functions as a lncRNA to activate transcription allowing for transcriptional recovery after UV damage.

exon had the opposite functional effect, enabling transcriptional recovery following UV damage.

Given this distinct role for the ~3,100-nucleotide-long RNA produced from the shorter isoform, the authors studied this RNA and its functional role in greater detail. While ectopic expression of the short isoform rescued the UV-dependent defect in transcriptional recovery, expression of the putative coding region contained within the short isoform did not, suggesting that the protein encoded by this mRNA may not be responsible for its function. Consistent with this interpretation, the RNA produced from the short isoform localized almost exclusively in the nucleus. In an elegant experiment to test more directly the rescue phenotype, the authors integrated a premature stop codon immediately downstream of the start codon that completely ablated protein production from the RNA. Despite producing no protein, there was complete rescue of the transcriptional recovery phenotype, demonstrating that the functional role of the short RNA isoform is mediated by a non-coding RNA encoded in the ASCC3 protein-coding gene.

This study reveals an unexpected source of functional lncRNAs that can be produced from within canonical protein-coding genes, which may increase the total number of possible functional lncRNAs encoded in mammalian genomes. The ASCC3 example is likely not the only case of a gene producing two distinct functional molecules—a protein and a lncRNA. Other genes examined by Svejstrup and co-workers produce unique isoforms that are formed by inclusion of alternative last exons (Williamson et al., 2017). Although there is no evidence that these other examples also work as lncRNAs, these isoforms generally contain significantly shorter open reading frames making them attractive candidates for further exploration of such roles. This dual functionality may also play important roles in other processes beyond the DNA damage response. In fact, alternative processing of mRNAs appears to be the rule rather than the exception in mammalian genomes (Kornblihtt et al., 2013; Wang et al., 2008). This raises the question of how many of the ~25,000 protein-coding mRNA genes yield isoforms that may function in some biological context as a lncRNA.

Given that most traditional functional screens generally decrease transcript levels, this study raises the intriguing question of how to properly define whether a given function ascribed to a gene is in fact mediated by an encoded protein product or through a functional RNA molecule. Currently, lncRNAs are defined by the *absence* of protein-coding potential within a gene. The ASCC3 example highlights the limitations of such a definition because even genes that may have coding potential (i.e., they can make proteins in some isoforms) can also generate isoforms in which the RNA functions as a non-coding RNA. One way to separate protein coding from lncRNA function within a gene is to use genetic tools to introduce frameshift mutations or premature stop codons. More generally, it may be necessary to develop new criteria that can detect the presence of positive signals of lncRNA function (rather than absence of coding potential) to more accurately classify such genes. These classifiers may include the presence of RNA structures, features of lncRNA-protein interactions, or increased levels of evolutionary selection beyond those defined by coding potential alone. However, such analytical criteria may prove challenging given our current state of knowledge about lncRNAs. Alternatively, genome-wide functional screens may provide a rapid approach to identify such cases by utilizing perturbations that decrease RNA levels (Gilbert et al., 2014) and comparing their effects to those that disrupt the encoded protein (Shalem et al., 2014).

The key question that remains unaddressed is why encode both the protein and lncRNA from within the same gene. Although the answer to this question is still unknown, there are several potential regulatory advantages that such a system might enable. Specifically, a lncRNA encoded from within a protein-coding gene may have an intrinsic regulatory advantage for controlling the function of the protein encoded from that same gene because lncRNAs have the unique capability of acting in close spatial proximity to their transcription loci (Engreitz et al., 2016). By being encoded from the same DNA as the protein, a lncRNA may be able to localize preferentially to its chromatin to activate or repress transcription or may act to generate a high local concentration to compete with protein binding at that locus.

These findings now add an extra layer of complexity to the study of lncRNAs. Instead of genes neatly falling into protein coding or non-coding categories, this report argues that we need to think carefully about whether a function ascribed to a gene may be mediated by a protein or RNA, regardless of the apparent coding potential of the gene. This study also raises the important questions of how pervasive such genes with dual functionality may be, what mechanisms such genes may contribute in genome regulation, and what regulatory and/or evolutionary advantage is provided by encoding two distinct molecules within the same transcriptional unit.

REFERENCES

Boeing, S., Williamson, L., Encheva, V., Gori, I., Saunders, R.E., Instrell, R., Aygün, O., Rodriguez-Martinez, M., Weems, J.C., Kelly, G.P., et al. (2016). Cell Rep. *15*, 1597–1610.

Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016). Nat. Rev. Mol. Cell Biol. *17*, 756–770.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Cell *159*, 647–661.

Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., Kobayashi, S., Payre, F., and Kageyama, Y. (2010). Science *329*, 336–339.

Kornblihtt, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M.J. (2013). Nat. Rev. Mol. Cell Biol. *14*, 153–165.

Pauli, A., Norris, M.L., Valen, E., Chew, G.L., Gagnon, J.A., Zimmerman, S., Mitchell, A., Ma, J., Dubrulle, J., Reyon, D., et al. (2014). Science *343*, 1248636.

Quinn, J.J., and Chang, H.Y. (2016). Nat. Rev. Genet. *17*, 47–62.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Science *343*, 84–87.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Nature *456*, 470–476.

Williamson, L., Saponaro, M., Boeing, S., East, P., Mitter, R., Kantidakis, T., Kelly, G.P., Lobley, A., Walker, J., Spencer-Dene, B., et al. (2017). Cell *168*, this issue, 843–855.

What Came First—the Virus or the Egg?

Robert W. Doms[1,*]

[1]The Children's Hospital of Philadelphia and the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA

*Correspondence: domsr@email.chop.edu

http://dx.doi.org/10.1016/j.cell.2017.02.012

Three recent studies find that the single-pass transmembrane protein HAP2 mediates gamete fusion and is remarkably similar to class II fusion proteins found in viruses such as dengue and Zika.

Biological membranes are inherently stable structures that segregate the cytoplasm from the extracellular environment and, in eukaryotes, organize content into membrane-delimited organelles. To share content, membranes must undergo carefully regulated membrane fusion events that are invariably mediated by highly specialized fusion proteins. Vesicular transport, placental syncytiotrophoblast formation, and gamete fusion are examples in which opposing cellular membranes must fuse with specificity and efficiency, overcoming the barriers that normally prevent lipid mixing and fusion pore formation. Mechanistically, much of what we know about how proteins elicit membrane fusion comes from the study of proteins that mediate entry of membrane-enveloped viruses into cells. Despite the large number and diversity of viruses, all viral fusion proteins identified to date fall into just three different structural classes, with some having clear parallels to cellular membrane fusion proteins. For example, placental cytotrophoblasts utilize a class I fusion protein derived from an endogenous retrovirus to form syncytiotrophoblasts (Mi et al., 2000), and SNARE proteins bear general similarities to class I viral fusion proteins as well (likely an example of convergent evolution), forming a bundle of alpha helices that bring opposing membranes into close proximity (Südhof and Rothman, 2009). Left unanswered, however, is how male and female gametes fuse with each other—a rather important event! Recent studies from the Clark, Grishin, Podbilewicz, Rey, and Snell laboratories (Fédry et al., 2017 [this issue of *Cell*]; Pinello et al., 2017; Valansi et al., 2017), using