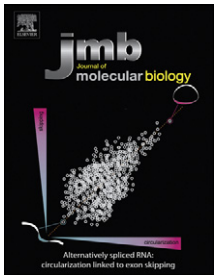


Exon Skipping Is Correlated with Exon Circularization

Steven Kelly [1], Chris Greenman [2], Peter R. Cook [3] and Argyris Papantonis [3,4]

*1 - **Department of Plant Sciences,** University of Oxford, South Parks Road, OX1 3RB Oxford, UK*
*2 - **The Genome Analysis Centre,** Norwich Research Park, NR4 7UH Norwich, UK*
*3 - **Sir William Dunn of Pathology,** University of Oxford, South Parks Road, OX1 3RE Oxford, UK*
*4 - **Center for Molecular Medicine,** University of Cologne, Robert-Koch-Straße 21, D-50931 Cologne, Germany*

***Correspondence to Steven Kelly and Argyris Papantonis:** A. Papantonis is to be contacted at: Center for Molecular Medicine, University of Cologne, Robert-Koch-Straße 21, D-50931 Cologne, Germany. steven.kelly@plants.ox.ac.uk; argyris.papantonis@uni-koeln.de*
http://dx.doi.org/10.1016/j.jmb.2015.02.018
Edited by A. Pyle

Abstract

Circular RNAs are found in a wide range of organisms and it has been proposed that they perform disparate functions. However, how RNA circularization is connected to alternative splicing remains largely unexplored. Here, we stimulated primary human endothelial cells with tumor necrosis factor α or tumor growth factor β, purified RNA, generated >2.4 billion RNA-seq reads, and used a custom pipeline to characterize circular RNAs derived from coding exons. We find that circularization of exons is widespread and correlates with exon skipping, a feature that adds considerably to the regulatory complexity of the human transcriptome.

Circular RNAs (circRNAs) were initially discovered as by-products of viral and mitochondrial RNA metabolism, tRNA intron excision, and ribosomal RNA processing [1]. Recently, circRNAs derived from protein-coding and non-coding genes have been identified in various cell types from organisms spanning the tree of life [2–6]. In mammals, circRNAs arising from exons of protein-coding genes have been identified, and some act as regulatory "miRNA sponges" during development [4,5,7]. It is now also understood that circularization occurs cotranscriptionally and may interfere with splicing [8]. To determine if circularization of exonic RNA was linked to exon skipping and whether these circular products are temporally regulated, we sampled RNA in 30-min intervals for 2 h from human umbilical vein endothelial cells (HUVECs) treated with tumor necrosis factor α (TNFα) or tumor growth factor β. The RNA was sequenced in two ways to provide insight into the relationship between the abundance of an exon in mature mRNA [using poly(A)$^+$ RNA] and its rate of circularisation (using rRNA-depleted total RNA) via a custom bioinformatics pipeline (Fig. 1a; for details, see Supplementary Methods).

All expressed exons were quantified in HUVECs at all times post-stimulation with either cytokine, using poly(A)$^+$ RNA-seq data (~120 million read pairs/time point; Supplementary Fig. 1a) and exon sequences from the Consensus Coding Sequence database [9]. circRNAs were quantified using total RNA-seq (depleted of rRNA and sequenced to the same depth). Here, reads that mapped to any expressed exon and produced a 25- to 75-nt overhang were selected, and overhangs remapped to the genome in an effort to determine their source. Exon ends joined "head to tail" were classified as circRNAs (Fig. 1a) as in previous approaches [4]. Of the 163,326 actively transcribed exons, 38,058 (23%) were detected as circularized with one or more supporting reads. To minimize "noise" in the data, we focused on the top 9025 (5.5%) circular transcripts [also not present in poly(A)$^+$; Supplementary Fig. 1b] derived from single exons in 2606 (of ~8800) actively transcribed genes. With the use of these stringent criteria, each circle is supported by ≥11 independent reads (Supplementary Table 1). Exonic circRNAs arise roughly equally from all exons (except those connected to 5′ or 3′ untranslated regions), almost invariably through

splicing at canonical splice sites (Supplementary Fig. 1c and d). Notably, exons with flanking introns longer than the genome average are more frequently circularized (Supplementary Fig. 1e) as described for other cell types [2].

To determine if these exon-derived circRNAs were related to exon skipping, we compared the abundance of each exon in poly(A)$^+$ mRNA to the extent of its circularization (i.e., to the ratio of circular to conventional connections assigned to each exon in total, rRNA-depleted, RNA). This revealed that the more an exon is circularized, the less it is represented in processed mRNA (Fig. 1b). To exclude that reduced exon abundance was due to reduced expression of the parent gene, we plotted the relative retention of each exon in mature transcripts as a function of its circularization; this confirmed that the absence of an exon from mRNA, as reflected in its "skipping" by alternative splicing [10], is proportional to its circularization rate (Supplementary Fig. 2a). Interestingly, genes producing circRNAs were particularly enriched for such Gene Ontology terms as "transporter activity", "translation" and "mRNA metabolism", and "nucleoside/-tide binding" (Supplementary Fig. 2b). However, it was recently shown that circRNAs derived from adjacent exons are more

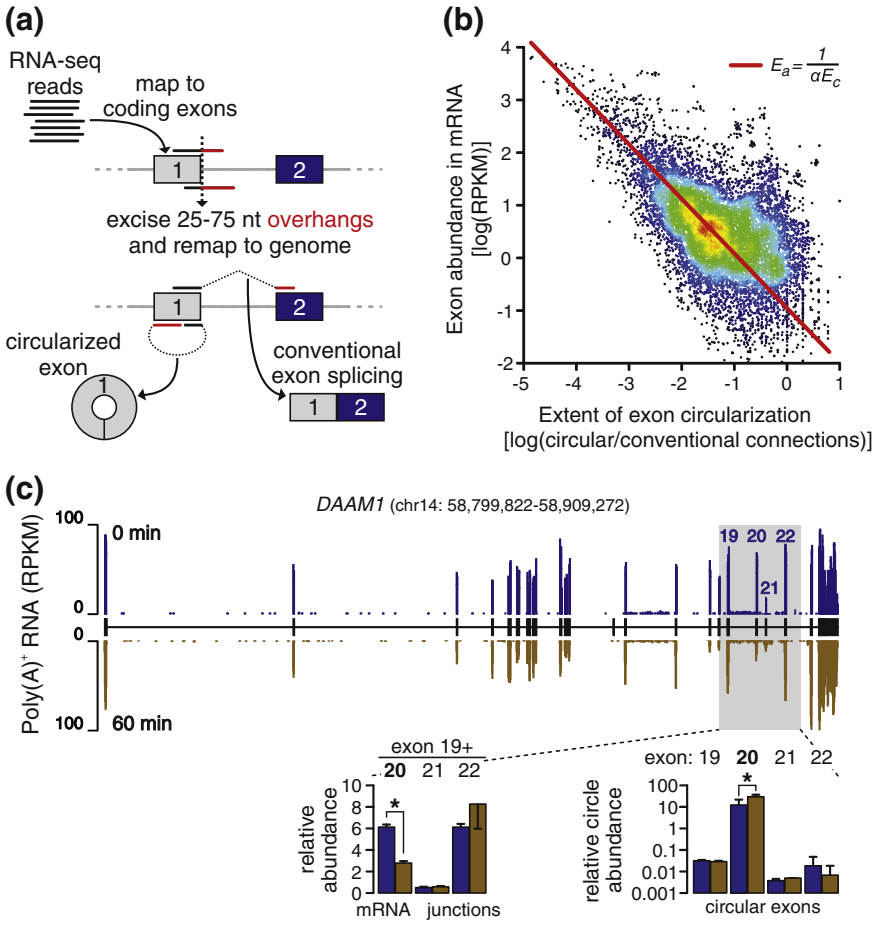


Fig. 1. Exon circularization is detected by RNA-seq and linked to exon skipping. (a) Strategy. Total, ribodepleted, RNA-seq 100-bp-long reads were mapped to expressed exons (orange and blue boxes); those producing a 25- to 75-nt overhang (red) were selected, overhangs were excised and remapped to the genome, and reads reporting exon circularization (i.e., "head-to-tail" mapping of the two read parts) or conventional splicing (i.e., exon–exon connections) were identified. (b) Relationship between exon circularization and exon abundance in mRNA. The equation describes the trendline (red), where E_a is relative abundance of an exon in poly(A)$^+$ RNA, E_c is the ratio of circular to conventional connections made by an exon, and α is the slope. (c) Validation of exon circularization in the alternatively spliced *DAAM1* gene. Top: The *DAAM1* gene model with juxtaposed tracks showing all reads in poly(A)$^+$ RNA at 0 min (blue) and at 60 min (brown) after TNFα stimulation. The levels of exon 20 drop after 60 min, indicating exon skipping. Bottom: quantitative RT-PCR quantitation of exon–exon connections (with primer pairs targeting exon 19 plus exon 20, or 21, or 22) and single-exon circles (with divergent primers targeting exons 19–22). Levels (±SD; $n = 4$) are normalized to those of *GAPDH* mRNA or the relevant exon–exon junction. *$P < 0.01$, two-tailed unpaired Student's t test.

common than those derived from a single exon [5]. To interrogate if production of multi-exon circles was also correlated with exon skipping, we compared their abundance to that of the same exons in mRNA and observed the correlation as for single-exon circles (Supplementary Fig. 2c). Moreover, consistent with previous analyses [5], our data revealed that circular connections between two different exons occurred >3-fold more frequently than between the two ends of the same exon (Supplementary Fig. 2c).

Our genome-wide analysis predicted that all skipped exons should produce circRNAs. However, inspection of RNA-seq data revealed that, for several genes with skipped exons, we could not robustly detect circular connections. To verify that circular products did form from these exons, we analyzed *DAAM1* transcripts using quantitative Reverse Transcription-PCR. This gene was chosen because exons 17, 18, and 20 are absent from its mRNA 60 min after TNFα stimulation, but their circularization was poorly detected in our RNA-seq data. In support of our genome-wide analysis, circular forms of all three skipped exons were detected here, and their rates of inclusion in the *DAAM1* mRNA inversely correlated to the abundance of the respective circRNAs in RNA prepared from pooled and single HUVEC donors (Fig. 1c and Supplementary Fig. 2d and e).

Previous reports have shown that there are substantial changes in mRNA abundance post-stimulation including changes in alternative splicing [11]. As circRNAs are thought to be long-lived compared to their linear counterparts [3,4], we examined whether their turnover was regulated following cytokine induction. After verifying that stimulation repressed or activated key responsive genes as expected (Supplementary Fig. 3a), we monitored circRNA appearance and disappearance along both time courses after grouping circRNAs according to expression level (Supplementary Fig. 3b). Of the different circRNAs detected at each time, the vast majority persist throughout the 2-h time course (regardless of expression quartile), while those that appear to turnover rapidly are supported, thus few reads that most likely represent sampling “noise” (Supplementary Fig. 3c and d). We conclude that enhanced circRNA stability also holds true during transcriptome remodeling in response to stimulation, with very few exceptions (e.g., exon 7 of *UBN2* or exon 5 of *ESCO1*).

Finally, given that circularization has recently been shown to be facilitated by the identity of the sequences flanking the circRNA on the linear transcript [12,13] and that here we uncover a clear link to exon skipping, it will be interesting to follow up on how these two components mechanistically combine to add yet another regulatory layer to the human transcriptome.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2015.02.018.

Acknowledgments

We thank the Biotechnology and Biological Sciences Research Council via ERASysBio+/FP7 (P.R.C.), the Genome Analysis Centre Capacity and Capability Challenge program (P.R.C. and S.K.), the Leverhulme Trust (S.K.), and the Centre for Molecular Medicine, Cologne (A.P.), for support. The authors would also like to thank the two anonymous reviewers whose advice and requested additions improved the manuscript.

Author Contributions: S.K., P.R.C., and A.P. conceived and designed experiments; A.P. performed experiments; C.G. performed RNA sequencing; S.K. wrote custom algorithms and performed data analysis with A.P.; S.K., P.R.C., and A.P. wrote the manuscript.

Received 3 December 2014;
Received in revised form 6 February 2015;
Accepted 13 February 2015
Available online 26 February 2015

Keywords:
circRNA;
cotranscriptional;
alternative splicing;
turnover

Abbreviations used:
circRNA, circular RNA; HUVEC, human umbilical vein endothelial cell; TNFα, tumor necrosis factor α; RT, reverse transcription.

References

[1] Lasda E, Parker R. Circular RNAs: diversity of form and function. RNA 2014;20:1829–42.
[2] Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One 2012;7:e30733.
[3] Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA 2013;19:141–57.
[4] Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 2013;495:333–8.
[5] Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. Genome Biol 2014;15:409.
[6] Wang PL, Bao Y, Yee MC, Barrett SP, Hogan GJ, Olsen MN, et al. Circular RNA is expressed across the eukaryotic tree of life. PLoS One 2014;9:e90859.

[7] Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. Nature 2013;495:384–8.
[8] Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, et al. circRNA biogenesis competes with pre-mRNA splicing. Mol Cell 2014;56:55–66.
[9] Farrell CM, O'Leary NA, Harte RA, Loveland JE, Wilming LG, Wallin C, et al. Current status and new features of the Consensus Coding Sequence database. Nucleic Acids Res 2014;42:D865–72.
[10] Dutertre M, Sanchez G, De Cian MC, Barbier J, Dardenne E, Gratadou L, et al. Cotranscriptional exon skipping in the genotoxic stress response. Nat Struct Mol Biol 2010;17:1358–66.
[11] Hao S, Baltimore D. RNA splicing regulates the temporal order of TNF-induced gene expression. Proc Natl Acad Sci USA 2013;110:11934–9.
[12] Liang D, Wilusz JE. Short intronic repeat sequences facilitate circular RNA production. Genes Dev 2014;28:2233–47.
[13] Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. Cell 2014;159:134–47.

SUPPLEMENTARY DATA

Exon skipping is correlated with exon circularization

Steven Kelly[1,*], Chris Greenman[2], Peter R. Cook[3], and Argyris Papantonis[3,4,*]

[1]*Department of Plant Sciences, University of Oxford, South Parks Road, OX1 3RB Oxford, UK*

[2]*The Genome Analysis Centre, Norwich Research Park, NR4 7UH Norwich, UK*

[3]*Sir William Dunn of Pathology, University of Oxford, South Parks Road, OX1 3RE Oxford, UK*

[4]*Center for Molecular Medicine, University of Cologne, Robert-Koch-Strasse 21, D-50931 Cologne, Germany*

**Corresponding authors*

Supplementary Data include:

—Materials and methods
—Supplementary Figures 1-3
—Supplementary Tables 1-2

Materials and Methods

Cell culture. Human umbilical vein endothelial cells (HUVECs) from pooled or single donors (**Supplementary Fig. 2d**) were grown at 37°C in EBM-2 basal medium complemented with 5% fetal calf serum and supplements (Lonza), starved overnight in EBM-2 plus 0.5% fetal calf serum (to synchronizes cells in G0 phase), stimulated for 0-120 min by addition of 10 and 50 ng/ml of TNFα and TGFβ (Peprotech), respectively, and harvested (at >85% confluence) for downstream analysis.

RNA isolation and qRT-PCR. $5x10^5$ HUVECs were harvested in QIAzol (Qiagen). Total RNA was isolated (RNAeasy kit; Qiagen), and DNase-treated (RQ1 DNase, 1 unit/µg total RNA for 30 min at 37°C; Promega). Where circRNAs were examined, total RNA was pretreated with RNase R (30 min at 37°C; Epicentre) to remove linear RNA. Following first-strand cDNA synthesis using oligo-dT or exon-specific primers, and Superscript II (Invitrogen) as per manufacturer's instructions, qPCR was performed on a RotorGene Q (Qiagen) using the QuantiTech SYBR Green Mix (Qiagen) in a typical three-step reaction with the primers listed in **Supplementary Table 2**.

RNA sequencing (RNA-seq) and data analysis. HUVEC total RNA (>200 nt) was isolated, DNase-treated (as above), ribodepleted (RiboMinus kit; Epicentre), chemically sheared to 250±100 nt, and reverse-transcribed using random hexamers (TrueSeq protocol; Illumina) with or without selection on poly(dT) beads, to produce "poly(A)$^+$-enriched" or "total RNA" libraries. Following adapter ligation, libraries were sequenced (~120 million read pairs per time point per stimulus; ~2.9 billion read pairs in total) on an Illumina HiSeq2000 platform. Individual 100-bp reads were mapped to the CCDS human exon set (hg19) using "stampy"[14], allowing 1-bp mismatch per 40 bp. Reads mapping to any individual exon producing a 25-75 nt overhang were retained for further analysis. The overhangs which extended off the end of the exon were excised and remapped to the human genome (using stampy as above) to determine their origin. Reads were then classified as resulting from (i) conventional splicing (with a 3'-exon end connected to a downstream 5'-exon end, or 5'-exon end with an upstream 3'-exon end), (ii) read-through (i.e., contiguous mapping across an exon-intron or intron-exon boundary), or (iii) exon circularization (i.e., with 3' and 5' ends of the same exon joined together in a "head-to-tail" arrangement; as in ref. 4).

Data access. RNA-seq data are freely available at http://bioinformatics.plants.ox.ac.uk/TGAC/

Supplementary references

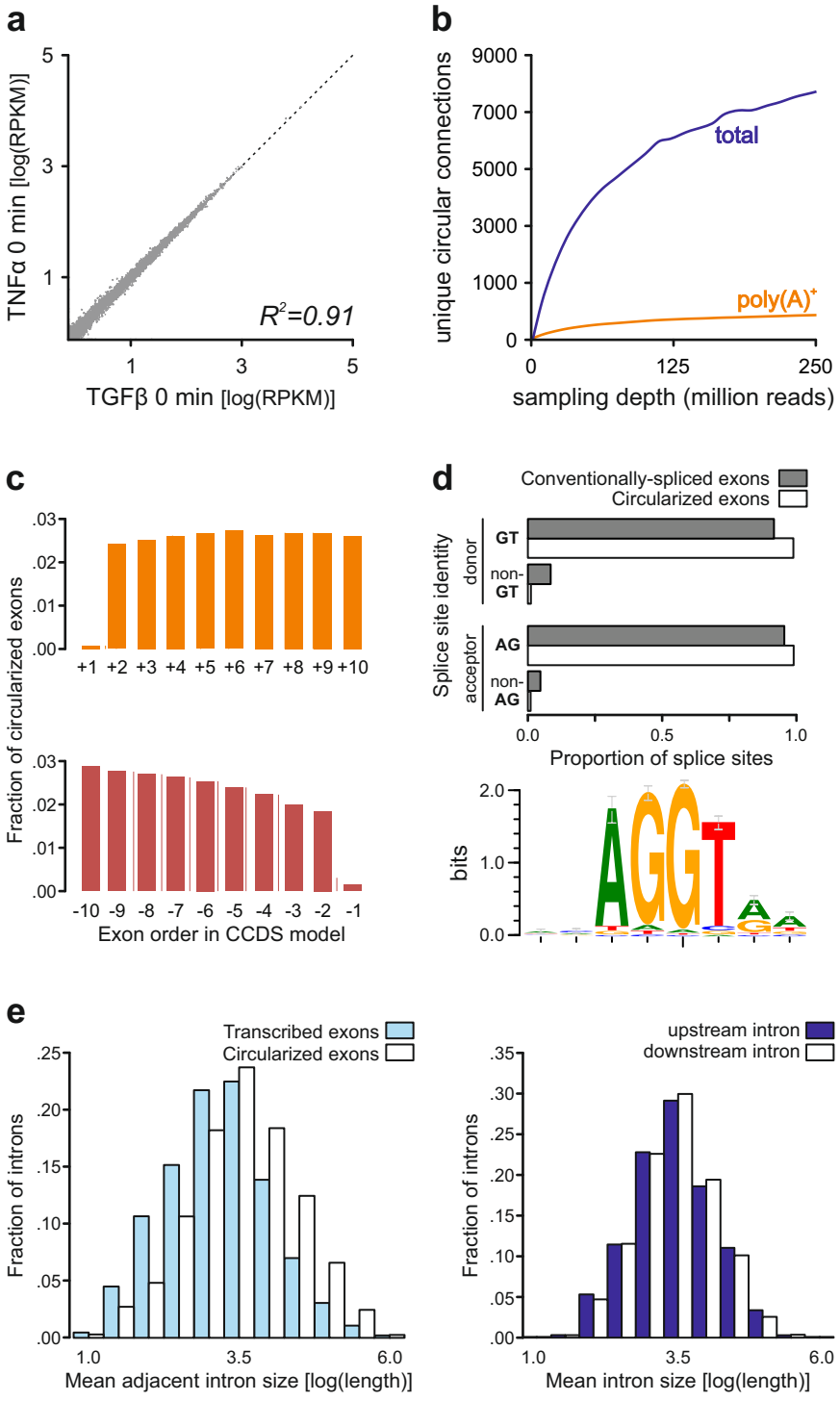
14. Lunter, G. & Goodson, M. (2011). Stampy, a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936-939.

15. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res*. **14**, 1188-1190.

16. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48.

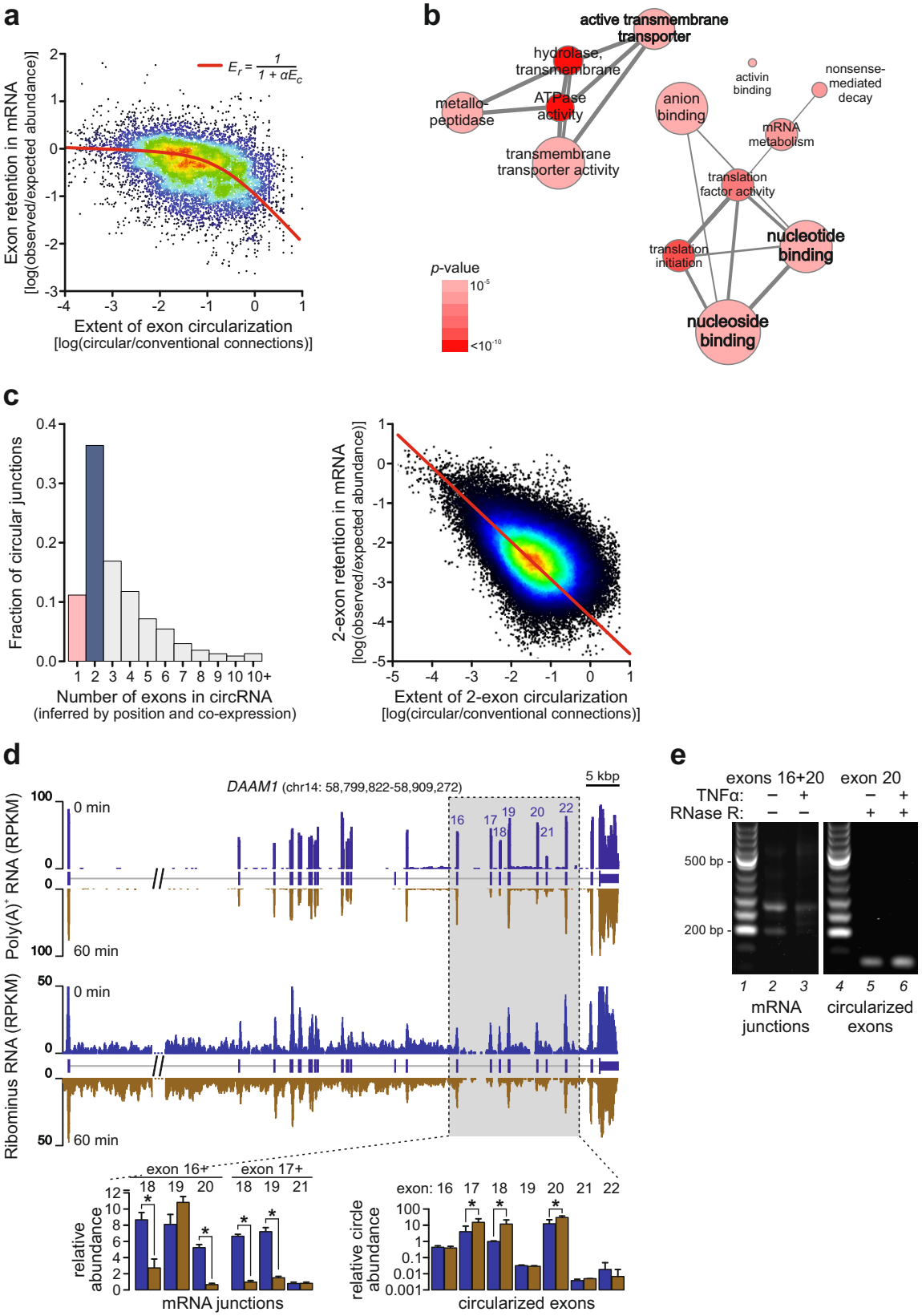
17. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800.

Supplementary Figure 1



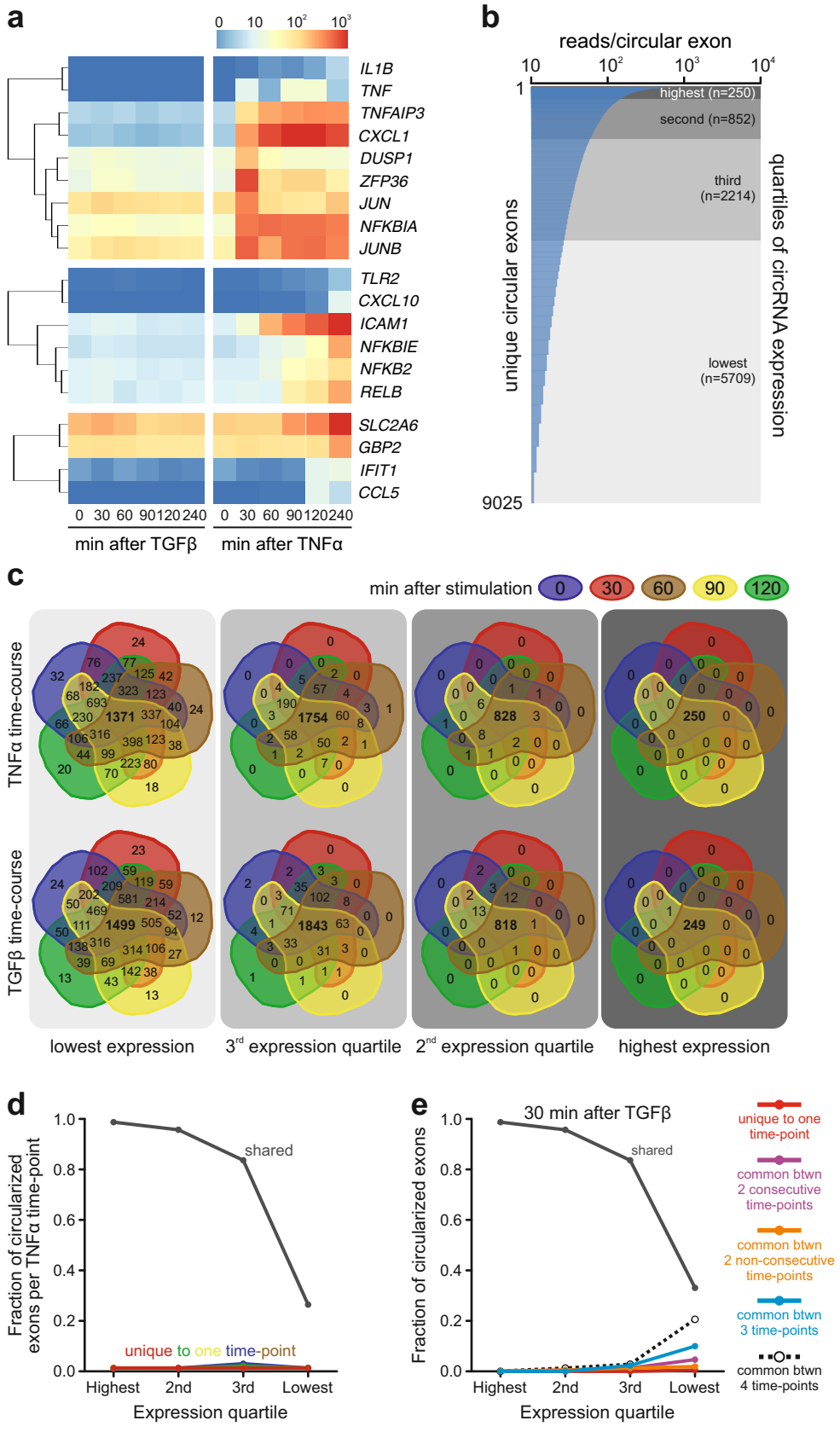
Supplementary Fig. 1. Reproducibility and features of circRNA analysis. (**a**) Log reads per kilobase per million (RPKM) from two biological replicates of total (ribodepleted) RNA-seq are well correlated (R^2=0.91, Spearman's correlation coefficient). (**b**) A plot describing the detection of any circularized exon in total (ribodepleted; *blue*) and poly(A)$^+$ (*orange*) RNA-seq data, as a function of sampling depth (in million reads). (**c**) Proportion of all recorded circular exons derived from the first (+1, …, +10) or last ten (-1, …, -10) exons in CCDS gene models. The apparent absence of circles derived from first or last exons is due to the absence of amenable "donor" or "acceptor" sites, respectively, as these exons include the 5' and 3' UTRs. (**d**) Splice donor/acceptor sequences. Bars show the frequency with which GT/AG and non-GT/non-AG dinucleotides occurs at the 5' "donor" or 3' "acceptor" splice site at linear (*grey*) and circular (*white*) exon ends (differences not statistically-significant; Fischer's exact test); a consensus logo derived from the data is also shown (generated using WebLogo[15]). (**e**) *Left*: introns flanking circularized exons (*white*) are longer than those flanking all transcribed exons (*light blue*) *Right*: no marked difference in length is seen between the up- (*blue*) and downstream intron (*white*) of a given circularized exon.

Supplementary Figure 2



Supplementary Fig. 2. Circular-exon GO terms, exon skipping, and validations. (**a**) Exclusion of an exon from an mRNA correlates with its appearance as a circle. The graph depicts the relationship between exon circularization (E_c, or the ratio of circular to non-circular connections made by an exon, assessed using total RNA) and exon retention (E_r, or the observed abundance of one particular exon in a mRNA compared to the abundance of all other exons in the same mRNA, assessed using poly(A)$^+$ RNA). The equation describing the trendline (*red*) is given at top right, where α (a constant) is 6.1. (**b**) GO term analysis on the 2,606 that produce the 9,025 circular exons analyzed here was performed using "GOrilla"[17] (using a cutoff of *p*-value $<10^{-5}$), and visualized as a network model via REViGO[18] (circle size represents relevant number of gene entries, color represents associated enrichment *p*-value, and line thickness represents significance of inter-term relevance). (**c**) Circular exons derived from more than a single exon. *Left*: The fraction of all "back-spliced" circular junctions occupied by single- (*pink*), two- (*blue*), or multi-exon (*grey*) circles is plotted. Note here that for ≥3 exons in a given circle their coexistence is inferred by their relative positions and co-expression (i.e., if the 3' end of exon 3 is found "back-spliced" to the 5' end of exon 1, and exons 1-3 are all skipped, then the circle is considered to contain all three of them). *Right*: The abundance of 2-exon-containing circles is plotted against the retention of these exons in mRNA; the correlation is essentially identical to that seen for single-exon circles (**Figure 1b**). (**d**) *Top*: The Ensemble gene model for *DAAM1* with juxtaposed tracks showing all reads in poly(A)$^+$ RNA at 0 (*blue*) and 60 min (*brown*) after TNFα stimulation. *Middle*: The same gene model with juxtaposed tracks showing all reads in total RNA-seq at the same times. In both, levels of exons 17, 18, and 20 are reduced after 60 min, indicative of exon skipping. *Bottom*: qRT-PCR quantitation of exon-exon connections (with primers targeting exon 17 plus exon 18, or 19, or 21 – or exon 19 plus exon 20, or 21, or 22) and single-exon circles (with divergent primers targeting one of exons 17-22). Levels (± S.D.; *n*=4) are normalized to those of *GAPDH* mRNA or the relevant exon-exon junction, respectively. Splicing of exon 17 to 18 (or 19) – and of exon 19 to 20 – falls as the corresponding exonic circles rise. *: $P<0.01$, two-tailed unpaired Student's *t*-test. (**e**) Typical images of gels containing linear exon-exon (*left*) and circular exon amplicons (*right*) derived from *DAAM1* RNA. HUVECs from a single donor were treated for 0 (-) or 60 (+) min with TNFα, total RNA isolated and treated ± RNase R, RT-PCR performed using primers targeting either exons 16 + 20 or the ends of exon 20, amplimers run on gels, and imaged. *Lanes 1, 4*: 50-bp DNA ladder (NEB) with 200- and 500-bp bands indicated. *Lanes 2, 3*: short and long amplimers contain exons 16 and 20, and 16-20, respectively, the levels of which are reduced by TNFα. *Lanes 5, 6*: amplimers derived from exon-20 circles are increased by the cytokine.

Supplementary Figure 3



Supplementary Fig. 3. Single-exon circularization along two cytokine-induced time-courses. (a) Heatmaps show that the expression profiles of some early (*top*), intermediate (*middle*), and middle-late TNFα-responsive genes (*bottom*; from ref. 11) are recapitulated in our data at the different times after stimulation with TGFβ or TNFα. The color code (upper-right corner) reflects RPKM expression levels. (**b**) Numbers of reads attached to each of 9,025 circular exons detected are plotted, and exons are the dinned into four quartiles based on their cumulative expression levels (*grey* color-code) so that each bin represents ~100,000 reads. (**c**) Venn diagrams depict exonic circRNAs, categorized according to expression levels as described in panel (b), shared or unique at different times after stimulation. (**d**) Turnover of circular exons (binned in four quartiles as in panel b) as illustrated by the fraction seen either throughout the time-course ("shared") or only at a particular time-point ("unique") after TNFα stimulation. (**e**) Turnover of circularized exons (binned in four quartiles as in panel b) detected 30 min after TGFβ stimulation as illustrated by the fraction seen throughout the time-course ("shared"; *grey*), or detected only at 30 min ("unique"; *red*), at 30 and 0 or 60 min (*purple*), at 30 and 90 or 120 min (*orange*), at 30 min and any two other times (*blue*), or at 30 min and any three other times (*black dashed line*).

Supplementary Table 1. Features of al detected circular exons. A list of 9,025 circular exons detected in this study (each with ≥11 in total). Gene IDs, chromosomal location, gene name and exon number (in the respective CCDS model) are shown alongside RPKM values mapping to the circular form per each time-point following TGFβ and TNFα stimulation; dinucleotide sequence at their 5' "donor" and 3' "acceptor" splicing sites and their genomic coordinates are also shown.

[*Table provided as an .xls file*]

Supplementary Table 2. Primers used in qRT-PCR. Oligonucleotides targeting exons 16-22 of *DAAM1* were designed using www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/ and are listed alongside their sequence and melting temperature. "For", sense primer; "Rev", antisense primer; "J", primer spans exon circle junction; "circ", targets the body of the exon.

Name	Sequence (5' to 3')	Tm (°C)
Exon16For	AACAGGAAGATCTGCCCAAG	59.3
Exon16circRev	TGCCCGTTTGATTTCGTC	59.7
Exon17For	AACACGAACTGGATCGGATG	60.9
Exon17-JRev	ATTTCAAGAGCGGCTCATCTC	60.9
Exon17Rev	GGCTCATCTCAAAAAGGAACC	60.1
Exon18-JFor	GTGGAAGAATTAATCACTATCAGC	60.0
Exon18Rev	GGTTTCACTTCTGCCACACG	61.7
Exon19For	GACACAAAATCCAGCATCGAC	60.5
Exon19Rev	AAAACCACCTCCAGCAACTG	60.1
Exon20For	TGCGAGATATTCCTCAAGCTG	60.5
Exon20Rev	TTTACTTTCGCAGCTTGAGG	58.3
Exon20circRev	ACACTGGGGTACTTATTTTCCAC	58.4
Exon21-JFor	AGAGACACATGACTGAGCTGGAC	60.9
Exon21Rev	CTTTCAAGCCACTTCTCAAGG	59.1
Exon22-JFor	AGCAGAAGCTAAAGACCTGGAG	59.3
Exon22Rev	GATGAACTGGCTGACAACAGAC	59.8