https://doi.org/10.1038/s41556-022-01066-9

A guide to naming eukaryotic circular RNAs

Ling-Ling Chen, Albrecht Bindereif, Irene Bozzoni, Howard Y. Chang, A. Gregory Matera, Myriam Gorospe, Thomas B. Hansen, Jørgen Kjems, Xu-Kai Ma, Jun Wei Pek, Nikolaus Rajewsky, Julia Salzman, Jeremy E. Wilusz, Li Yang & Fangqing Zhao

Alternative splicing of eukaryotic messenger RNA transcripts often leads to the production of several mature RNAs — including linear RNAs and circular RNAs (circRNAs) — from a single gene locus. The names given to circRNAs are often ambiguous and lack consistency across studies. This Comment calls on the community to embrace a common nomenclature for naming circRNAs to ensure clarity and reproducibility.

As precursor RNAs (pre-RNAs) are being transcribed by RNA polymerase II, they must be spliced by the spliceosome in order to produce mature, functional RNAs. In the canonical splicing process, exon regions are sequentially joined to one another in a linear fashion, while introns are removed as lariats that are rapidly debranched and degraded (Fig. 1a, bottom). However, it has long been recognized that alternative splicing events can generate several mature RNAs from one gene locus, thereby expanding the diversity of functions encoded by a single gene[1]. For example, a pre-RNA can be subjected to 'back-splicing', when a downstream 5' splice site is covalently linked with an upstream 3' splice site to form a circular RNA (circRNA)[2–5] (Fig. 1a, top). The sequences of these transcripts often fully overlap with a subset of the exons (and introns) of their cognate linear RNAs, but the circRNAs can be distinguished by their unique back-splice junctions (BSJs). In addition, cells can produce a distinct set of circRNAs when introns fail to be debranched and instead accumulate[6,7] (Fig. 1a, bottom); these are known as circular intronic RNAs (ciRNAs) or stable intronic sequence RNAs (sisRNAs). Whereas circRNAs generated by back-splicing are joined by a 3',5'-phosphodiester bond, ciRNAs have a 2',5'-phosphodiester bond[4,5].

In recent years, there has been an explosion in the number of annotated circRNAs, thanks to the development of biochemical enrichment strategies, as well as computational algorithms that detect sequencing reads spanning BSJs and lariat branch points in RNA-sequencing (RNA-seq) data[2–5]. Hundreds of ciRNAs have been identified, with reported roles in gene transcription and early embryogenesis[6,7]. Many more circRNAs (hundreds of thousands) have been discovered across eukaryotes, showing that back-splicing is by far the most common mechanism through which endogenous RNA circles are generated (reviewed in refs. [2–5]).

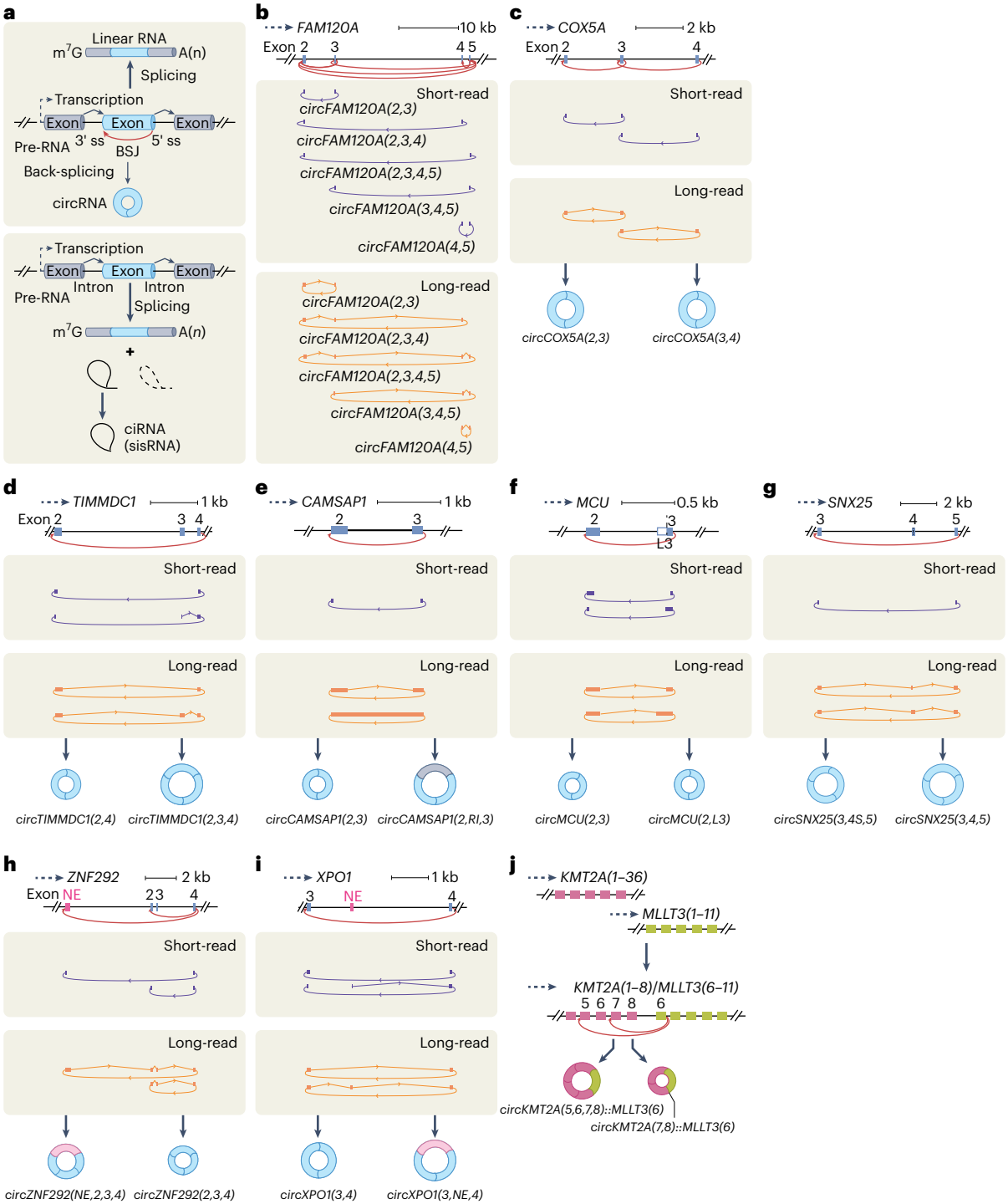
Most mature circRNAs are expressed at low levels, but certain genes produce them at much higher levels than their canonically spliced linear RNAs. This is due, in part, to the fact that circRNAs are inherently resistant to exonucleolytic decay and can have long half-lives in cells. The vast majority of circRNAs have yet to be studied in detail, but individual examples are known to serve as decoys for microRNAs or proteins, to facilitate the formation of functional ribonucleoprotein complexes, or to encode peptides/proteins that are translated (for example, in response to certain cellular stresses). In addition, circRNAs affect many physiological processes, including early development, immune responses, neurogenesis and tumorigenesis (reviewed in refs. [4,5]).

Dilemmas regarding the naming of circRNAs

Among all annotated circRNAs generated by back-splicing, perhaps the most well studied is derived from a long noncoding RNA (LINC00632)[8], and is known by two distinct names: *CDR1as* and *ciRS-7* (refs. [9–11]). This circRNA originates from a region antisense to the gene encoding cerebellar degeneration-related protein 1 (CDR1) — hence the name *CDR1as*. It also functions as a circRNA sponge for miRNA-7 — hence the name *ciRS-7*. The vast majority of other reported circRNAs have been named by adding the prefix 'circ' to their host gene's name; for example, *circFAM120A* indicates a circRNA generated from the *FAM120A* gene. One can clearly distinguish the circRNA from its cognate linear RNA in this manner.

However, this naming strategy does not include information regarding which exons (and introns) are included; nor does it make it possible to differentiate between multiple circRNAs produced from a single host gene. For example, in human HEK293/293FT cells, at least five distinct circRNAs are detected from the *FAM120A* locus in short-read[12] and long-read[13] sequencing data, with each mature circRNA containing a different combination of exons (Fig. 1b). Note that short-read sequencing reveals the sequences around each BSJ, but often cannot cover the internal region of circRNAs produced from the *FAM120A* locus (Fig. 1b, purple). In contrast, long-read sequencing can provide direct detection of all included exons (Fig. 1b, orange). One of these circRNAs (that containing exons 2–5) is substantially more abundant than the others, and binds the translation inhibitor IGF2BP2 on ribosomes[12]. So, the simple term *circFAM120A* is not sufficient to distinguish all five circRNAs from one another, let alone indicate which one has a characterized function.

A clear naming system for circRNAs is necessary — especially one that would allow both the host gene and the exact exons and introns present in the mature circRNA to be easily known. Nonetheless, the naming of RNA circles is beyond the scope of the HUGO Gene Nomenclature Committee at present[14]. Of note, many circRNAs have been archived by publicly available databases, such as CIRCpedia (http://yang-laboratory.com/circpedia/)[15], circAtlas (http://159.226.67.237:8080/new/index.php)[16], circBank



(http://www.circbank.cn)[17] and circBase (http://circbase.org)[18], and can be retrieved by database entry. However, different strategies were used to name circRNAs in distinct databases, and the resulting names are often not easily recalled. For example, the functional *FAM120A* circRNA (Fig. 1b) is known as HSA_CIRCpedia_64725 in CIRCpedia, hsa-FAM120A_0006 in circAtlas, hsa_circFAM120A_007 in circBank, and hsa_circ_0001875 in circBase (Table 1).

Here, we call on the community to discuss and embrace standards for naming circRNAs, so that a common nomenclature is used in databases and future studies to ensure clarity and reproducibility.

Suggested criteria for circRNA naming

To name circRNAs generated by back-splicing, we suggest including both the gene symbol and the numbers of exon components (as well as

Fig. 1 | Suggested naming scheme to clarify the complexity of patterns of expression of circRNAs. a, Top, in addition to generating linear RNAs, a gene can generate circRNAs when a 5' splice site (ss) is joined to an upstream 3′ splice site via a process known as back-splicing. Of note, the efficiency of back-splicing (thin arrow) is often low compared with that of canonical splicing (heavy arrow) at the same gene locus. Bottom, pre-RNAs can be spliced to generate mature linear RNAs that are capped (m^7G) and polyadenylated (A(*n*)), as well as intron lariats that are debranched and degraded (dashed loop). However, some lariats fail to be debranched, leading to the accumulation of stable ciRNAs and/or sisRNAs. **b**, Naming of human circRNAs produced from the gene *FAM120A* (ENST00000277165.11 annotation) with different back-splice junction (BSJ) sites. **c**, Naming of circRNAs produced from the gene *COX5A* (ENST00000322347.11 annotation) with different BSJ sites. **d**, Naming of two alternative circRNAs produced from the gene *TIMMDC1* (ENST00000494664.6 annotation) that differ in their inclusion of a cassette exon. **e**, Naming of circRNAs produced from the gene *CAMSAP1* (ENST00000389532.9 annotation) that differ in their inclusion of a retained intron (RI). **f**, Naming of circRNAs produced from the gene *MCU* (ENST00000603649.5 annotation) that differ in their use of an alternative 3' splice site. A longer exon 3 (L3) with additional sequence at the 5' end of the annotated exon 3 can be alternatively spliced within the circRNA produced from the *MCU* locus. **g**, Naming of circRNAs produced from the gene *SNX25* (ENST00000618785.4 annotation) that differ in their use of an alternative 5' splice site. **h**, Naming of circRNAs produced from the gene *ZNF292* (ENST00000369578.6 annotation) that differ in their BSJ and their inclusion of a novel exon (NE). **i**, Naming of circRNAs produced from the gene *XPO1* (ENST00000401558.7 annotation) that differ in their inclusion of a novel cassette exon. **j**, Naming of fusion circRNAs produced from the *KMT2A::MLLT3* (also known as *MLL::AF9*) translocation in leukaemic cell lines. *KMT2A(1–36)*, *KMT2A* exons 1–36; *MLLT3(1–11)*, *MLLT3* exons 1–11; *KMT2A(1–8)*, *KMT2A* exons 1–8; *MLLT3(6–11)*, *MLLT3* exons 6–11. Blue, grey, pink and green bars in gene assemblies represent exons; black lines represent introns. Canonical splice junctions ('polylines') and BSJs (arc lines) are shown in sequencing data. Short-read (purple) and long-read (orange) examples[12,13] are shown.

Table 1 | The functional *FAM120A* circRNA has a different identity in each circRNA database

Database	circID in each database	Genome assembly	Genomic location of circRNAs	Host gene	Exon components	Gene annotation
CIRCpedia	HSA_CIRCpedia_64725	GRCh38/hg38; GRCh37/hg19	chr9:93471140-93498886 (hg38); chr9:96233422-96261168 (hg19)	*FAM120A*	2,3,4,5	GENCODE plus RefSeq
circAtlas	hsa-FAM120A_0006	GRCh38/hg38	chr9:93471141-93498886 (hg38)	ENSG00000048828.16	N/A	GENCODE
circBank	hsa_circFAM120A_007	GRCh37/hg19	chr9:96233422-96261168 (hg19)	*FAM120A*	N/A	RefSeq
circBase	hsa_circ_0001875	GRCh37/hg19	chr9: 96233422-96261168 (hg19)	*FAM120A*	N/A	RefSeq

CIRCpedia (hg38, hg19), circBank (hg19) and circBase (hg19) use the 0-based coordinate system, while circAtlas (hg38) uses the 1-based coordinate system. Hence, annotation of the genomic start location of a circRNA will differ by one base depending on the database.

introns, if retained in the mature transcript; see below) after the prefix 'circ'. Such a naming system requires clear information regarding which linear transcript was used as a reference. This is because alternative promoters, alternative splice sites, and/or alternative polyadenylation signals often lead to a variety of distinct linear transcripts being produced from a given gene locus[1]. Here, one GENCODE/Ensembl transcript per gene locus is selected as the reference and, when possible, we suggest the use of standardized reference transcripts, such as MANE (Matched Annotation from NCBI and EMBL-EBI) transcripts[19].

To illustrate these proposed naming standards, we provide examples of different types of circRNAs below. Beyond using these names in future published studies, it is imperative that the linear reference transcripts used for all circRNA names are also provided (for example, when names are first used in the text or figures). In addition, gene-locus schematics that include the linear reference transcript (similar to the top part of Fig. 1b, for example) and genome coordinates, along with the assembly version, should be provided. Given that not all transcript data will necessarily be available in perpetuity, providing the actual sequence of the circRNA will help to ensure long-term clarity.

Multiple circRNAs produced from the same gene locus with different BSJ sites. Back-splicing of the human *COX5A* locus produces two distinct circRNAs that differ in their composition owing to alternative BSJ selection: one circRNA consists of exons 2 and 3 (according to the ENST00000322347.11 transcript), the other of exons 3 and 4 (Fig. 1c). Rather than calling either transcript *circCOX5A*, we instead suggest naming them *circCOX5A(2,3)* and *circCOX5A(3,4)*, respectively, where (2,3) and (3,4) indicate exons 2,3 and exons 3,4. Likewise, in the case of the *FAM120A* gene locus (Fig. 1b), the five circRNAs can be named to indicate their different exonic components: *circFAM120A(2,3)*, *circFAM120A(2,3,4)*, *circFAM120A(2,3,4,5)*, *circFAM120A(3,4,5)*, and *circFAM120A(4,5)*, according to the ENST00000277165.11 transcript. Beyond allowing these circRNAs to be distinguished from one another, the names also illustrate the position of the BSJ: the end of the last listed exon was back-spliced to the beginning of the first listed exon. So, for *circFAM120A(2,3)*, the end of exon 3 was back-spliced to the beginning of exon 2, while the end of exon 4 was back-spliced to the beginning of exon 2 in *circFAM120A(2,3,4)* (Fig. 1b).

In situations in which only one circRNA from a gene has yet been found, we still recommend including the exact exon(s) from the mature circRNA in the name. Doing so will ensure consistency across all circRNA names, and will help to avoid ambiguities if additional circRNA isoforms are found in the future.

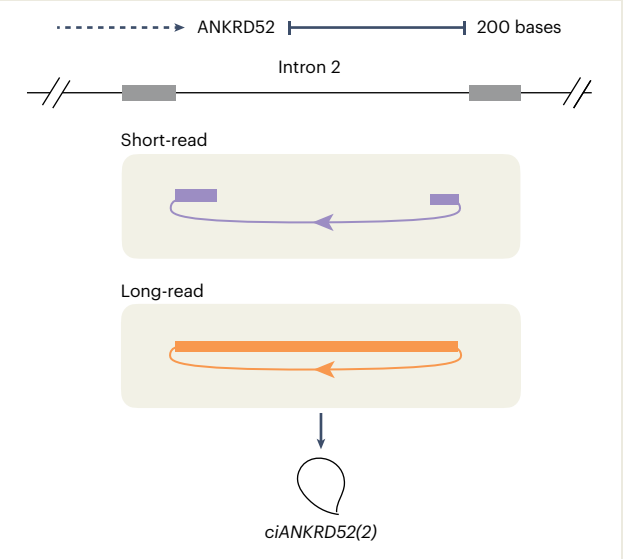
CircRNAs with the same BSJs but different internal splicing patterns. Genes can generate several circRNAs with distinct BSJ sites (Fig. 1b,c), yet can also produce circRNAs with the same BSJs but differing internal sequences[4]. This is because all four standard types of alternative splicing event — such as inclusion of a cassette exon (or the converse, skipping of such an exon), intron retention, and use of alternative 5' or 3' splice sites — have been identified in the internal regions of circRNAs[4].

Inclusion of a cassette exon has frequently been observed in alternatively spliced circRNAs[4]. For example, the human *TIMMDC1* gene

BOX 1

Suggested guidelines for ciRNA naming

Analogous to our suggested guidelines for naming circRNAs produced by back-splicing, a similar strategy can be used to name ciRNAs that accumulate when intron lariats fail to be debranched (Fig. 1a, bottom). We suggest the alternative prefix 'ci' for ciRNAs, followed by the gene symbol and the specific intron number. For example, the name *ciANKRD52(2)* denotes a ciRNA produced from the second intron of the *ANKRD52* gene (according to the ENST00000267116.8 transcript annotation)[6] (see Box Figure). In general, ciRNAs and sisRNAs[6,7] have not been as extensively studied as circRNAs generated by back-splicing, and more work is needed to understand their modes of biogenesis, localization and function. Regardless, the linear reference transcript used for all ciRNA names should be provided in published work to ensure that this simple naming system can easily distinguish ciRNA isoforms from one another.



Naming of a ciRNA produced from an intron lariat. *ciANKRD52(2)* is produced from the second intron of the *ANKRD52* gene.

generates two circRNAs that share the same BSJ (with the end of exon 4 joined to the start of exon 2) but differ in their inclusion or exclusion of the cassette exon 3 (Fig. 1d). According to the ENST00000401558.7 transcript of *TIMMDC1*, we suggest calling these transcripts *circTIMMDC1(2,3,4)* and *circTIMMDC1(2,4)*, depending on the inclusion or exclusion of exon 3.

For circRNAs that have a retained intron (RI) — previously referred to as exon–intron circular RNAs[20] — we suggest the same naming scheme as above, but including 'RI' to indicate the existence of an internal RI. As an example, two circRNAs are produced from the human *CAMSAP1* gene (Fig. 1e). Both have the end of exon 3 back-spliced to the beginning of exon 2, but one isoform contains an RI between these exons (observed directly in long-read sequencing data; Fig. 1e). Hence, we suggest naming these transcripts *circCAMSAP1(2,3)* and *circCAMSAP1(2,RI,3)* (according to the ENST00000389532.9 annotation) to indicate the presence of the retained intron only in the latter.

For circRNAs that contain 5' or 3' alternatively spliced exons, we suggest adding 'S' (for 'short') or 'L' (for 'long') to the number of the exon with the alternative splice-site selections. For example, two circRNAs are produced from the human *MCU* gene that have the same BSJ site but different internal 3' splice site selections (Fig. 1f): one contains the annotated exon 3 according to the ENST00000603649.5 transcript, and the other contains a longer exon 3. These circRNAs can be referred to as *circMCU(2,3)* and *circMCU(2,L3)*, respectively, so that their differences can be clearly noted. Note that 'L' is written before the exon number, to indicate the usage of an alternative 3' splice site that results in a longer exon 3. In the case of alternative 5' splice site selections within circRNAs, the S or L labelling can be added after the number of the alternatively spliced exon. For example, two circRNAs are produced from the human *SNX25* gene that have the same BSJ site but different 5' splice site selections for the internal exon 4 (Fig. 1g). Use of *circSNX25(3,4S,5)* and *circSNX25(3,4,5)* can distinguish these two circRNA isoforms, which have either a short version of exon 4 or a full-length version, respectively, according to the ENST00000618785.4 annotation. In those rare cases where there are more than two alternative splice sites, consecutive letters (A, B, C and so on) may instead need to be used to distinguish among isoforms.

CircRNAs with previously unannotated exons. Recent work has indicated that a number of circRNAs include previously unannotated exons that are not observed in linear RNAs[4,21,22]. In these cases, we suggest marking such a sequence as a 'new exon' (NE) when naming the circRNA. As shown in Fig. 1h, the human *ZNF292* gene generates a three-exon circRNA (*circZNF292(2,3,4)*, according to the ENST00000369578.6 annotation) as well as an additional circRNA that contains a previously unannotated exon upstream of exon 2. As well as not being present in the ENST00000369578.6 transcript, this new exon is not annotated in any assembled *ZNF292* transcript in GENCODE V37. Hence, we suggest naming this circRNA *circZNF292(NE,2,3,4)*.

In another example, the human *XPO1* gene generates two circRNAs that have the same BSJ but differ in the presence of an internal novel cassette exon (Fig. 1i). According to the ENST00000401558.7 annotation, we suggest naming these two circRNAs *circXPO1(3,4)* and *circXPO1(3,NE,4)*, indicating whether the novel exon is skipped or included, respectively. If multiple unannotated exons are present in a circRNA, the NE notation can be used multiple times to indicate each of their positions. We recognize, however, that this label is not ideal, and it should be viewed as a temporary placeholder. Further efforts to annotate the transcriptome — especially by incorporating circRNAs into databases such as GENCODE and RefSeq — should allow more unambiguous names to be given in the future. On rare occasions, unannotated splice sites may occur in circRNAs, and including their genomic positions in the naming scheme may be necessary to specify the exact splice sites and sequence contents of the mature transcript.

Fusion circRNAs from translocated gene loci. Most circRNAs are produced from a single gene, but some contain exons derived from two different genes, mostly in regions that have experienced chromosomal translocations and gene fusion[23]. As an example, at least two distinct fusion circRNAs have been reported to originate from the *KMT2A::MLLT3* (also known as *MLL::AF9*) translocation in leukaemic

cell lines[24] (Fig. 1j). We suggest naming these circular transcripts *circKMT2A(5,6,7,8)::MLLT3(6)* and *circKMT2A(7,8)::MLLT3(6)*, including both host gene names but separating them by a double colon to designate a gene fusion (as recommended in ref. [23]). This allows one to distinguish between these transcripts, and to infer that exon 6 (according to the ENST00000380338.9 annotation) of *MLLT3* (*AF9*) has been back-spliced to exon 5 or 7 (according to the ENST00000534358.8 transcript) of *KMT2A* (*MLL*).

Conclusions

Given the wide variety of circular transcripts observed in eukaryotic cells — and the fact that numerous distinct RNA circles can be derived from a single host gene — it is becoming increasingly clear that a systematic naming scheme for circRNAs is needed. To limit ambiguity, we recommend that the names of circRNAs produced from back-splicing should include the prefix 'circ' followed by the host gene symbol and exon (and, if present, intron) information. An analogous strategy could be used to name ciRNAs derived from intron lariats by using the alternative prefix 'ci' (Box 1).

These naming schemes are intuitive and can be used to name circular transcripts in any eukaryotic species. Nonetheless, they depend on existing annotations of linear transcripts. Our proposed scheme thus represents an interim solution until circRNAs are incorporated into annotation databases such as GENCODE and RefSeq. Such a step will be the only true long-term solution, as it will remove ambiguity in circRNA names (for example, by enabling the removal of NE labels) and eliminate the need to have a reference linear transcript. In the meantime, it is essential that the reference transcripts used for all circRNA names be provided in publications.

We call on authors, reviewers and editorial staff to ensure that manuscripts that report circRNAs provide: (1) the genomic coordinates and associated genome assembly; (2) the full sequences of circRNAs; (3) any previously given names, for example in circRNA databases; and (4) clear diagrams that depict where circRNAs of interest map in the genome. This will help greatly to facilitate communication and eliminate confusion when several circRNAs are produced from the same gene locus. More and more circRNAs continue to be identified, including some that are presumably derived from intron self-ligation[21] and some from noncoding RNAs, including transfer RNAs, ribosomal RNAs and mitochondrial RNAs (reviewed in refs. [4,25]). As the field progresses, additional naming criteria can be proposed for these new transcript classes.

Ling-Ling Chen[1] ✉, Albrecht Bindereif[2], Irene Bozzoni[3], Howard Y. Chang[4], A. Gregory Matera[5], Myriam Gorospe[6], Thomas B. Hansen[7], Jørgen Kjems[7], Xu-Kai Ma[8], Jun Wei Pek[9], Nikolaus Rajewsky[10], Julia Salzman[11], Jeremy E. Wilusz[12] ✉, Li Yang[8] ✉ & Fangqing Zhao[13]

[1]State Key Laboratory of Molecular Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [2]Institute of Biochemistry, Faculty of Biology and Chemistry, Justus Liebig University of Giessen, Giessen, Germany. [3]Department of Biology and Biotechnologies 'Charles Darwin' and IIT Center for Life Nano- & Neuro-Science@ Sapienza, Sapienza University of Rome, Rome, Italy. [4]Center for Personal Dynamic Regulomes, Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. [5]Integrative Program in Biological and Genome Sciences, University of North Carolina, Chapel Hill, NC, USA. [6]Laboratory of Genetics and Genomics, National Institute on Aging Intramural Research Program, National Institutes of Health, Baltimore, MD, USA. [7]Interdisciplinary Nanoscience Center (iNANO), Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark. [8]Center for Molecular Medicine, Children's Hospital, Fudan University and Shanghai Key Laboratory of Medical Epigenetics, International Laboratory of Medical Epigenetics and Metabolism, Institutes of Biomedical Sciences, Fudan University, Shanghai, China. [9]Temasek Life Sciences Laboratory, 1 Research Link National University of Singapore, Singapore, Singapore. [10]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Hemholtz Association, Berlin, Germany. [11]Department of Biochemistry, Stanford University, Stanford, CA, USA. [12]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Therapeutic Innovation Center, Baylor College of Medicine, Houston, TX, USA. [13]Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing, China.
✉e-mail: linglingchen@sibcb.ac.cn; jeremy.wilusz@bcm.edu; liyang_fudan@fudan.edu.cn

Published online: 19 January 2023

References

1. Nilsen, T. W. & Graveley, B. R. *Nature* **463**, 457–463 (2010).
2. Szabo, L. & Salzman, J. *Nat. Rev. Genet.* **17**, 679–692 (2016).
3. Kristensen, L. S. et al. *Nat. Rev. Genet.* **20**, 675–691 (2019).
4. Yang, L., Wilusz, J. E. & Chen, L. L. *Annu. Rev. Cell Dev. Biol.* **38**, 263–289 (2022).
5. Liu, C. X. & Chen, L. L. *Cell* **185**, 2016–2034 (2022).
6. Li, X. et al. *Sci. China Life Sci.* **64**, 1795–1809 (2021).
7. Tay, M. L. & Pek, J. W. *Curr. Biol.* **27**, 1062–1067 (2017).
8. Barrett, S. P., Parker, K. R., Horn, C., Mata, M. & Salzman, J. *PLoS Genet.* **13**, e1007114 (2017).
9. Memczak, S. et al. *Nature* **495**, 333–338 (2013).
10. Hansen, T. B. et al. *Nature* **495**, 384–388 (2013).
11. Hansen, T. B. et al. *EMBO J.* **30**, 4414–4422 (2011).
12. Li, S. et al. *Nat. Methods* **18**, 51–59 (2021).
13. Xin, R. et al. *Nat. Commun.* **12**, 266 (2021).
14. Seal, R. L. et al. *EMBO J.* **39**, e103777 (2020).
15. Dong, R., Ma, X. K., Li, G. W. & Yang, L. *Genom. Proteom. Bioinf.* **16**, 226–233 (2018).
16. Wu, W., Ji, P. & Zhao, F. *Genome Biol.* **21**, 101 (2020).
17. Liu, M., Wang, Q., Shen, J., Yang, B. B. & Ding, X. *RNA Biol.* **16**, 899–905 (2019).
18. Glazar, P., Papavasileiou, P. & Rajewsky, N. *RNA* **20**, 1666–1670 (2014).
19. Morales, J. et al. *Nature* **604**, 310–315 (2022).
20. Li, Z. et al. *Nat. Struct. Mol. Biol.* **22**, 256–264 (2015).
21. Zhang, J. et al. *Nat. Biotechnol.* **39**, 836–845 (2021).
22. Rahimi, K., Veno, M. T., Dupont, D. M. & Kjems, J. *Nat. Commun.* **12**, 4825 (2021).
23. Bruford, E. A. et al. *Leukemia* **35**, 3040–3043 (2021).
24. Guarnerio, J. et al. *Cell* **165**, 289–302 (2016).
25. Lasda, E. & Parker, R. *RNA* **20**, 1829–1842 (2014).

Competing interests

J.E.W. is a consultant for Laronde. H.Y.C. is a co-founder of Orbital Therapeutics. The other authors declare no competing interests.

Additional information

Peer review information *Nature Cell Biology* thanks Igor Ulitsky, Shenglin Huang, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.