



Published in final edited form as:

*Nat Cell Biol.* 2023 January ; 25(1): 1–5. doi:10.1038/s41556-022-01066-9.

## A guide to naming eukaryotic circular RNAs

Ling-Ling Chen<sup>1,\*</sup>, Albrecht Bindereif<sup>2</sup>, Irene Bozzoni<sup>3</sup>, Howard Y. Chang<sup>4</sup>, A. Gregory Matera<sup>5</sup>, Myriam Gorospe<sup>6</sup>, Thomas B. Hansen<sup>7</sup>, Jørgen Kjems<sup>8</sup>, Xu-Kai Ma<sup>9</sup>, Jun Wei Pek<sup>10</sup>, Nikolaus Rajewsky<sup>11</sup>, Julia Salzman<sup>12</sup>, Jeremy E. Wilusz<sup>13,\*</sup>, Li Yang<sup>9,\*</sup>, Fangqing Zhao<sup>14</sup>

<sup>1</sup>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 200031, China

<sup>2</sup>Institute of Biochemistry, Faculty of Biology and Chemistry, Justus Liebig University of Giessen, 35392 Giessen, Germany

<sup>3</sup>Department of Biology and Biotechnologies “Charles Darwin” and IIT Center for Life Nano- & Neuro-Science@Sapienza, Sapienza University of Rome, Italy

<sup>4</sup>Center for Personal Dynamic Regulomes, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA

<sup>5</sup>Integrative Program in Biological and Genome Sciences, University of North Carolina, Chapel Hill, NC 27599, USA

<sup>6</sup>Laboratory of Genetics and Genomics, National Institute on Aging Intramural Research Program, National Institutes of Health, Baltimore, MD 21224, USA

<sup>7</sup>Interdisciplinary Nanoscience Center (iNANO), Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

<sup>8</sup>Department of Molecular Biology and Genetics, Interdisciplinary Nanoscience Centre (iNANO), Aarhus University, Aarhus, Denmark

<sup>9</sup>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 201102, China

<sup>10</sup>Temasek Life Sciences Laboratory, 1 Research Link National University of Singapore, Singapore 117604, Singapore

<sup>11</sup>Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Hemholtz Association, Berlin 13125, Germany

<sup>12</sup>Department of Biochemistry, Stanford University, Stanford, CA 94305, USA

<sup>13</sup>Verna and Marris McLean Department of Biochemistry and Molecular Biology, Therapeutic Innovation Center, Baylor College of Medicine, Houston, TX 77030, USA

\*Correspondence should be addressed to linglingchen@sibcb.ac.cn (L.-L.C.); jeremy.wilusz@bcm.edu (J.E.W.); liyang\_fudan@fudan.edu.cn (L.Y.).

<sup>14</sup>Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100088, China

## Abstract

Alternative splicing of eukaryotic transcripts often leads to production of multiple mature RNAs from a single gene locus. In addition to encoding linear RNAs, genes can produce stable circular RNAs (circRNAs) that are often co-expressed with their cognate linear RNAs. Multiple distinct circRNAs are frequently generated from a gene locus via back-splicing, with each mature transcript having a potentially unique function due to its distinct combination of exons and sometimes retained introns. However, names currently given to circRNAs are often ambiguous and lack consistency across studies. Here, we call on the community to embrace standards for naming circRNAs so that a common nomenclature is used to ensure clarity and reproducibility.

## Keywords

Alternative splicing; back-splicing; back-splice junction; circRNA; ciRNA; nomenclature

## Circular RNAs are produced from many eukaryotic precursor RNAs

Splicing of RNA polymerase II-transcribed precursor RNAs (pre-RNAs) by the spliceosome is critical for the production of mature, functional RNAs. Exons are sequentially joined to one another in a linear fashion, whereas introns are removed as lariats that are rapidly debranched and degraded (Figure 1a, top). It has long been recognized that alternative splicing events can generate multiple mature RNAs from a gene locus, thereby expanding the diversity of functions encoded by a single gene<sup>1</sup>. Besides being spliced to generate a linear RNA, a pre-RNA can be subjected to “back-splicing” when a downstream 5' splice site (ss) is covalently linked with an upstream 3' ss to form a circular RNA (circRNA) (Figure 1a, bottom)<sup>2-5</sup>. 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 (BSJ). In addition, cells can produce a distinct set of circRNAs from introns when an intron fails to be debranched and instead accumulates (Figure 1a, top)<sup>6,7</sup>. These are known as circular intronic RNAs (ciRNAs) or stable intronic sequence RNAs (sisRNAs). While circRNAs generated by back-splicing are joined by a 3',5'-phosphodiester bond, ciRNAs have a 2',5'-phosphodiester bond.

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

Most mature circRNAs are expressed at low levels, but certain genes express circRNAs 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 formation of functional ribonucleoprotein complexes, or to encode peptides/proteins that are translated, e.g. in response to certain cellular stresses. In addition, circRNAs impact many physiological processes including early development, immune responses, neurogenesis and tumorigenesis (see reviews<sup>4,5</sup>).

## Current dilemmas for naming circRNAs

Among all annotated circRNAs generated by back-splicing, perhaps the most well-studied transcript is derived from a long ncRNA (LINC00632)<sup>8</sup> and is known by two distinct names: *CDRIas* and *ciRS-7*<sup>9-11</sup>. This circRNA originates from a region antisense to the cerebellar degeneration-related protein 1 (*CDRI*) gene, hence the name *CDRIas*. It functions as a circular RNA sponge for miRNA-7, hence the name *ciRS-7*. The vast majority of other reported circRNAs have been named by adding the prefix “circ” in front of their host genes’ names, e.g. *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 enable differentiation among multiple circRNAs produced from a host gene. For example, in human HEK293/293FT cells, at least five distinct circRNAs are detected from the *FAM120A* locus in short-read<sup>12</sup> and long-read<sup>13</sup> sequencing data, with each mature circRNA containing a different combination of exons (Figure 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 (Figure 1b, top). In contrast, long-read sequencing can provide direct detection of all included exons (Figure 1b, bottom). One of these circRNAs (the one containing exons 2, 3, 4, and 5) is significantly more abundant than the others and binds the translation inhibitor IGF2BP2 on ribosomes<sup>12</sup>. The simple term “*circFAM120A*” is thus not sufficient to distinguish all five circRNAs from one another, let alone indicate which one is the isoform with a characterized function.

A clear naming system for circRNAs is thus necessary, especially one that allows both the host gene as well as the exact exons and introns present in the mature circRNA to be easily known. The naming of RNA circles is nonetheless currently beyond the scope of the HUGO Gene Nomenclature Committee (HGNC)<sup>14</sup>. Of note, many circRNAs have been archived by publicly available databases, such as CIRCpedia (<http://yang-laboratory.com/circpedia/>)<sup>15</sup>, circAtlas (<http://159.226.67.237:8080/new/index.php>)<sup>16</sup>, circBank (<http://www.circbank.cn>)<sup>17</sup> and circBase (<http://circbase.org>)<sup>18</sup>, and can be retrieved by database entry. However, different strategies were used to name circRNAs in distinct databases and the resulting names often are not easily recalled. For example, the functional *FAM120A* circRNA (Figure 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 circular RNAs 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 exon component numbers (as well as 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<sup>1</sup>. Here, one GENCODE/Ensembl transcript per gene locus is selected as the reference and, when possible, we suggest the use of standardized reference transcripts, e.g. MANE (Matched Annotation from NCBI and EMBL-EBI) transcripts<sup>19</sup>.

To illustrate these proposed naming standards, we provide examples of different types of circRNAs in the following sections. Beyond employing these names in future published studies, it is imperative that the linear reference transcripts used for all circRNA names are also provided, e.g., 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 of Fig. 1b, for example) and genome coordinates (along with assembly) should be provided. Considering that not all transcript data will necessarily be available in perpetuity, providing the actual sequence of the circRNA will help 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 due to alternative BSJ selection: one circRNA consists of exons 2 and 3 (according to the ENST00000322347.11 transcript), while the other consists of exons 3 and 4 (Figure 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 (Figure 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 illustrate the position of the BSJ: the end of the last listed exon was back-spliced to the beginning of the first listed exon. In the case of *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)* (Figure 1b).

In situations where only one circRNA from a gene has been found, we still recommend that the exact exon(s) included in the mature circRNA be included in the name. Doing so will ensure consistency across all circRNA names and help avoid ambiguities as additional circRNA isoforms may be found in the future.

### CircRNAs with the same BSJ, but with different internal splicing patterns

Genes can generate multiple circRNAs that have distinct BSJ sites (Figures 1b, c), but can also generate circRNAs that share the same BSJ yet differ in their internal sequences<sup>4</sup>.

This is because all four standard types of alternative splicing events, such as inclusion of a cassette exon (or the converse, exon skipping), intron retention, and use of alternative 5' or 3' splice sites, have been identified in the internal regions of circRNAs<sup>4</sup>.

Inclusion of a cassette exon has been frequently observed in alternatively spliced circRNAs<sup>4</sup>. For example, the human *TIMMDC1* gene generates two circRNAs that share the same BSJ (the end of exon 4 is joined to beginning of exon 2) but differ in their inclusion or exclusion of the cassette exon 3 (Figure 1d). According to the ENST00000401558.7 transcript of the *TIMMDC1* gene, we suggest calling these transcripts *circTIMMDC1(2,3,4)* and *circTIMMDC1(2,4)*, depending on inclusion or exclusion of exon 3, respectively.

For circRNAs having a retained intron (RI), previously referred to as exon-intron circular RNAs<sup>20</sup>, we suggest the same naming scheme as described above but to include the abbreviation “RI” in the circRNA name to indicate the existence of an internal RI. As an example, two circRNAs are produced from the human *CAMSAP1* gene (Figure 1e). Both have the end of exon 3 back-spliced to the beginning of exon 2, but one isoform contains a RI between exons 2 and 3 (directly observed in long read sequencing data, Figure 1e). Hence, we suggest naming these transcripts *circCAMSAP1(2,3)* and *circCAMSAP1(2,RI,3)* (according to the ENST00000389532.9 transcript annotation) to indicate the presence of the retained intron only in the latter circRNA.

For circRNAs that contain 5' or 3' alternatively spliced exons, we suggest to add an “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' ss selections (Figure 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 usage of an alternative 3' ss that results in a longer exon 3. In the case of alternative 5' ss selections within circRNAs, the “S” or “L” labeling 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' ss selections for the internal exon 4 (Figure 1g). Use of the names *circSNX25(3,4S,5)* and *circSNX25(3,4,5)* can distinguish these two circRNA isoforms that have either a short version of exon 4 or full-length exon 4, respectively, according to the ENST00000618785.4 transcript annotation. In rare cases where more than two alternative splice sites are used, consecutive letters (e.g. A, B, C...) 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<sup>4, 21, 22</sup>. In these cases, we suggest to mark such a sequence as a “new exon” (NE) when naming the circRNA. As shown in Figure 1h, the human *ZNF292* gene generates a three-exon circRNA (*circZNF292(2,3,4)*) according to the ENST00000369578.6 transcript of the *ZNF292* gene) as well as an additional circRNA that contains a previously unannotated exon upstream of exon 2. Besides not being present in the ENST00000369578.6 transcript, this NE is not annotated in any assembled *ZNF292*

transcript in GENCODE V37. Hence, we suggest this circRNA with the NE to be named *circZNF292(NE,2,3,4)*.

As an additional example, the human *XPO1* gene generates two circRNAs with the same BSJ, but they differ by the presence of an internal novel cassette exon (Figure 1i). According to the ENST00000401558.7 transcript of the *XPO1* gene, we suggest naming these two circRNAs, *circXPO1(3,4)* and *circXPO1(3,NE,4)* to indicate whether the novel exon is skipped or included, respectively. If multiple unannotated exons are present in a circRNA, the NE notation should be used multiple times to indicate each of their positions. We recognize that the NE label is not ideal and should be viewed as a temporary placeholder. Further annotation efforts of the transcriptome, especially incorporation of circRNAs into databases such as GENCODE and RefSeq, should allow more unambiguous names to be given in the future. In 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 with chromosomal translocations by gene fusion<sup>23</sup>. 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 leukemic cell lines<sup>24</sup> (Figure 1j). Accordingly, we suggest the names of these circular transcripts to be *circKMT2A(5,6,7,8)::MLLT3(6)* and *circKMT2A(7,8)::MLLT3(6)*, respectively, where both host gene names are included but separated by a double colon to designate a gene fusion (as recommended in Ref. 23). This allows one to distinguish between these transcripts and infer that exon 6, according to the ENST00000380338.9 transcript, of *MLLT3 (AF9)* has been back-spliced to exon 5 or 7, according to the ENST00000534358.8 transcript, of *KMT2A (MLL)*, respectively.

### Conclusions

Given the wide variety of circular transcripts observed in eukaryotic cells and the fact that multiple distinct RNA circles can be derived from a single host gene, it is becoming increasingly clear that a systematic naming scheme for circular RNAs is needed. To limit ambiguity, we recommend the names of circRNAs produced from back-splicing to include the prefix “circ” followed by the host gene symbol as well as exon (and, if present, intron) information. An analogous strategy can 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 linear transcript annotations. The proposed naming scheme thus represents an interim solution until circular RNAs are incorporated into annotation databases such as GENCODE and RefSeq. Such a step is the only true long-term solution as it will remove ambiguity in circRNA names (e.g., by enabling removal of NE labels) as well as eliminate the need to have a reference linear transcript. In the meantime, it is critical that the reference transcripts used for all circular RNA names be provided in publications.

We call on authors, reviewers, and editorial staff to ensure that manuscripts that report circRNAs provide (i) the genomic coordinates and associated genome assembly, (ii) the full sequences of circular RNAs, (iii) any previously given names, e.g., in circRNA databases, as well as (iv) clear diagrams that depict where circular RNAs of interest map in the genome. This will greatly help facilitate communication and eliminate confusion when multiple circular RNAs are produced from the same gene locus. More and more circular RNAs continue to be identified, including some presumably from intron self-ligation<sup>21</sup> as well as some derived from noncoding RNAs, including tRNAs, rRNAs, and mitochondrial RNAs (see reviews<sup>4, 25</sup>). As the field progresses, additional naming criteria can be proposed for these novel transcript classes.

## References

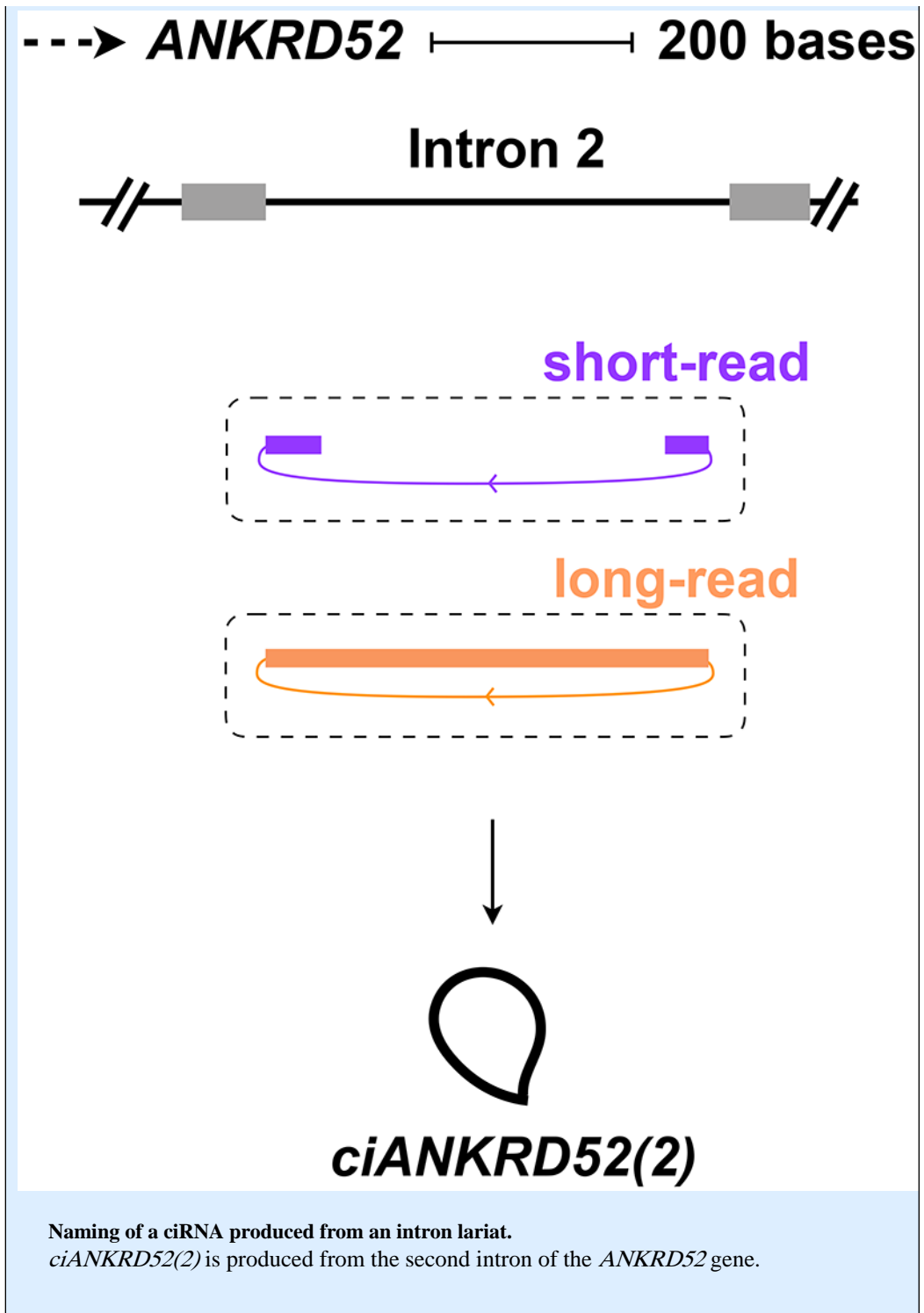
1. Nilsen TW & Graveley BR Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463, doi:10.1038/nature08909 (2010). [PubMed: 20110989]
2. Szabo L & Salzman J Detecting circular RNAs: bioinformatic and experimental challenges. *Nat Rev Genet* 17, 679–692, doi:10.1038/nrg.2016.114 (2016). [PubMed: 27739534]
3. Kristensen LS et al. The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* 20, 675–691, doi:10.1038/s41576-019-0158-7 (2019). [PubMed: 31395983]
4. Yang L, Wilusz JE & Chen LL Biogenesis and Regulatory Roles of Circular RNAs. *Annu Rev Cell Dev Biol*, doi:10.1146/annurev-cellbio-120420-125117 (2022).
5. Liu CX & Chen LL Circular RNAs: Characterization, cellular roles, and applications. *Cell* 185, 2016–2034, doi:10.1016/j.cell.2022.04.021 (2022). [PubMed: 35584701]
6. Li X et al. Linking circular intronic RNA degradation and function in transcription by RNase H1. *Sci China Life Sci* 64, 1795–1809, doi:10.1007/s11427-021-1993-6 (2021). [PubMed: 34453665]
7. Tay ML & Pek JW Maternally Inherited Stable Intronic Sequence RNA Triggers a Self-Reinforcing Feedback Loop during Development. *Curr Biol* 27, 1062–1067, doi:10.1016/j.cub.2017.02.040 (2017). [PubMed: 28343963]
8. Barrett SP, Parker KR, Horn C, Mata M & Salzman J ciRS-7 exonic sequence is embedded in a long non-coding RNA locus. *PLoS Genet* 13, e1007114, doi:10.1371/journal.pgen.1007114 (2017). [PubMed: 29236709]
9. Memczak S et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338, doi:10.1038/nature11928 (2013). [PubMed: 23446348]
10. Hansen TB et al. Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388, doi:10.1038/nature11993 (2013). [PubMed: 23446346]
11. Hansen TB et al. miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J* 30, 4414–4422, doi:10.1038/emboj.2011.359 (2011). [PubMed: 21964070]
12. Li S et al. Screening for functional circular RNAs using the CRISPR-Cas13 system. *Nat Methods* 18, 51–59, doi:10.1038/s41592-020-01011-4 (2021). [PubMed: 33288960]
13. Xin R et al. isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. *Nat Commun* 12, 266, doi:10.1038/s41467-020-20459-8 (2021). [PubMed: 33436621]
14. Seal RL et al. A guide to naming human non-coding RNA genes. *EMBO J* 39, e103777, doi:10.15252/emboj.2019103777 (2020). [PubMed: 32090359]
15. Dong R, Ma XK, Li GW & Yang L CIRCpedia v2: An Updated Database for Comprehensive Circular RNA Annotation and Expression Comparison. *Genomics Proteomics Bioinformatics* 16, 226–233, doi:10.1016/j.gpb.2018.08.001 (2018). [PubMed: 30172046]
16. Wu W, Ji P & Zhao F CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol* 21, 101, doi:10.1186/s13059-020-02018-y (2020). [PubMed: 32345360]

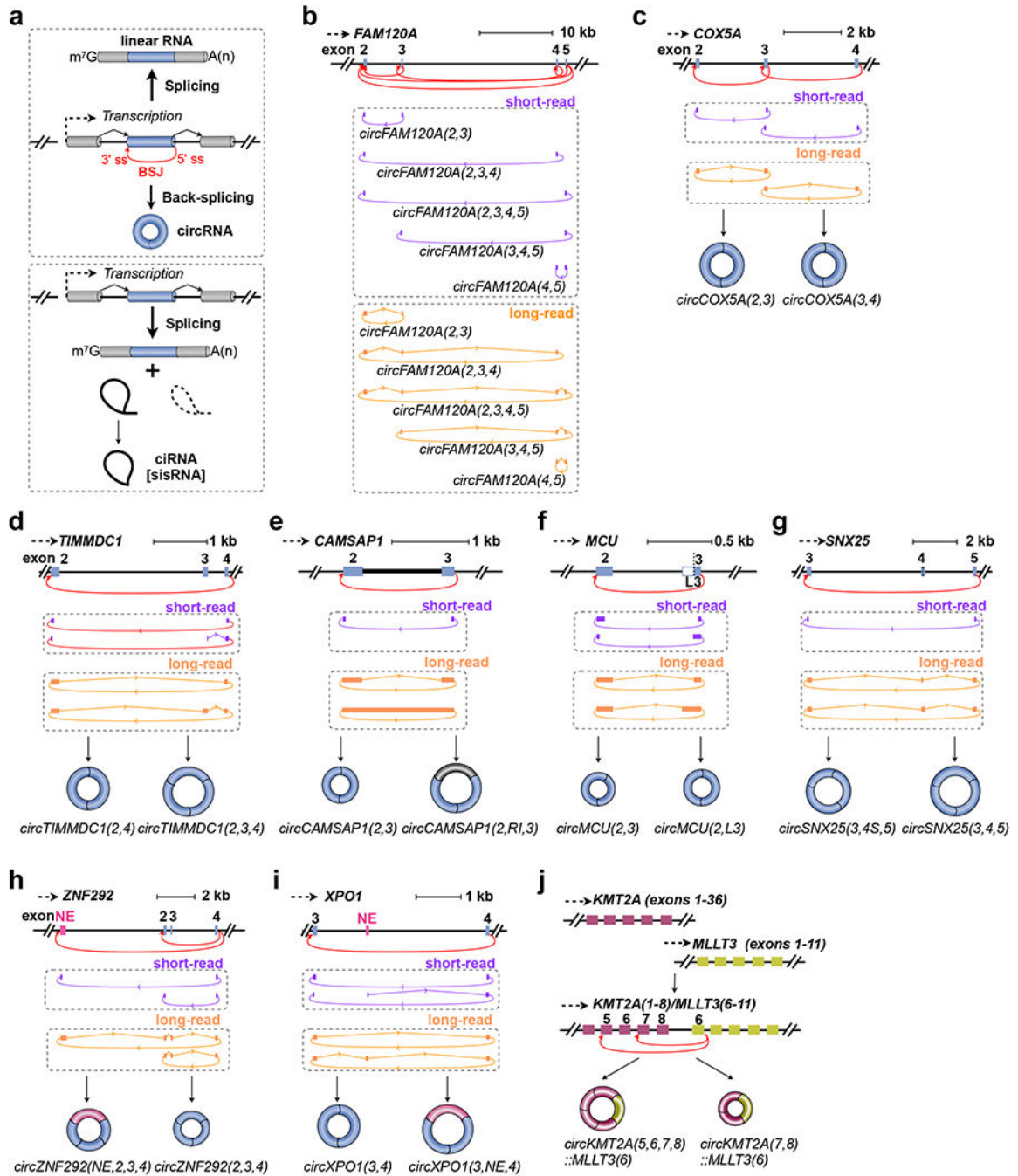
17. Liu M, Wang Q, Shen J, Yang BB & Ding X Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biol* 16, 899–905, doi:10.1080/15476286.2019.1600395 (2019). [PubMed: 31023147]
18. Glazar P, Papavasileiou P & Rajewsky N circBase: a database for circular RNAs. *RNA* 20, 1666–1670, doi:10.1261/rna.043687.113 (2014). [PubMed: 25234927]
19. Morales J et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature* 604, 310–315, doi:10.1038/s41586-022-04558-8 (2022). [PubMed: 35388217]
20. Li Z et al. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* 22, 256–264, doi:10.1038/nsmb.2959 (2015). [PubMed: 25664725]
21. Zhang J et al. Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. *Nat Biotechnol* 39, 836–845, doi:10.1038/s41587-021-00842-6 (2021). [PubMed: 33707777]
22. Rahimi K, Veno MT, Dupont DM & Kjems J Nanopore sequencing of brain-derived full-length circRNAs reveals circRNA-specific exon usage, intron retention and microexons. *Nat Commun* 12, 4825, doi:10.1038/s41467-021-24975-z (2021). [PubMed: 34376658]
23. Bruford EA et al. HUGO Gene Nomenclature Committee (HGNC) recommendations for the designation of gene fusions. *Leukemia* 35, 3040–3043, doi:10.1038/s41375-021-01436-6 (2021). [PubMed: 34615987]
24. Guarnerio J et al. Oncogenic Role of Fusion-circRNAs Derived from Cancer-Associated Chromosomal Translocations. *Cell* 165, 289–302, doi:10.1016/j.cell.2016.03.020 (2016). [PubMed: 27040497]
25. Lasda E & Parker R Circular RNAs: diversity of form and function. *RNA* 20, 1829–1842, doi:10.1261/rna.047126.114 (2014). [PubMed: 25404635]



**Box 1:****Suggested guidelines for ciRNA naming.**

Analogous to the 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 (Figure 1a, top). 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]<sup>6</sup> (Figure in Box 1). In general, ciRNAs and sisRNAs<sup>6,7</sup> 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 this simple naming system can easily distinguish ciRNA isoforms from one another.





**Figure 1. Suggested naming scheme to clarify the complexity of circular RNA expression patterns.**

a. (Top) Pre-RNAs can be spliced to generate mature linear RNAs that are capped and polyadenylated as well as intron lariats that are debranched and degraded (denoted by dashed lines). However, some lariats fail to be debranched leading to the accumulation of stable ciRNAs. (Bottom) In addition to generating linear RNAs, a gene can generate circRNAs when a 5' splice site (ss) is joined to an upstream 3' ss 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.

- b. Naming of human circRNAs produced from the gene *FAM120A* (ENST00000277165.11) that have different BSJ sites.
- c. Naming of circRNAs produced from the gene *COX5A* (ENST00000322347.11) that have different BSJ sites.
- d. Naming of two alternative circRNAs produced from the gene *TIMMDC1* (ENST00000494664.6) that differ in their inclusion of a cassette exon.
- e. Naming of circRNAs produced from the gene *CAMSAP1* (ENST00000389532.9) that differ in their inclusion of a retained intron.
- f. Naming of circRNAs produced from the gene *MCU* (ENST00000603649.5) that differ in their use of an alternative 3' ss. 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* gene locus.
- g. Naming of circRNAs produced from the gene *SNX25* (ENST00000618785.4) that differ in their use of an alternative 5' ss.
- h. Naming of circRNAs produced from the gene *ZNF292* (ENST00000369578.6) that differ in their BSJ and presence of a novel exon.
- i. Naming of circRNAs produced from the gene *XPO1* (ENST00000401558.7) 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 leukemic cell lines.
- Gene assembly: bars, exons; black lines, introns. Canonical splice junctions (SJs, polylines) and back-splice junctions (BSJs, arc lines) are shown in sequencing data. Short-read (purple) and long-read (orange) examples 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/<br>hg38;<br>GRCh37/<br>hg19 | chr9:93471140-93498886<br>(hg38);<br>chr9:96233422-96261168<br>(hg19) | <i>FAM120A</i>     | 2,3,4,5         | GENCODE<br>+ RefSeq |
| circAtlas | hsa-FAM120A_0006        | GRCh38/<br>hg38                     | chr9:93471141-93498886<br>(hg38)                                      | ENSG00000048828.16 | N/A             | GENCODE             |
| circBank  | hsa_circFAM120A_007     | GRCh37/<br>hg19                     | chr9:96233422-96261168<br>(hg19)                                      | <i>FAM120A</i>     | N/A             | RefSeq              |
| circBase  | hsa_circ_0001875        | GRCh37/<br>hg19                     | chr9: 96233422-96261168<br>(hg19)                                     | <i>FAM120A</i>     | 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 start genomic location of a circRNA will differ by one base depending on the database.