Winter 1-15-2021

Discovery and Characterization of Circular and Chimeric RNAs in Complex Diseases

Xiaoxin Liu
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Computer Science and Technology
Computer Science

Dissertation Examination Committee:
Weixiong Zhang, Chair
Sharlee Climer
Tao Ju
Brendan Juba
Paul Shaw

Discovery and Characterization of Circular and Chimeric RNAs in Complex Diseases

by

Xiaoxin Liu

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

January 2021
St. Louis, Missouri

Table of Contents

List of Figures ..... *iv*

List of Tables ..... *v*

Acknowledgments ..... *vi*

Abstract ..... *vii*

Chapter 1. Introduction *1*

1.1 Circular RNA **1**

1.2 Chimeric RNA **3**

1.3 Hypothesis and research objectives **5**

Chapter 2. Discovery of all types of circular and chimeric RNAs *7*

2.1 Identification method **7**

2.1.1 Identification of circRNAs of all types 7

2.1.2 Performance of CAT for circRNA identification 10

2.1.3 Identification of chimeric RNAs of all types 12

2.2 Interior circRNAs are ubiquitous in eukaryotes **15**

2.3 Interior circRNA predominantly expressed in nascent RNA **18**

Chapter 3. Characteristics of circular and chimeric RNA *20*

3.1 Characteristics of circRNA **20**

3.1.1 Short homologous sequences associated with circRNAs 20

3.1.2 Complementary sequences flanking circRNAs 21

3.1.3 More than one interior circRNA from the same genomic locus 22

3.2 Characteristics of chimeric RNAs **24**

3.2.1 Short homologous sequences associated with chimeric RNAs 24

3.2.2 More than one chimeric RNA from the same genomic locus 24

Chapter 4. Functions of circular and chimeric RNAs *26*

4.1 circRNAs function as competing endogenous RNAs in psoriatic skin **26**

4.1.1 circRNAs in psoriatic skin 27

4.1.2 Aberrantly expressed circRNAs in PS 29

4.1.3 Putative circRNA functions in PS 32

4.2 Potential functions of chimeric RNAs in prostate cancer and psoriasis **35**

Chapter 5. Biogenesis models and summary *40*

5.1 A splicesome-independent mechanism for circular and chimeric RNA production **40**

5.2 Summary **43**

References ***46***

Appendix ***53***

List of Figures

Figure 1: Mechanisms of circular RNA and chimeric RNA formation …….3

Figure 2: Two key elements for identification of circRNAs and classification of circRNAs.……8

Figure 3: Specificity and sensitivity of circRNA identification methods …….……12

Figure 4: Identification and classification of chimeric RNAs ……..….13

Figure 5: Distribution and validation of circRNAs ……..….17

Figure 6: Distribution of circRNAs in nascent RNA ……..…..18

Figure 7: Sequence and structural features characteristic of i-circRNAs ……..……21

Figure 8: Multiple i-circRNAs from the same genomic locus ……23

Figure 9: Sequence and structural characteristics of chimeric RNAs ……..25

Figure 10: Distribution and characteristics of abundant circRNAs in PS …….28

Figure 11: Experimental validation of three identified circRNAs in psoriatic skin and HaCaT keratinocyte cells ……..……31

Figure 12: Putative functions of three circRNAs as ceRNAs ……34

Figure 13: Pathways enriched by host genes of chimeric RNAs ……..36

Figure 14: Differentially expressed chimeric RNAs ……..…37

Figure 15: Pathways enriched by host genes of DE chimeric RNAs …….39

Figure 16: Mechanisms of circular and chimeric RNAs ……42

List of Tables

Table 1: circRNA identification algorithms ……..10

Table 2: Chimeric RNA identification algorithms ……14

Table 3: Descriptions of four datasets of stranded RNA-seq used for chimeric RNA identification ……...……14

Table 4: Descriptions of five collections of stranded RNA-seq data ……16

Acknowledgments

I have received a great deal of support and assistance throughout the writing of this dissertation.

Firstly, I would like to express my sincere gratitude to my advisor Professor Weixiong Zhang for the continuous support of my Ph.D. research. His motivation, immense knowledge, and patience encouraged and pushed me to sharpen my thinking and brought my work to a higher level. He was always there when I had questions and needed a discussion. His guidance helped me in all the time of my research and writing of this thesis.

I would also like to thank the rest of my thesis committee – Professor Sharlee Climer, Professor Tao Ju, Professor Brendan Juba, and Professor Paul Shaw – for their insightful comments and encouragement. I also thank Professor Jeremy Buhler and Professor Sanmay Das for their suggestions in formulating the research questions.

I would like to thank my collaborators: Professor Anne Bowcock, Professor Zhangfeng Hu, and Dr. Jacqueline Frost, for their stimulating discussions and supportive experiments.

I gratefully acknowledge the assistance of all the staff members in the Department of Computer Science and Engineering. They are always nice and helpful and have provided a positive and pleasant work environment.

Many thanks to my fellow labmates for sharing their expertise and thoughts with me: Dr. Jing Xia, Dr. Zheng Chen and Zhuangzhuang Zhang. I would also like to thank my friends who went through hard times together with me and believed in me all the time. Finally, I would like to thank my family, especially my parents and my sister, for always supporting and loving me.

Xiaoxin Liu

Washington University in St. Louis

January 2021

Abstract

Discovery and Characterization of Circular and Chimeric RNAs in Complex Diseases

by Xiaoxin Liu

Doctor of Philosophy in Computer Science

Washington University in St. Louis, 2021

Professor Weixiong Zhang, Chair

Non-co-linear RNAs, including circular and chimeric RNAs, have been investigated in a broad range of cellular processes and various diseases. The current biogenesis models of circular and chimeric RNAs depend on RNA splicing mechanisms. We hypothesized that some of the non-co-linear RNAs are products of an alternative mechanism other than RNA splicing. To test this hypothesis, we developed a method, which is independent of genome annotations and splicing signals, to identify Circular and Chimeric RNAs of All Types (C^2AT or CAT for simplicity) by analyzing the RNA species profiled by whole genome RNA sequencing. Statistical analysis and experimental validation revealed that both circular RNAs (circRNAs) and chimeric RNAs may arise from the interior regions of exons and introns across the human genome, supporting a potential novel mechanism of biogenesis of circular and chimeric RNAs. Moreover, analysis of differential expressions of circRNAs and chimeric RNAs and their host genes suggested that these non-co-linear RNAs may play an important role in complex diseases, such as cancer and psoriasis. All the results from this systematic study of non-co-linear RNAs significantly expanded our understanding of the origin, diversity, and complexity of circRNAs and chimeric RNAs as well as their potential functions in complex diseases.

Chapter 1. Introduction

The advent and development of high-throughput RNA sequencing technology have revolutionized the research on linear RNAs and non-co-linear RNAs, which consist of two or more RNA segments that are topologically inconsistent with their DNA counterparts in the reference genome. While linear RNAs, e.g., messenger RNAs (mRNAs) and long noncoding RNAs, have been studied for decades, non-co-linear RNAs, mainly circular and chimeric RNAs, are a relatively less scrutinized class, which are more complex in their structures and biogenesis. Chimeric RNAs are composed of sequence segments from two or more primary RNA transcripts, whereas circular RNAs (circRNAs) are formed in a closed circle structure by linking the 5' and 3' ends of RNA molecules.

1.1 Circular RNA

Circular RNA (circRNA) is a class of non-coding RNA, which appears in a single-stranded, covalently closed circular form [1, 2]. As first reported in RNA viruses [3] and eukaryotes [2] decades ago, circRNAs were thought to be a result of erroneous splicing of transcripts for a long time. With the development of RNA sequencing technology (RNA-seq), many genome-wide sequencing-based RNA profilings have revealed that circRNAs exist broadly in animals [4-7], plants [8, 9], and nearly all eukaryotic species [10].

Functional analysis has suggested many molecular functions of circRNAs, including promoting transcription [11], competing with canonical splicing [12], and influencing the expression of mRNA [13] and protein [12]; some circRNAs may even be translated into polypeptide [14, 15]. circRNAs exhibit expression patterns that are specific to cell types and tissues [16-21] as well as development stages [17, 22]. Some circRNAs have been shown to have aberrant expressions in complex diseases, such as cancer [23-25], osteoarthritis [26], cardiovascular diseases [27, 28], and

Alzheimer's disease (AD) [29], and may consequently play a role in disease onset and progression. The most well-studied circRNA *CDR1as* [7], also known as *ciRS-7* [6], originates from the noncoding transcript antisense to a gene encoding Cerebellar Degeneration-Related protein 1 (*CDR1*) on human chromosome X. It carries more than 70 binding sites of microRNA (miRNA) *miR-7*, and thus acts as a sponge of *miR-7* [6, 7]. *miR-7* is a "master" regulator of many signaling pathways via mediating other regulating factors, including epidermal growth factor receptor (*EGFR*), insulin receptor substrate-1 (*IRS-1*), *IRS-2*, p21-activated kinase-1 (*Pak1*), and *Raf1* [23, 25, 30], to name a few. *CDR1as* is enriched in human and mouse brains and neurons [6, 7, 17] and thus expected to be involved in neurodegenerative diseases through the *CDR1as*-*miR-7*-targets gene regulatory cascade.

Most of the reported circRNAs contain exclusively exons or exons and introns [31-33], and these circRNAs join at intron-exon junctions, suggesting that they are products of the splicing apparatus [34-39] (Figure 1A). Indeed, the formation of some of these circRNAs has been shown to require canonical splicing signals [40] and to involve exon skipping [41, 42]. Furthermore, it has been revealed that trans-acting RNA-binding proteins and reverse complementary repeat sequences, e.g., Alu elements in human, in adjacent introns help bring the 3'-end of an exon and the 5'-end of the exon or an upstream exon to close vicinity in the cell and subsequently join the two ends of the linear transcript to help form a circRNA [43-47].

Intergenic regions have been indicated to host circRNAs [35]. Yet most recent studies have focused on circRNAs originated from intron-exon junctions. Particularly, the existing circRNA identification methods zoom onto genomic loci of intron-exon boundaries and capitalize on the annotated splicing signals of splicing donor and receptor sites on the genome [48-50]. However, it

remains unresolved if intron-exon junctions are the only genomic origin for circRNA production and if splicing is the only mechanism underlying circRNA biogenesis.

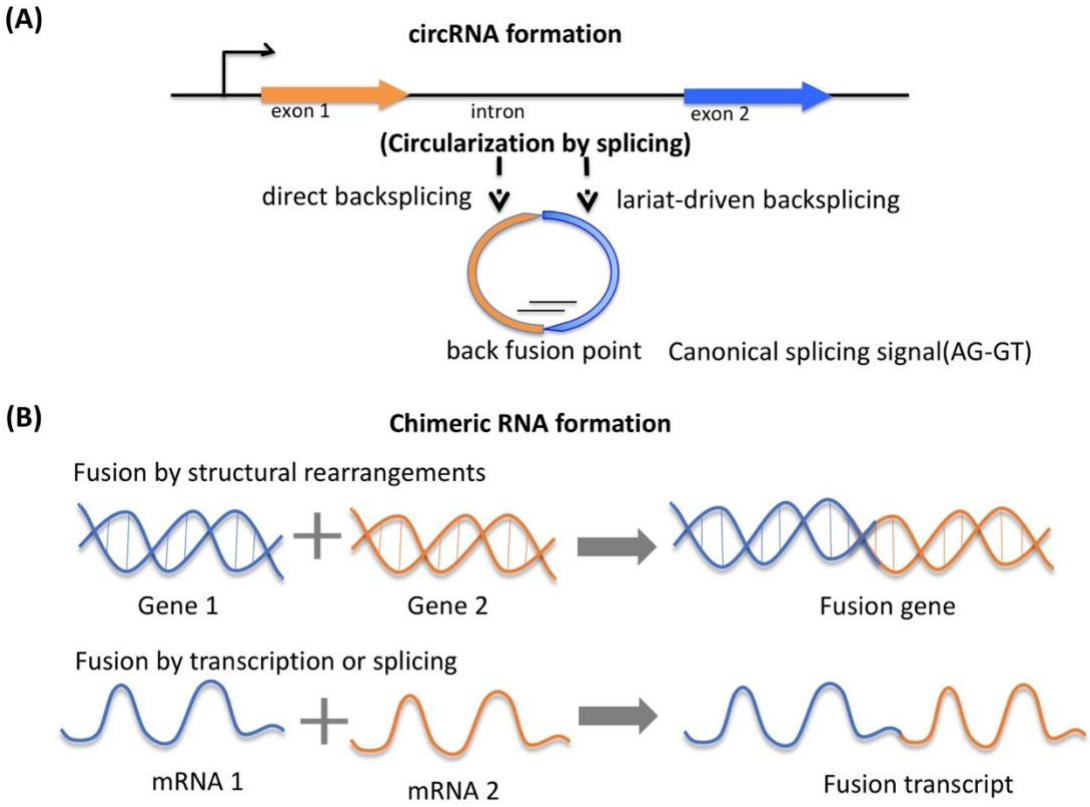


Figure 1. Mechanisms of circular RNA and chimeric RNA formation. **(A)** Direct backsplicing and lariat-driven backsplicing can lead to the formation of circRNAs, and splicing signals, such as AG-GT, are always involved in the process of circularization. **(B)** Chimeric RNAs can be formed by both structural rearrangements and non-structural rearrangements. Structural rearrangements of chromosomes include transforms such as insertion, deletion, translocations and inversions. Non-structural rearrangements may happen during the process of transcription, i.e. transcription read-through of neighboring genes, or after the process of transcription, i.e. splicing of mRNA molecules.

1.2 Chimeric RNA

Chimeric RNA, also known as fusion transcript, is composed of sequence segments from two or more different transcripts (Figure 1B). The first described chimeric RNA was *BCR/ABL1* in cancer cells in the early 1980s, which was generated by chromosomal rearrangement between chromosome 9 and chromosome 22 [51]. Since then, chimeric RNAs have been discovered broadly in almost all types of human neoplasia and even in lower eukaryotes such as nematodes and

trypanosomes [52, 53]. Recently, advances in next-generation sequencing have helped identify more putative chimeric RNAs. For instance, the ChimerDB database has maintained more than 60,000 fusion transcripts in various types of cancer as well as normal tissues of humans [54].

The most well-known function of chimeric RNAs is to encode novel chimeric proteins [55]. Many chimeric RNAs and their corresponding protein products are associated with oncogenic properties and often act as driver genes in various cancers. For example, chimeric RNA *SLC45A3-ELK4* and *TMPRSS2-ERG* are promising biomarkers for prostate cancer [56]; *BCR-ABL*, *FUS-ERG*, *MLL-AF6*, and *MOZ-CBP* are expressed in acute myeloid leukemia (AML) [57, 58]; *EML4–ALK* may function as a therapeutic target and a diagnostic molecular marker in non-small-cell lung cancer (NSCLC) [59]. In addition to cancers, some chimeric RNAs are also found to be related to some other diseases. Chimeric RNA *CFHR5–CFHR2* was identified to cause familial C3 glomerulonephritis [60]. *TEL-Syk* was reported to be expressed in patients with myelodysplastic syndrome and involved in the pathogenesis of hematopoietic malignancies [61]. A comprehensive genetic study of two siblings affected by Autism spectrum disorder (ASD) and their unaffected parents suggested that chimeric transcript *ELMOD3-SH2D6* may play a role in ASD, together with other well-established variants such as copy number variants [62]. However, the research on chimeric RNAs in complex diseases remains to be relatively rare compared to that on various types of cancers.

Most chimeric RNAs can be translated into proteins. However, fusion proteins are often expressed with low abundance and tissue specific [63]. Despite that a large number of chimeric transcripts have been reported, only a few fusion proteins have been confirmed so far. Fusion transcripts and proteins are not only expressed in cancers but also present in normal tissues with unclear functions [64].

Two biogenesis models have been established for chimeric RNA formation, i.e., structural rearrangement in DNA and transcriptional rearrangement in RNA [63] (Figure 1B). In the model of structural rearrangement, the structures of chromosomes may change in different ways, such as translocations, inversions, deletions, and insertions, leading to the formation of fusion genes. These fusion genes may then be transcribed into chimeric RNAs and further translated into fusion proteins. In transcriptional rearrangement, trans-splicing or cis-splicing of RNAs may occur between different mRNA molecules and transcription read-through may happen between adjacent genes, resulting in fusion transcripts that may be translated into fusion proteins. According to these two mechanisms, chimeric RNAs are composed of exons from different genes or mRNAs, and the existing chimeric RNA identification methods focus on genomic loci of intron-exon boundaries and depend on the annotated exons, introns, and intergenic regions. However, it remains an open question if intron-exon junctions are the only genomic origin for chimeric RNA generation, and if the above two mechanisms are the only biogenesis mechanisms for chimeric RNA production.

1.3 Hypothesis and research objectives

The existing studies of circular and chimeric RNAs focus exclusively on the candidate transcripts originating from canonical intron-exon junctions. Inspired by the result that intergenic regions can host circRNAs [35], we challenge this common wisdom and hypothesize that circular and chimeric RNAs can arise in other genomic loci across the genome, particularly from the regions inside of introns and exons.

To validate this hypothesis, we develop an integrative computational algorithm without using any information of splicing signals and genome annotation to search for fusion junctions of circRNAs and chimeric RNAs from any genomic locus which are captured in RNA-seq data. With this algorithm, we carry out systematic studies using several large sets of RNA-seq data collected from

plants and humans. Besides canonical circRNAs and chimeric RNAs originating from intron and exon junctions, we are particularly interested in noncanonical circular and chimeric RNAs that are produced from other regions of the genome, such as interior regions of exons and introns. Some of these novel, noncanonical RNAs are further experimentally validated using Polymerase Chain Reactions (PCR) followed by the Sanger sequencing.

Furthermore, we further hypothesize that the noncanonical circular and chimeric RNAs are functional. This is inspired by the results that canonical circular RNAs can regulate gene transcription by acting as miRNA sponges [65] or competing with canonical splicing [12] and chimeric RNAs may also function as competing endogenous RNAs by retaining miRNA binding sites [66]. Therefore, circular and chimeric RNAs may potentially act as regulators in the cell.

To study the second hypothesis, we adopt systems biology approaches to explore the characteristics, functions, and biogenesis of circular and chimeric RNAs in complex diseases. Statistical methods are applied to search for sequence and structure patterns from the identified circular and chimeric RNAs, especially from the sequences near the fusion points. We perform differential expression analysis and pathway enrichment on circular and chimeric RNAs in complex diseases such as prostate cancer and psoriasis. Moreover, significant differentially expressed (DE) circRNAs and chimeric RNAs, particularly the noncanonical ones, are scanned for potential functions, such as binding miRNAs and competing with mRNAs. Some of these DE RNAs are further experimentally validated using PCR.

Chapter 2. Discovery of all types of circular and chimeric RNAs

Here, we develop a new algorithm to identify circular and chimeric RNAs that arise from inside of exons, introns, and intergenic transcripts without genome annotation and apply the method to analyze these non-co-linear RNAs in humans (Homo sapiens) and Arabidopsis (Arabidopsis thaliana). To distinguish them from the reported circular and chimeric RNAs over intron-exon junctions, we name such novel type as interior circRNAs (i-circRNAs for short) and interior chimeric RNAs.

2.1 Identification method

Many methods have been developed for circRNA identification, but most of them utilize splicing signals or gene annotations to facilitate the discovery of back fusion points (BF points), such as find_circ [7], CIRCexplorer2 [67], ACFS [68] and KNIFE [50]. A new circRNA identification method was developed to search for all possible back fusion points (BF points) that are captured in RNA-seq data without using any information of splicing signals and genome annotation.

2.1.1 Identification of circRNAs of all types

A new method was developed to comprehensively search for circular RNAs of all types (i.e., the CAT method) using RNA-seq data without information of splicing signals. CAT first maps sequencing reads to the reference genome with Bowtie2 [69]. The mappable reads are discarded since they are from linear RNAs. CAT takes a left and a right *x*-mer on the 5'-end and 3'-end of an unmapped read, called the left and right anchors, respectively (Figure 2A). It attempts to *split-map* the read to the genome – the left anchor maps to a locus that is downstream from the locus to which the right anchor maps. If successful, it then searches for the split point within the sequence between the two anchors such that when the read is split at the split point, the two segments of the read can be aligned to the genomic loci that the anchors determine (Figure 2A). In the current CAT

implementation for sequencing read length of 100bp, *x* was initially set to 20, and no more than two mismatches were allowed in mapping.

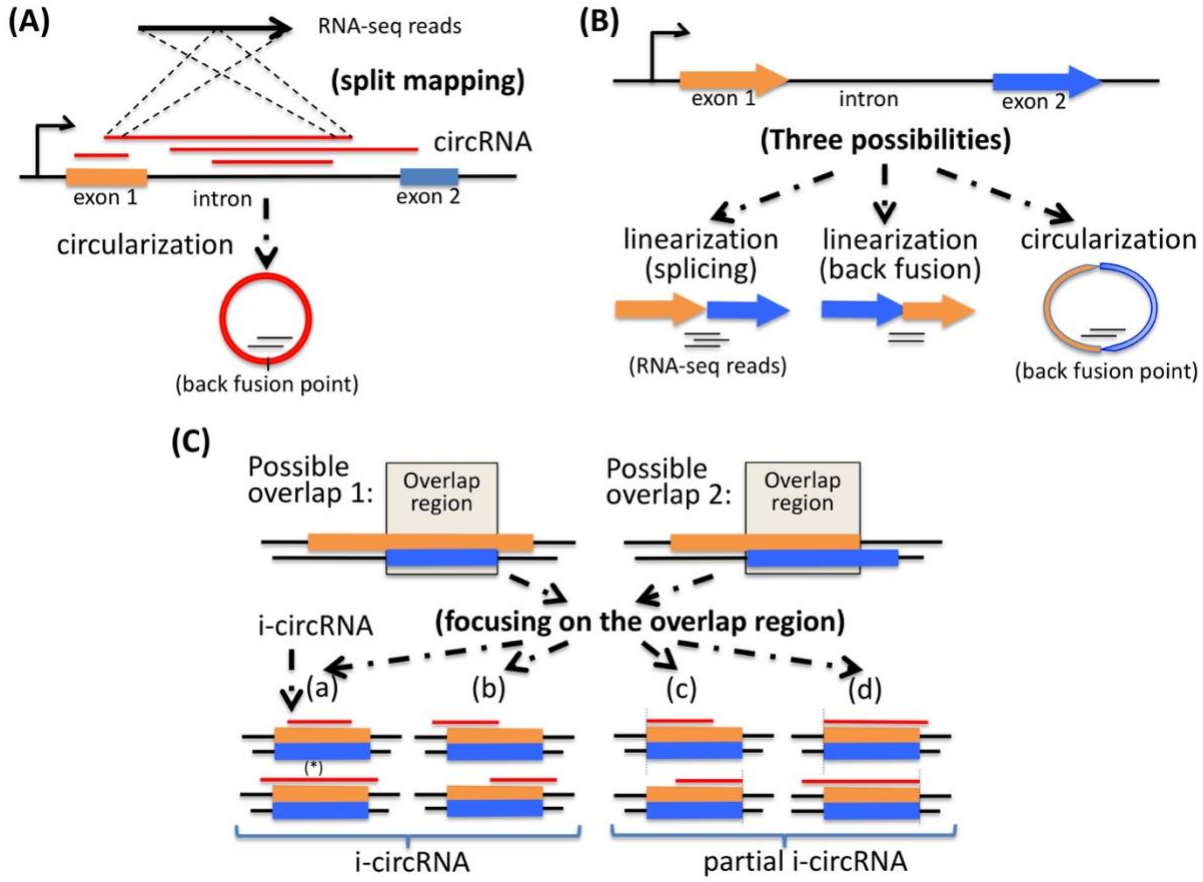


Figure 2. Two key elements for identification of circRNAs and classification of circRNAs. **(A)** The first key element is split mapping of RNA-seq reads to candidate circRNAs, particular i-circRNAs, which may originate from an interior region of an exon, an intron or a pair of adjacent exon and intron. **(B)** The second key element is to avoid ambiguity in distinguishing a canonical circRNA or i-circRNA spanning over more than one exon using RNA-seq data alone, particularly data from an RNA library preparation protocol where circRNAs are not enriched. Shown here are two exons from which two possible linear RNAs, one from splicing and the other from back splicing, and a circular RNA may be produced independently. It is difficult to distinguish the circRNA from the linear RNAs when they appear together using RNA-seq data alone. To this end, the analysis here focuses on candidate circRNAs from one exon, one intron, or an adjacent exon and intron pair. **(C)** Classification of i-circRNAs. An i-circRNA may have four possible positions relative to the intron-exon boundaries of its originating transcript. A more elaborate scenario involves an i-circRNA from locus with more than one gene annotation. Two types of gene annotation may appear, one where an exon encompasses another exon and the other where two exons are partially overlapped. Regardless of which of the two types occurs, it is sufficient to consider only the overlapping region. Based on the relative position of the i-circRNA and the overlapping region, the i-circRNA can be classified into a *complete i-circRNA* (or i-circRNA) or *partial i-circRNA* where one of the back fusion points of the i-circRNA is aligned to one of the boundaries of the overlapping region.

CAT retains those BF points that are supported by at least *k* (e.g., *k*>2 in most of our analyses) split-mapped reads in an RNA-seq library. Furthermore, only unambiguous candidate BF points are considered by setting a threshold for quality score. The quality score is defined as the score difference between the best mapping and the second-best mapping of the anchors, so the larger the better. In our analysis, we usually set the threshold for the quality score as 40, which is the largest possible value for a quality score.

Specifically, while there certainly exist circRNAs spanning more than one exon, it is difficult to infer such circRNAs with certainty based on RNA-seq data alone, particularly on RNA libraries not enriched for circRNAs (Figure 2B). Therefore, only candidate circRNAs from single introns, single exons, pair of adjacent exons and introns, and intergenic non-coding transcripts are considered, including canonical circRNAs over intron-exon junctions and interior circRNAs (i-circRNAs) arising from the interior regions of exons and introns (Figure 2C), if RNA-seq libraries are not RNase R treated.

An i-circRNA may have a few possible positions relative to its originating RNA transcript (Figure 2C). Consider a simple case of a transcript consisting of a single exon. In addition to possibly giving rise to a circRNA by circularization on its 5'-end and 3'-end, it may host an i-circRNA completely interior to the transcript, which is referred to as *complete i-circRNA* or simply i-circRNA. Besides, it may generate an i-circRNA starting from its 5'-end and ending in the middle of the transcript or an i-circRNA starting in the middle and ending at its 3'-end, any of which is referred to as a *partial i-circRNA* (Figure 2C). It is more involved when the originating transcript is from a genomic locus that is annotated to have more than one gene structure. Regardless, it is sufficient to focus on the common region of two or multiple overlapping exons and/or introns (Figure 2C), and reason about the position of the circRNA with respect to the common region

following the same rules for classifying i-circRNAs into *complete i-circRNAs* and *partial i-circRNAs* (Figure 2C). i-circRNAs from intergenic regions are analyzed and classified in reference to the currently annotated intergenic noncoding transcripts, if available, and transcripts assembled using the same RNA-seq data that are employed to detect circRNAs.

2.1.2 Performance of CAT for circRNA identification

CAT supports RNA-seq data from different RNA library preparation protocols, including RNA library preparation with ribosome depletion and RNase R treatment as well as stranded or unstranded sequencing. We compared CAT with four most popular existing methods for circRNA detection, i.e., CIRI2 [70], CIRCexplorer2 [67], DCC [71], and find_circ [7] (Table 1). These methods share a common feature of relying on the genome annotation and/or the splicing signals for calling circRNAs, and all these methods, except DCC, ignore the strand specificity even if such information is available in data from strand-specific sequencing. These features led to two serious consequences, which were adequately dealt with in CAT. First, they completely ignore the circRNAs arising from the interior regions of introns and exons and intergenic transcripts which are typically poorly annotated and are often not associated with splicing signals. Second, ignoring strand specificity can erroneously designate a circRNA to the wrong strand.

Table 1. circRNA identification algorithms

Tool	Version	Language	Mapper	Signal?	Annotation?	URL
find_circ	N/A	Python	Bowtie2	Yes	No	http://www.circbase.org/
CIRI2	2.0.6	Perl	Bwa	Yes	No	https://sourceforge.net/projects/ciri/files/CIRI2/
CIRCexplorer2	2.3.8	Python	Tophat	No	Yes	https://circexplorer2.readthedocs.io/en/latest/
DCC	0.4.7	Python	STAR	No	No	https://github.com/dieterich-lab/DCC
CAT	N/A	Python	Bowtie2	No	No	https://github.com/xiaoxin8712/CAT

CAT had an estimated error rate that was comparable with that of the other four methods. In a comparative analysis for estimating the error rates of these five methods, we used two datasets, a single-end dataset of HeLa cell (GSE130905) and a paired-end dataset of prostate cancer cell lines

(GSE113120). Each dataset contains two sets of RNA-seq data, one set from a ribosome-depletion protocol and the other from a protocol of ribosome-depletion followed by RNase R treatment which removes linear RNAs (Figures 3A and 3B). In this comparison, all these methods used the most stringent read-mapping quality and a circRNA was claimed to be identified by a method if it detected at least *k* (*k*=3 for HeLa and *k*=12 for prostate cancer) RNA-seq reads across the back-splicing or back-fusion point of the circRNA with a threshold of a quality score as 40. Taken the RNase R treatment as the gold standard for circRNA detection, we estimated the error rate of a method on the ribosome-depletion data by comparing the results from the ribosome-depletion data and data of ribosome depletion followed by RNase R treatment, a scheme proposed in [72]. Among the methods compared in HeLa, DCC had the highest estimated error rate of 13.93%, find_circ had the smallest estimated error rate of 8.35%, and the estimated error rate of CAT was ~3% lower than that of DCC and ~2% higher than that of find_circ. In prostate cancer, CIRCexplorer2 had the highest estimated error rate of 8.70%, find_circ had the smallest estimated error rate of 2.49%, while the estimated error rate of CAT was ~4% lower than that of CIRCexplorer2 and ~1.5% higher than that of find_circ. Moreover, the comparative analysis revealed that to reduce the possible erroneous calling of all types of circRNA, it was important to adopt stringent criteria for detecting circRNAs, including the read mapping quality and the minimal number of supporting reads mapped to the back-fusion point of a circRNA (Figures 3C-E).

The ability to detect i-circRNAs enables CAT to report more novel circRNAs and improves the error rate as well. In HeLa cell, CAT had an estimated error rate of 9.06% for boundary circRNAs, and the estimated error rate for i-circRNAs was 17.83%, higher than that for boundary circRNAs (Figure 3F). Most of the i-circRNAs are bonafide circRNAs, even though many i-circRNAs are

not as stable as boundary circRNAs and the expressions of these i-circRNAs may be transient and abundant in nascent RNAs, which we will discuss in detail in Section 2.3.

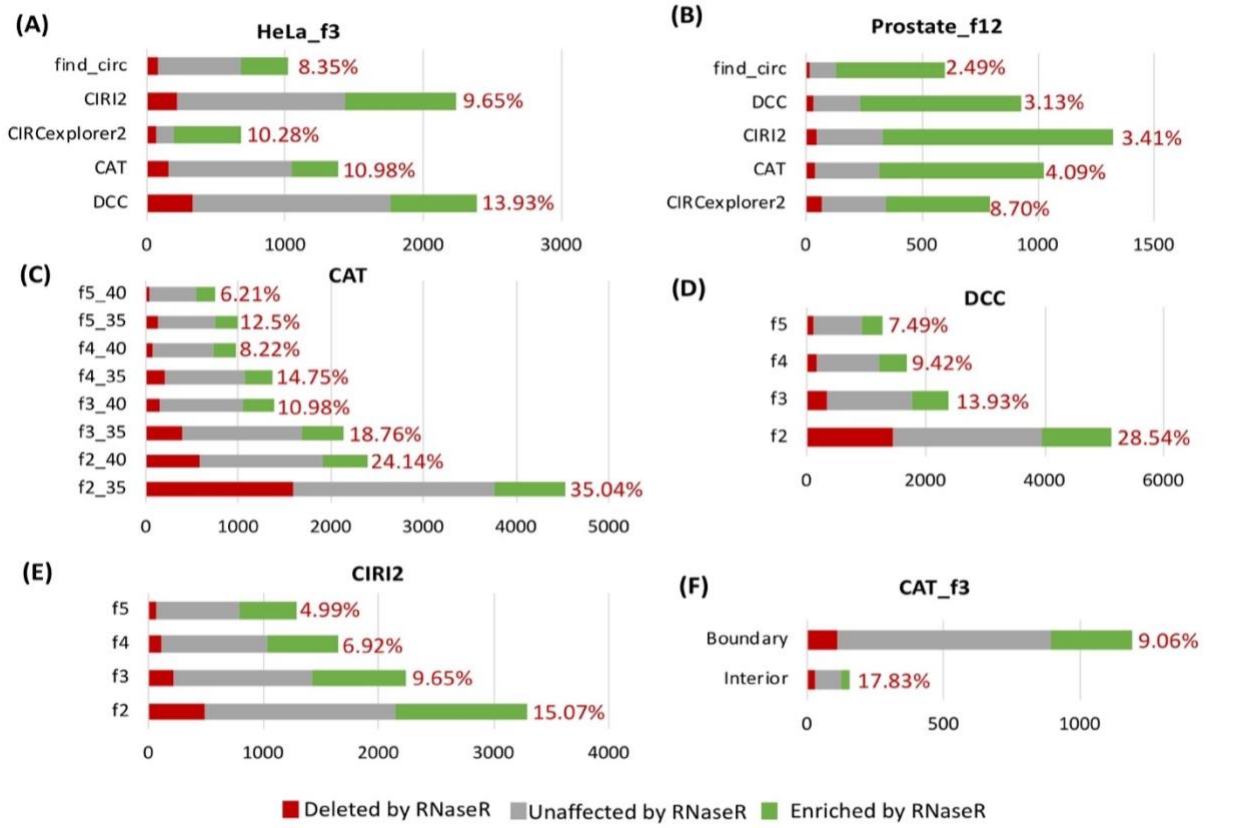
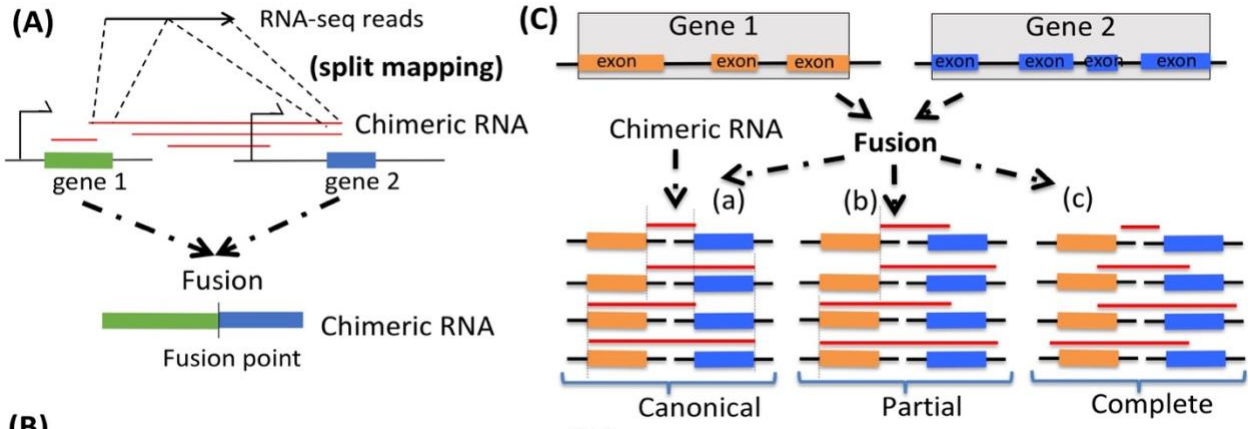


Figure 3. Specificity and sensitivity of circRNA identification methods. Stacked barplot of all predicted circRNAs in (**A**) HeLa (RNase R vs. Control) and (**B**) prostate cancer cell lines (RNase R vs. Control) stratified by RNase R resistant (>= 5 fold enrichment, green), unaffected (1-5 fold enrichment, gray) and RNase R sensitive (depleted in RNase R treated samples, red), as denoted. Percentage reflects the fraction of RNAse R sensitive circRNAs defined as false positives. The minimal number of supporting reads was set to be 3 in (**A**) HeLa and 12 in (**B**) prostate for all algorithms. The influence of different parameters on the specificity and sensitivity is shown for (**C**) CAT, (**D**) DCC, and (**E**) CIRI2 in HeLa. For example, f2_35 means that the minimal number of supporting reads is set to be 2 and the quality score is set to be 35, while f3_40 means that the minimal number of supporting reads is set to be 3 and the quality score was set to be 40. (**F**) Stacked barplot of boundary and interior circRNAs predicted by CAT in HeLa when the minimal number of supporting reads is set to be 3.

2.1.3 Identification of chimeric RNAs of all types

The method CAT for finding circRNAs can be revised to detect chimeric RNAs. If the unmappable read can be split mapped to more than one gene, which are on the same strand, on different strands, or even on different chromosomes, then the fusion point is considered as an important piece of evidence for chimeric RNA (Figure 4A). The filtering criteria for circRNAs, i.e., the number of

supporting reads *k* and mapping quality scores, are also applied to increase identification accuracy. As a chimeric RNA is originated from two genes, gene annotations are used in our method to filter out candidates from the same genes. However, we are not using splicing signals or prior knowledge of genome annotation to find the fusion junctions as other methods do, such as FusionFinder [73], FusionHunter [74], and JAFFA [75]. The information of gene structures is only used to determine the types of chimeric RNAs identified, either canonical chimeric RNAs or interior chimeric RNAs.



(B)

Tool	Sensitivity (%)	False Positives
FusionFinder	82 (41/50)	10
TopHat-fusion	80 (40/50)	73
MapSplice	78 (39/50)	23
defuse	64 (32/50)	4
FusionHunter	40 (20/50)	4
SOAPfuse	74 (37/50)	1
JAFFA	88 (44/50)	0
CAT	86 (43/50)	0

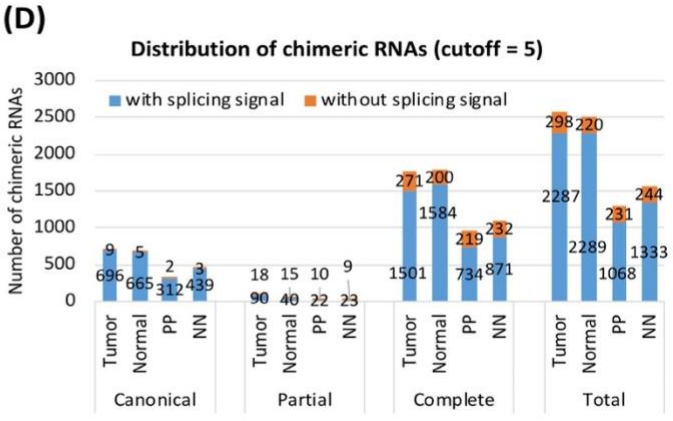


Figure 4. Identification and classification of chimeric RNAs. **(A)** Split mapping of RNA-seq reads to candidate chimeric RNAs, which may originate from an interior region of an exon or an intron of two or more genes. **(B)** Sensitivity and number of false positives for eight chimeric RNA identification methods. The number of true positives was shown together with the total number of artificially created true positives in parenthesis. **(C)** Classification of chimeric RNAs. A chimeric RNA may have three possible positions relative to the intron-exon boundaries of its originating transcript. Based on the relative position of the chimeric RNA and the known boundaries, the chimeric RNAs can be classified into a canonical chimeric RNA, *complete interior chimeric RNA* or *partial interior chimeric RNA* where one of the fusion points of the chimeric RNA is aligned to one of the intron-exon boundaries. **(D)** Distribution of chimeric RNAs in tumor and normal tissues of prostate cancer, and psoriatic-involved (PP) lesions and normal (NN) controls of skin tissues. Cutoff for the number of supporting reads in at least one sample is set to be 5.

To evaluate the performance of CAT on chimeric RNA detection, we compared CAT with seven popular existing tools (Table 2) for fusion transcript detection. We tested and compared these methods using an artificially created fusion transcript dataset [76], which contained 50 true fusion transcripts. CAT had the lowest number of false positives and the second-highest sensitivity among all eight methods compared (Figure 4B).

Table 2. Chimeric RNA identification algorithms

Tool Name	Group	Language	Exon?	URL
FusionFinder	Paired/Single-end	Perl	Yes	https://sourceforge.net/projects/fusionfinder/
TopHat-Fusion	Paired/Single-end	Tophat	No	http://ccb.jhu.edu/software/tophat/fusion_index.shtml
MapSplice	Paired/Single-end	Python	Yes	http://www.netlab.uky.edu/p/bioinfo/MapSplice
defuse	Paired-end	C++, Perl, R	No	https://shahlab.ca/projects/defuse/
FusionHunter	Paired-end	Perl	Yes	https://github.com/leofountain/FusionHunter
SOAPfuse	Paired-end	Perl	No	https://github.com/Nobel-Justin/SOAPfuse
JAFFA	Paired/Single-end	Bash, R	Yes	https://github.com/Oshlack/JAFFA/wiki
CAT	Paired/Single-end	Python	No	https://github.com/xiaoxin8712/CAT

To reiterate, CAT uses no information of splicing signals or exon/intron boundaries for chimeric RNA detection but instead uses genome annotation to classify chimeric RNAs into the categories of conventional chimeric RNAs with two fusion points at exon or intron boundaries, *complete interior chimeric RNAs* with both fusion points not at exon or intron boundaries, and *partial interior chimeric RNA*s with one fusion point from exon or intron boundaries (Figure 4C).

Table 3. Descriptions of four datasets of stranded RNA-seq used for chimeric RNA identification

Dataset	Organism	Tissue	Library	No. of Samples	Accession
Tumor	Homo sapiens	Prostate	Ribominus	30	GSE133626
Normal	Homo sapiens	Prostate	Ribominus	30	GSE133626
PP	Homo sapiens	Skin	Ribominus	28	GSE121212
NN	Homo sapiens	Skin	Ribominus	38	GSE121212

Besides, we applied CAT to four collections of stranded RNA-seq data from tumor and normal tissues of human prostate cancer, psoriatic-involved (PP) lesions, and normal (NN) controls of

human skin (Table 3). We identified a large number of chimeric RNAs, many of which were interior chimeric RNAs residing inside of introns, exons, and intergenic regions. In the tumor, normal, PP and NN dataset, 1880 (72.8% of the total), 1839 (73.3%), 985 (75.8%) and 1135 (72.0%) of the chimeric RNAs that we detected were interior chimeric RNAs when the number of supporting reads in at least one sample was not less than 5, among which 1772, 1784, 953 and 1103 were *complete interior chimeric RNAs*, respectively (Figure 4D). Interestingly, there exist more interior chimeric RNAs than canonical chimeric RNAs, and *complete interior chimeric RNAs* constitute a large part of all chimeric RNAs of the two tissues that we profiled (Figure 4D). Not surprisingly, a portion of interior chimeric RNAs does not carry splicing signals. In particular, among the *complete interior chimeric RNAs*, 271 (15.3%), 200 (11.2%), 219 (23.0%) and 232 (21.0%) are not adjacent to splicing signals in the tumor, normal, PP and NN dataset, respectively (Figure 4D).

2.2 Interior circRNAs are ubiquitous in eukaryotes

We applied CAT to five collections of stranded RNA-seq data from human (HeLa cells, prostate cancer cell lines, normal brain, and normal liver tissues) and Arabidopsis (inflorescences) (Table 4). We identified a large number of circRNAs, many of which were i-circRNAs residing inside of introns, exons, and intergenic regions.

In HeLa cells, prostate cancer cell lines, Arabidopsis inflorescences, human normal brain and liver tissues, 641 (92.2% of the total), 664 (55.1%), 304 (95.9%), 87 (56.5%) and 32(69.6%) of the circRNAs that we detected were i-circRNAs, among which 515, 313, 251, 72 and 27 were *complete i-circRNAs*, respectively (Figure 5A). Interestingly, there exist more i-circRNAs than canonical circRNAs, and *complete i-circRNAs* constitute a large part of all circRNAs of the two organisms that we profiled (Figure 5A). Remarkably, most i-circRNAs do not carry splicing

signals, i.e., their back fusion points are not adjacent to splicing signals of the major or minor splicing donor and receptor sites (Figure 5A).

Table 4. Descriptions of five collections of stranded RNA-seq data

Dataset	Library	Samples	Accession
HeLa cells (in-house)	RNase R treated	4	GSE119938
Prostate cancer cell lines	RNase R treated	2 samples for each cell line: 22Rv1, V16A, PC-3 and 42D, 4 samples for LNCaP.	GSE113120
Normal brain tissues	Ribominus	2	GSE77661
Normal liver tissues	Ribominus	2	GSE77661
Arabidopsis	RNase R treated	4	GSE117416

The circRNAs detected in HeLa cells were genuine circRNAs since a stranded RNA extraction method with a ribo-zero protocol (to preserve noncoding RNAs) plus RNase R treatment (to remove linear transcripts) was adopted in RNA-seq library preparation. To further confirm the i-circRNA candidates that we identified to be *bona fide* i-circRNAs, 14 *complete i-circRNAs* and 2 canonical circRNAs from HeLa cells were chosen for experimental validation. Divergent PCR primers were designed and applied separately to the RNA and DNA of the circRNAs to be validated. The circRNAs that were detected in RNA but not in DNA were further subjected to Sanger sequencing for confirmation (Figures 5B-C and Figures 8B-D). Note that Sanger sequencing successfully recovered the full-length structure of an intronic i-circRNAs *hsa_circ_0092378*. In our experiments, we also included *hsa_circ_0005035* and *has_circ_0092379*, a canonical circRNA from the second exon of gene *IGF1R* on human chromosome 15 and a canonical circRNA from the first exon of gene *RMRP* on human chromosome 9, respectively, as positive controls (Figures 5B and 8B). In total, 5 (35.7% of 14) i-circRNAs and 2 (100% of 2) canonical circRNAs in HeLa cells were experimentally validated. One possible reason for not being able to detect all 14 i-circRNAs using PCR was the significantly

lower expression abundance of i-circRNAs than canonical circRNAs in HeLa cells – the former was on average nearly one order of magnitude lower than the latter (Figure 5D). Combined, the results showed that i-circRNAs exist in animal and plant species, which are in concordance with the previous results that circRNAs appear in nearly all branches of life [10].

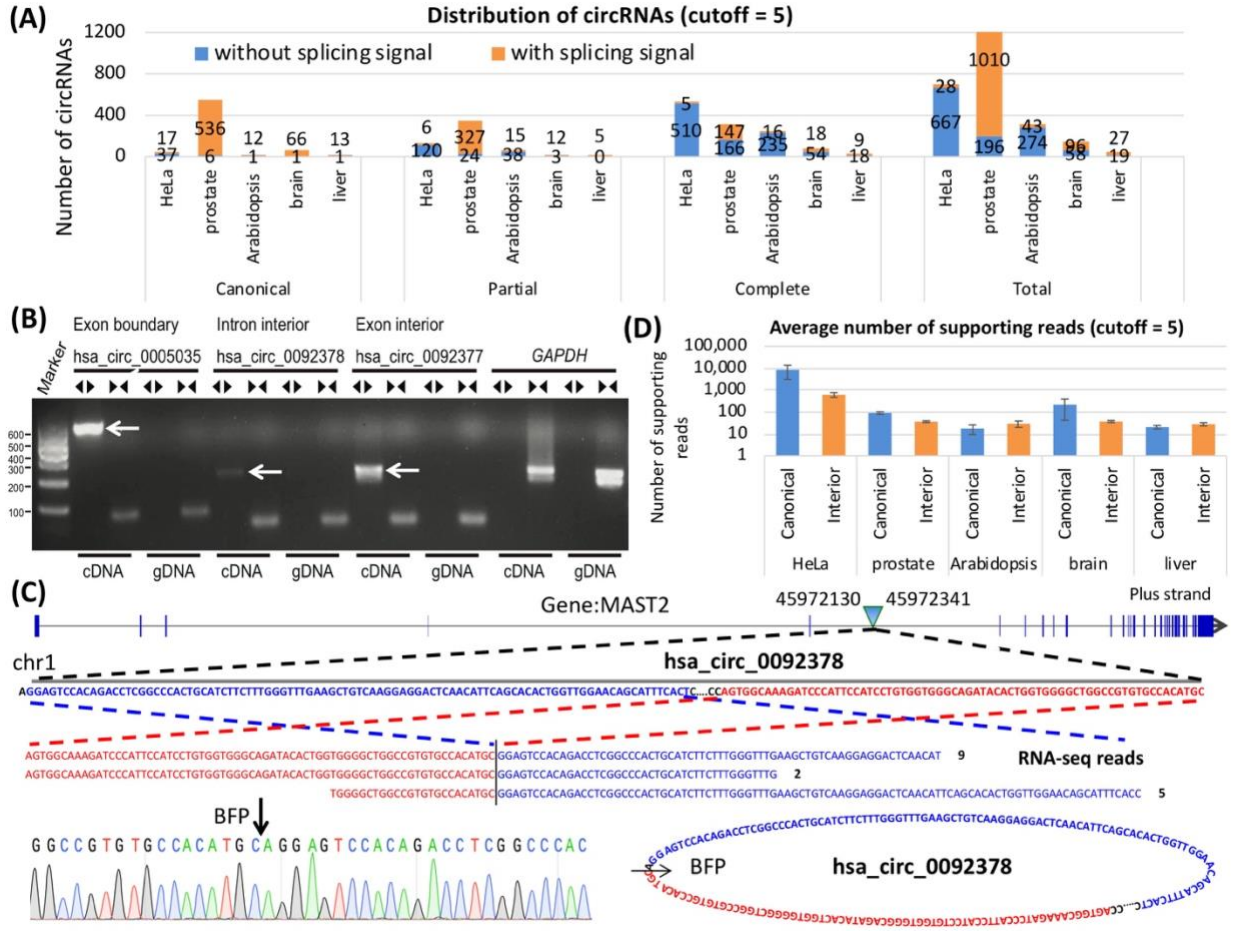
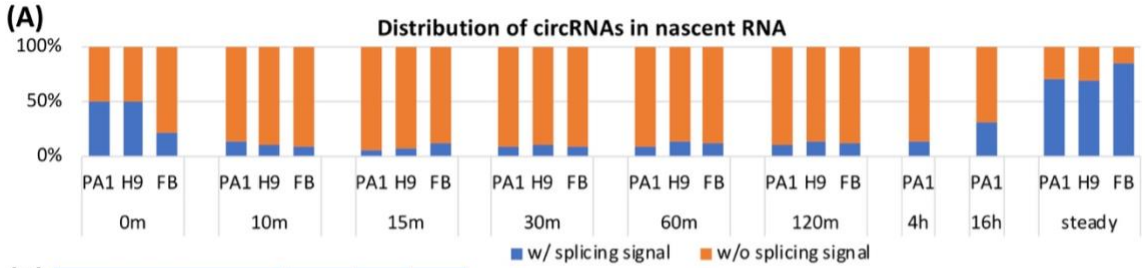


Figure 5. Distribution and validation of circRNAs. **(A)** Distribution of circRNAs in HeLa cells, prostate cancer cell lines, Arabidopsis inflorescences, human normal brain and liver tissues when number of supporting reads in at least one sample $k = 5$ and threshold for quality score is set as 35. **(B)** Validation of a canonical circRNA (*has_circ_0005035*) and two i-circRNAs in HeLa cells by PCR. The divergent and convergent arrows above the gel image represent, respectively, the divergent and convergent PCR primers used, and the white arrows in the gel image point to circRNAs. Here, *hsa_circ_0005035*, a canonical (exonic boundary) circRNA, was used as a positive control, *hsa_circ_0092378* is an intronic i-circRNAs and *hsa_circ_0092377* is an exonic i-circRNA. Gene *GRAPDH* was used as an internal control. **(C)** The genomic origin of intronic i-circRNA *hsa_circ_0092378*, its circular structure and back-fusion points (BFPs), and the results from RNA-seq and the Sanger sequencing. **(D)** Average expression levels (quantified by RNA-seq reads) of canonical circRNAs and i-circRNAs (including *complete i-circRNAs* and *partial i-circRNAs*).

2.3 Interior circRNA predominantly expressed in nascent RNA

It was unexpected and surprising that i-circRNAs expressed predominantly in nascent (newly-transcribed) RNA (Figure 6). We found that more than 50% of the total detected circRNAs were i-circRNAs without splicing signals (Figure 6A) by applying the CAT method to a non-stranded Ribominus RNA-seq dataset (Figure 6B). These results from the three cancer cell lines that we considered and from the first 120 minutes when nascent RNAs were profiled [77] were consistent (Figure 6A).



(B)

GSE73325 # samples		PA1	H9	FB
Ribominus	0m	1	1	1
	10m	1	2	1
	15m	2	2	2
	30m	2	2	2
	60m	3	2	2
	120m	3	2	2
	4h	2	-	-
	16h	2	-	-
	Steady	2	1	1
RNaseR	30m	4	-	-
	60m	2	-	-
	120m	3	-	-

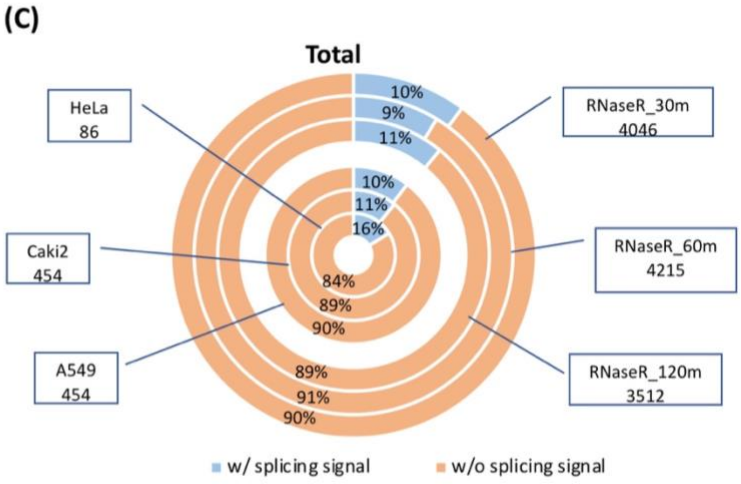


Figure 6. Distribution of circRNAs in nascent RNA. (**A**) Distribution of circRNAs in different time points of three cell lines, PA1, H9, and FB, sequenced from non-stranded and Ribominus libraries. (**B**) The number of samples for each time point in the non-stranded nascent RNA dataset GSE73325. (**C**) Distribution of circRNAs in three stranded Ribominus cancer cell lines, HeLa, Caki2 and A549, and three-time points of non-stranded RNase R treated libraries of PA1. All circRNAs are annotated as to whether being adjacent to splicing signals or not. The total number of circRNAs detected for each dataset, including boundary and interior circRNAs after classification, are indicated below the name of the dataset.

Interestingly, the expression of circRNAs, particularly i-circRNAs, was transient in nascent RNA. During the period of nascent RNA profiling, only a small percentage of circRNAs persisted from

one time-point to the next, indicating a high turn-around rate of circRNA. To confirm the validity of the transient circRNAs in nascent RNAs, we also used RNase R treated RNA-seq data of PA1 (Figure 6B). This set of data covered three time-points that matched the RNA-seq dataset from the ribosome depletion protocol. The interior and canonical circRNAs from the RNase R treated dataset had similar expression patterns (Figure 6C) as that from the ribosome-depletion dataset (Figure 6A), i.e., ~90% of the detected circRNAs, were not associated with splicing signals.

To further validate that interior circRNA predominantly expressed in nascent RNAs, we also studied the stranded ribo-minus RNA-seq data from three cancer cell lines (GSE92250), HeLa (cervical cancer), A549 (lung carcinoma), and Caki2 (epithelial carcinoma of the kidney) [78]. Similar expression patterns were also observed in these cancer cells (Figure 6C). It is important to highlight that the RNA-seq data for these three cancer cell lines were collected at the M phase of the cell cycle, indicating that interior circRNAs may potentially function in cancer cell cycle control.

Chapter 3. Characteristics of circular and chimeric RNA

Recent genome-wide sequencing-based RNA profiling has revealed properties and potential functions of some circRNAs and chimeric RNAs, but the characteristics and functions of most circRNAs and chimeric RNAs are still elusive. Here, we analyze circRNAs and chimeric RNAs identified by CAT and study their characteristics, particularly the characteristics of interior circRNAs and interior chimeric RNAs.

3.1 Characteristics of circRNA

Some characteristics of circRNAs have been investigated and validated so far. For example, circRNAs are stable in cells, especially most circRNAs from exons exhibit a half-life more than 48h [35]; expressions of circRNAs are specific to cell types and tissues [16-21] as well as developmental stages and ages [17, 22]; evolutionarily conserved sequences of circRNAs have been found between human and mouse. Analyzing circRNAs detected from five datasets of humans and Arabidopsis (Table 4), we have found three new characteristics of canonical circRNAs and i-circRNAs, i.e., many circRNAs are associated with short homologous sequences, some circRNAs are flanked by complementary sequences, and more than one circRNA may originate from the same genomic locus.

3.1.1 Short homologous sequences associated with circRNAs

One interesting finding of circRNAs was short homologous sequences (SHSs) at the back fusion points of canonical circRNAs and i-circRNAs (Figure 7A). The lengths of SHS vary from 1- to 45-nt with the majority in the range of 1- to 6-nt and peaked at 1-nt, 2-nt or 5-nt (Figure 7B). SHSs are overall abundant, with the percentage of circRNAs with SHSs ranging from 45.3% (315 out of 695 in HeLa) to 97.8% (45 out of 46 in the liver) and the abundance is more pronounced in i-circRNAs (Figure 7C), indicating a potential role that SHSs may play in the formation of circRNAs.

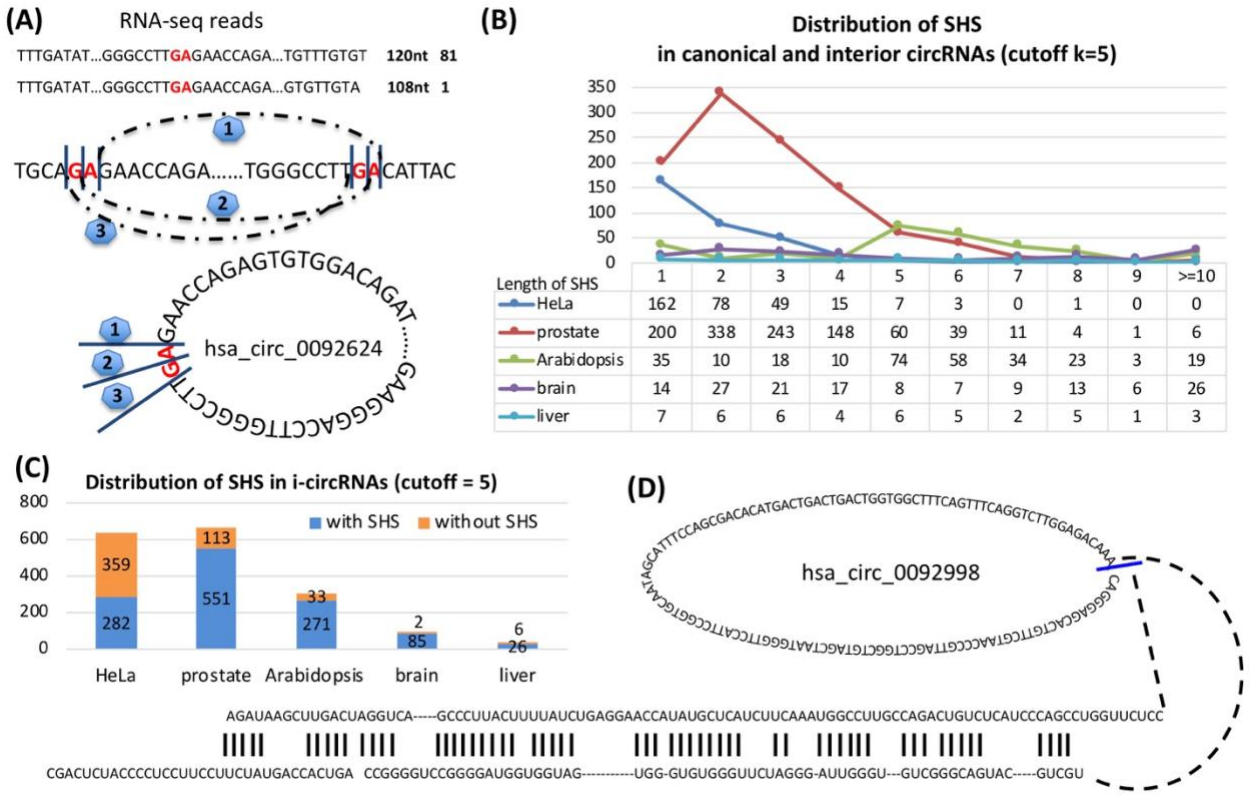


Figure 7. Sequence and structural features characteristic of i-circRNAs. **(A)** An example of the effect of SHS on possible back fusion points on i-circRNA *hsa_circ_0092624*, where due to dinucleotides GA and sequencing reads across GA, there may exist three candidate back fusion points, marked as 1, 2 and 3. **(B)** Distribution of canonical and interior circRNAs with SHSs of different lengths in 5 tissues and cell types in human and Arabidopsis. **(C)** Distribution of SHS in i-circRNAs detected from the five datasets. **(D)** An example of RNA folding structure and complementary sequences flanking an intronic i-circRNA, *hsa_circ_0092998*.

3.1.2 Complementary sequences flanking circRNAs

RNA folding structures over the flanking sequences near the back splicing or back fusion points of circRNAs constitute constructs for the production of some canonical circRNAs [43-47]. To test if this is also the case for some of the i-circRNAs identified, we fold the 100-bp sequences up- and down-stream of the back fusion points of canonical and interior circRNAs with RNAfold [79]. For each circRNA, 100-nt sequences outside of the 5'-end and 3'-end of the back fusion point are extracted separately. The two sequences are concatenated together in the 5'- to 3'-end order, and RNAfold [79] is then applied to the concatenated sequence to find the secondary structure of the

merged sequence with the minimum free energy. The secondary structure is divided into two equal parts from the center. The two parts are filtered with the following criteria: 1) at least 50-nt (i.e., 50% of the 100-nt sequence) are paired with nucleotides from the other part, and 2) no more than 10-nt are paired with nucleotides from the same part. The sequences passing through these criteria are considered to have complementary structures that may favor the generation of circRNAs. A total of 26 (4.1% of 641), 44 (6.6% of 664), 13 (4.3% of 304), 4 (4.6% of 87), and 7 (21.9% of 32) i-circRNAs (k=5) in HeLa, prostate, Arabidopsis, brain, and liver, respectively, have paired flanking sequences (see Figure 7D for an example). This suggested that some i-circRNAs might be generated through fold-back structures of complementary sequences.

3.1.3 More than one interior circRNA from the same genomic locus

Our results also revealed that more than one i-circRNA may originate from inside of the same intron or exon for the two species we analyzed (Figure 8A). Interestingly, one transcript may even host multiple circRNAs, some of which share some common sequences. Specifically, 534 (52.5%), 57 (6.8%), 6 (13.3%), 3 (6.5%) and 21 (47.7%) of all circRNA generating transcripts may produce more than one circRNA in HeLa cells, human prostate cancer cells, Arabidopsis inflorescences, human brain, and human liver, respectively (Figure 8A). For example, four i-circRNAs emerge from the locus of the first exon of gene *RMRP* in the HeLa cells, and the exon itself produces a canonical circRNA *hsa_circ_0092379* (Figures 8B-D). The largest i-circRNA from this locus, *hsa_circ_0092379_i1*, spans across the 5'UTR and the first exon of *RMRP* (Figure 8C). This i-circRNA (*hsa_circ_0092379_i1*) and the circRNA from the first exon were identified by RNA-seq profiling experiments, and the i-circRNA was also detected by the Sanger sequencing, even though its expression level was relatively lower than the canonical circRNA. The Sanger sequencing detected three additional smaller i-circRNAs, all of which are embedded within the

largest i-circRNA (Figure 8D). The smallest i-circRNA (i.e., *hsa_circ_0092379_i4*) also resides completely inside the other two i-circRNAs (*hsa_circ_0092379_i2* and *has_circ_0092379_i3*, Figure 8D).

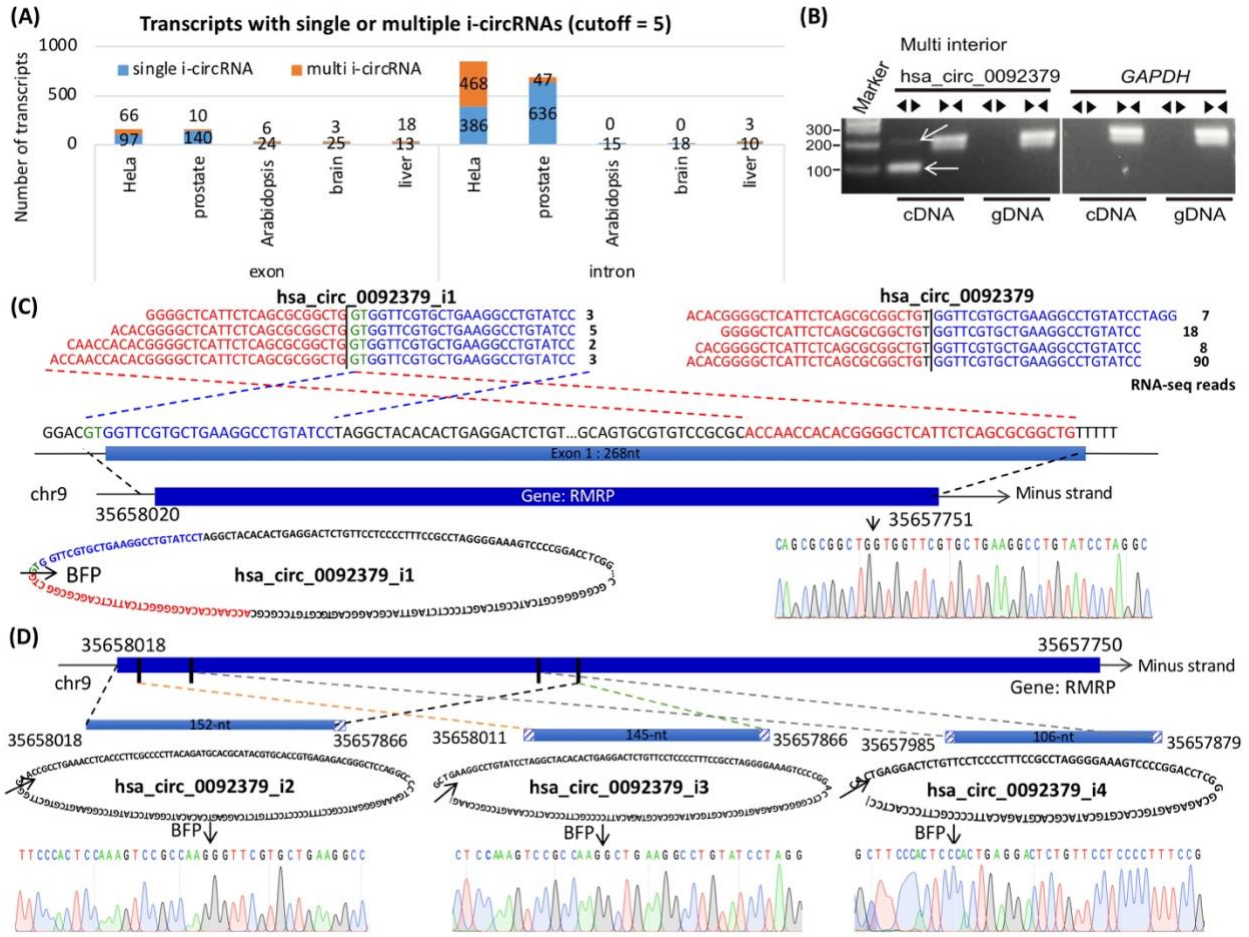


Figure 8. Multiple i-circRNAs from the same genomic locus. (**A**) Distributions of multiple and single i-circRNAs originated within a single exon or intron in the HeLa cells, prostate cancer cell lines, Arabidopsis inflorescences, human normal brain and liver tissues when the minimal number of supporting reads in at least one sample is set to $k = 5$. (**B**) Detection of multiple i-circRNAs around the locus of the first exon of gene *RMRP* in the HeLa cells by PCR. The divergent and convergent arrows above the gel image represent divergent and convergent PCR primers used, respectively, and the white arrows in the gel image indicate detected i-circRNAs. Gene *GRAPDH* is used as an internal control. (**C**) The genomic origin of the largest i-circRNA *hsa_circ_0092379-i1*, its circular structure, the results from RNA-seq (the four lines of sequencing reads on the top left) and the Sanger sequencing (on the lower right). The two back fusion points of this circRNA are 2-nt before the exon and 1-nt before the end of the exon. Included here is the RNA-seq result of the canonical circRNA *hsa_circ_0092379* from the exon (the four lines of reads on the top right). (**D**) The genomic origins of three smaller i-circRNAs *hsa_circ_0092379_i2*, *_i3* and *_i4*, their circular structures, back-fusion points and the results from the Sanger sequencing. All of these i-circRNAs reside within the largest i-circRNA *hsa_circ_0092379-i1* shown in (**C**).

Note that some i-circRNAs were not initially detected by RNA-seq profiling, but rather by subsequent PCR and/or the Sanger sequencing. For example, the back-fusion points of the three smaller i-circRNAs *hsa_circ_0092379_i2* to *hsa_circ_0092379_i4* (Figure 8D) were not recorded in the RNA-seq data but instead were serendipitously discovered by PCR using the same pairs of divergent primers designed to validate their companion canonical circRNA, *hsa_circ_0092379*, and the largest i-circRNA, *hsa_circ_0092379_i1*.

3.2 Characteristics of chimeric RNAs

Chimeric RNAs were analyzed based on the genomic loci of their fusion junctions and host genes. In addition to splicing signals, short homologous sequences were also found at the fusion junctions of most chimeric RNAs. For the same pair of host genes, more than one chimeric RNA may be generated by different combinations of exons and introns.

3.2.1 Short homologous sequences associated with chimeric RNAs

Short homologous sequences (SHSs) have been identified to associate with circRNAs (see Chapter 2). In chimeric RNAs, we also observed the existence of short homologous sequences overlapping with the fusion points, which may facilitate the formation of chimeric RNAs. In the tumor, normal, PP and NN dataset (Table 3), 2288 (88.5% of the total), 2213 (88.2%), 1176 (90.5%), and 1432 (90.8%) of the chimeric RNAs that we identified were associated with short homologous sequences (Figure 9A). Interestingly, not only chimeric RNAs without splicing signals but also most chimeric RNAs with splicing signals were detected to have short homologous sequences, indicating a facilitating role that SHSs may play in the formation of chimeric RNAs.

3.2.2 More than one chimeric RNA from the same genomic locus

Another property of chimeric RNAs is that more than one chimeric RNA may be generated from the same genomic locus. In tumor, normal, PP and NN, 117, 106, 78, and 94 host genes produced

more than one chimeric RNA, i.e., 293, 251, 178, and 218 chimeric RNAs, respectively (Figure 9B). Moreover, a well-known chimeric RNA in prostate cancer, *TMPRSS2-ERG* [56, 80, 81], also has multiple alternative chimeric transcripts in our tumor data (Figure 9C). The junctions of chimeric RNAs from *TMPRSS2-ERG* may be combinations of exon-exon, intron-intron, or even antisense of exon and intron, suggesting the complexity and diversity of fusion structures. Note that SHSs are also observed in five of the six examples, further confirming the involvement of SHSs in the formation of chimeric RNAs.

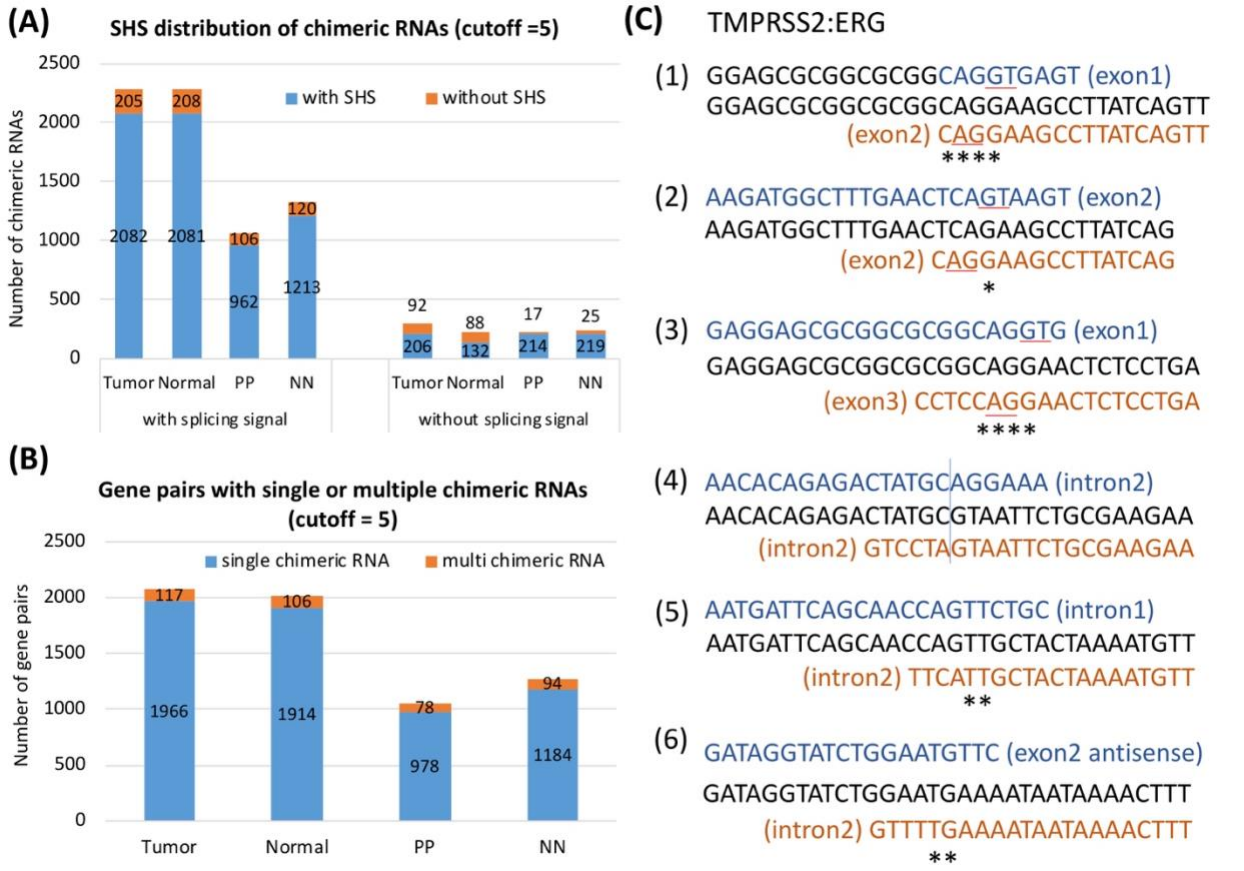


Figure 9. Sequence and structural characteristics of chimeric RNAs. (**A**) SHS distribution of chimeric RNAs in the datasets of tumor, normal, PP and NN when the number of supporting reads in at least one sample is set as $k = 5$. (**B**) Distribution of host gene pairs with single or multiple chimeric RNAs. (**C**) An example of multiple chimeric RNAs originating from the same pair of genes, *TMPRSS2* and *ERG*. Black sequences overlap the fusion points of chimeric RNAs, while blue and orange sequences are from the corresponding 5' host and 3' host genes, respectively. Splicing signals are underlined and SHSs are indicated with asterisks.

Chapter 4. Functions of circular and chimeric RNAs

Various functions have been studied for both circular and chimeric RNAs in cancer [23, 58, 66]. In this study, we focused on autoimmune disease, psoriasis, which has not been well studied for circular and chimeric RNAs. Our results revealed that circRNAs may function as competing endogenous RNAs and chimeric RNAs may compete with their host genes in psoriatic skin.

4.1 circRNAs function as competing endogenous RNAs in psoriatic skin

Psoriasis (PS) is a chronic, inflammatory, and immune-mediated skin disease, characterized by raised, red scaly plaques [82]. Its prevalence ranges from 0.09% to 11.4%, affecting at least 100 million individuals worldwide [83]. Besides the long-lasting and high recurrence rate, PS may increase the risk of stroke [84], myocardial infarction [85], type 2 diabetes [86], and cancer [87, 88]. The etiology and pathogenesis of PS remain poorly understood. Studies in the past decades have identified many genetic risk factors [89, 90] and aberrant expression of many transcripts, including non-coding RNAs such as microRNAs (miRNAs) [91-94] and long non-coding RNAs (lncRNAs) [95]. The expression of many PS related mRNAs may be regulated by miRNAs [93, 94]. For example, *miR-21* contributes to T-cell derived psoriatic skin inflammation [96], and overexpression of *miR-31* enhances the production of inflammatory cytokines and chemokines [97]. Many lncRNAs also function in PS, e.g., *PRINS* [98, 99], *lnc-IL7R* [100], and *LincR-Ccr2-5'AS* [101].

However, the expression and potential function of circRNAs in PS are understudied and poorly understood and only three studies have been described so far. A microarray-based analysis of psoriatic lesions identified 4,956 differentially expressed (DE) circRNAs [102], a study of mesenchymal stem cells (MSC) of psoriatic skin described 129 DE circRNAs [103], and a recent study of six paired lesional and non-lesional skin samples reported 148 DE circRNAs [104]. The

microarray study missed novel circRNAs, the MSC study was unlikely to have captured many circRNAs of psoriatic skin, and the skin study only compared circRNAs between lesional and non-lesional skin but not those of healthy controls.

We set forth to identify novel circRNAs and study their potential functions in PS. We studied a cohort of 93 skin tissue samples (GSE121212) from which a large collection of circRNA-enriched RNA-seq libraries from 28 psoriatic-involved (PP) lesions, 38 normal (NN) controls, and 27 psoriatic-uninvolved (PN) skin samples were derived. Besides studying circRNAs arising from intron-exon boundaries, we also studied i-circRNAs arising from the interior regions of introns, exons, and intergenic regions [105]. We confirmed our discoveries with three validation sets of RNA-seq data from psoriatic samples and cell lines. We also experimentally validated and studied three circRNAs in psoriatic skin with RT-PCR and Sanger sequencing. Furthermore, we studied the potential function of circRNAs as competing-endogenous RNAs (ceRNAs) in PS.

4.1.1 circRNAs in psoriatic skin

Four paired-end RNA-seq datasets from human skin were used to investigate circRNAs in PS. One was used for discovery, the other three were used for validation. The discovery dataset, *hsa_skin1* (GSE121212), was derived from human skin biopsies where the TruSeq Stranded Total RNA Protocol was used in combination with the RiboZero rRNA removal Kit. It consisted of 28 PP, 38 NN, and 27 PN samples. One of the validation datasets, *hsa_skin2* (GSE74697), was generated from human skin samples with the ScriptSeq complete kit from Epicenter and contained 18 PP and 16 NN samples. The other two datasets, *hsa_MSC1* (GSE81106) and *hsa_MSC2* (GSE89725), were from skin-derived Mesenchymal stem cells (MSC) and derived with the TruSeq Stranded Total RNA Library Prep kit using the CircRNA Enrichment kit and ribo-zero rRNA Removal Kit, respectively. Each of the MSC datasets consisted of 3 PP and 3 NN samples.

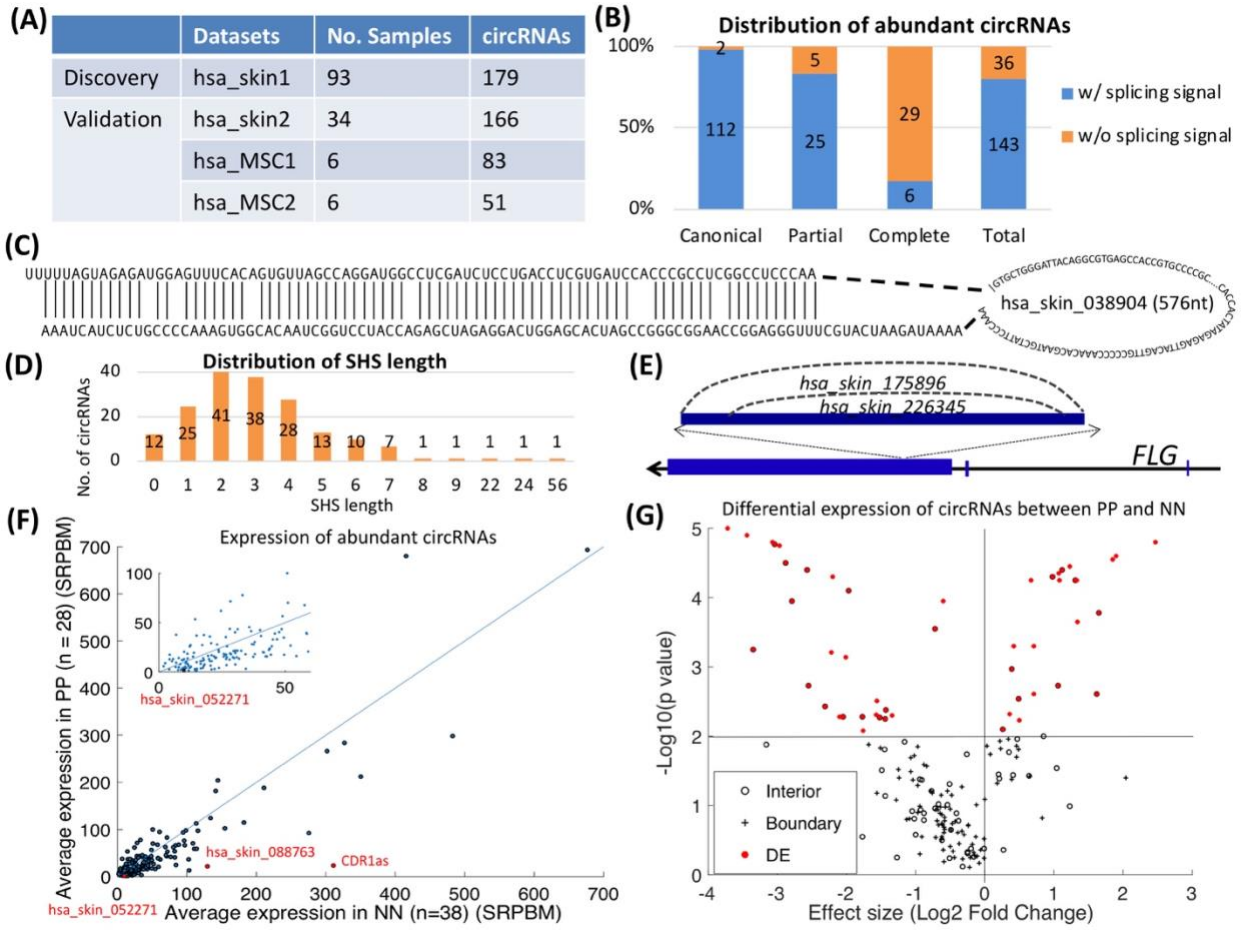


Figure 10. Distribution and characteristics of abundant circRNAs in PS. **(A)** Names and number of samples in the discovery and validation datasets. In total, 179 abundant circRNAs were identified in the discovery dataset, among which 166, 83, and 51 were independently validated by the three validation datasets. **(B)** Distribution of the 179 circRNAs in the discovery dataset. All circRNAs are annotated as to whether they are adjacent to splicing signals. **(C)** An example of RNA folding structure and complementary sequences flanking an exon i-circRNA, *hsa_skin_038904*. **(D)** Distribution of the lengths of short homologous sequences (SHS) in the discovery dataset. **(E)** An example of two circRNAs, *hsa_skin_175896* and *hsa_skin_226345*, arising from gene *FLG* (Filaggrin). **(F)** Average expression (spliced reads per billion mappings [SRPBM]) of the 179 highly expressed circRNAs in PP and NN skin. **(G)** Volcano plot visualizing differential expression of circRNAs when PP and NN skin samples were compared. The red and black dots in the plot represent significantly differentially expressed (p-value < 0.01) and not significantly expressed circRNAs, respectively. Circle and cross marks represent interior and boundary circRNAs (canonical circRNAs), respectively.

From the discovery dataset, we identified 179 circRNAs expressed in at least 10 samples and supported by at least 5 reads in one of these samples. Among the 179 circRNAs, 64 (35.8%) were identified for the first time and 65 (36.3%) were i-circRNAs, 49 of which were novel. As a validation of the result, we examined the expression of the 179 circRNAs in the three validation

RNA-seq datasets (Figure 10A). In total, 170 circRNAs were expressed in at least one of the three validation datasets. In particular, 166 (92.7%) of the 179 circRNAs were expressed in another large RNA-seq dataset *hsa_skin2* of 34 psoriatic and normal skin biopsy samples, and 37 circRNAs were highly expressed in all three validation datasets (Figure 10A). Among these, three were exon or intron i-circRNAs with no associated splicing signal. A substantial portion of the canonical and *partial i-circRNAs*, but very few of the *complete i-circRNAs*, of the 179 circRNAs carried the splicing signal (Figure 10B).

Not surprisingly, many circRNAs exhibit the three characteristics that we have revealed in Section 3.1. First, they have complementary sequences flanking their BF points; 24 circRNAs had paired flanking sequences, among which 17 were i-circRNAs including 15 with no splicing signal (see Figure 10C for an example). This suggested that the production of these i-circRNAs may be assisted by fold-back structures of complementary sequences. Second, the BF points of many circRNAs, particularly i-circRNAs, resided within short homologous sequences (SHSs), with lengths of 1- to 56-nt and peaked at 2-nt (Figure 10D). Among the 179 abundant circRNAs, 101 (56.4%) were associated with SHS that were longer than 2-nt, 29 of which had no splicing signal, indicating a potential role of SHS in circRNA biogenesis. Third, more than one circRNA may arise from a genomic locus. Among the 161 host genes of the 179 circRNAs, 13 (including *FLG*, *FLG2*, *KRT10*, *KRT2*, and *LPP*) produced multiple circRNAs. For example, *hsa_skin_175896* contained *hsa_skin_226345* (Figure 10E).

4.1.2 Aberrantly expressed circRNAs in PS

Two criteria were introduced to identify abundantly expressed circRNAs. First, the minimal number of reads mapped to a BF point in at least one sample is no less than k=5. Second, the total number of samples in which the circRNA appears is no less than m, a parameter is adjusted for

different sample sizes; in our experiment, m=10. The circRNAs that did not meet these two criteria were considered as having low expression. The expression levels of highly expressed circRNAs were normalized by SRPBM (spliced reads per billion mappings), i.e., the number of circular reads/number of mapped reads (units in billion), to ameliorate the effects of sequencing depth and batch effects. Differentially expressed (DE) circRNAs were identified by the RankSum method [106] with a threshold of p-value ≤ 0.01 based on SPRBM. DE genes of PP versus NN and PN were obtained with the Tophat and Cufflinks pipelines [107], with the threshold being adjusted to $q < 0.05$ and fold change > 2. Using SRPBM, 47 (26.3%) of the 179 circRNAs were expressed at higher levels between PP vs. NN (Figure 10F) and 51 (28.5%) circRNAs were differentially expressed (DE) in PP vs. NN skin (Figure 10G, see Appendix).

A putative function of circRNAs is acting as competing-endogenous RNAs (ceRNAs) [108] for regulating mRNA genes that share common regulating miRNAs with circRNAs. We named these circRNAs ce-circRNAs and the associated mRNAs circRNA-associated genes. More precisely, circRNA-associated mRNAs were defined as DE mRNAs that were positively regulated by circRNAs via circRNA-miRNA-mRNA regulatory cascades. That is, a DE mRNA was associated with a circRNA if both were bound and regulated by a common miRNA. mRNA and circRNA targets of mature miRNAs [109] were predicted with miRanda [110] using the default parameters of a score of 140.0 and energy of -1.0. CircRNA-associated genes were selected from the DE genes that were positively correlated with their corresponding circRNAs, i.e., Spearman's rank correlation coefficient > 0.0 and p-value < 0.01. In total, there existed 8,096 mRNA genes significantly DE between PP vs. NN skin, among which 4,034 were associated with 51 DE circRNAs.

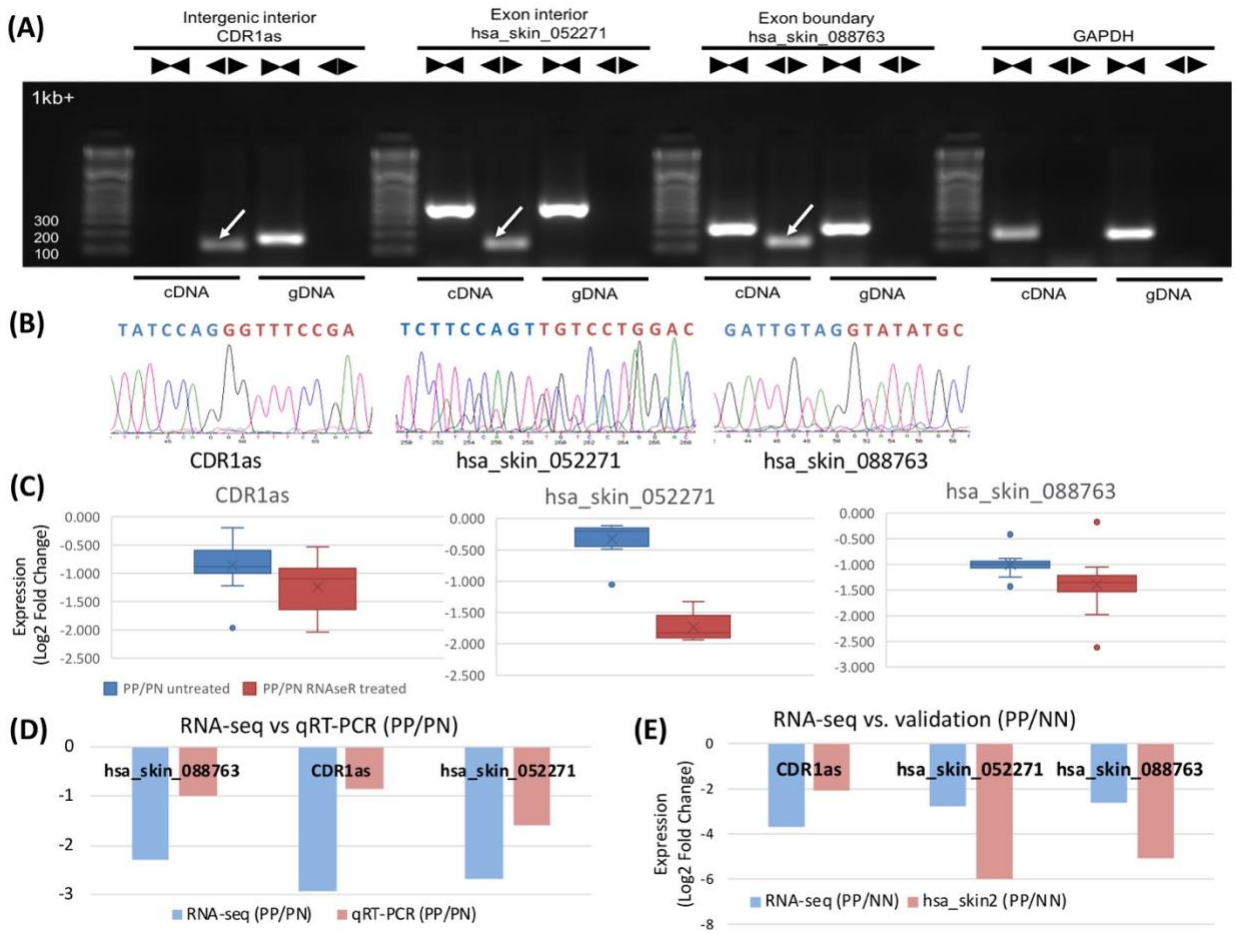


Figure 11. Experimental validation of three identified circRNAs in psoriatic skin and HaCaT keratinocyte cells. **(A)** Validation of a well-known intergenic i-circRNA (*CDR1as*), a novel exon i-circRNA (*hsa_skin_052271*), and a canonical circRNA (*hsa_skin_088763*) by PCR. The divergent and convergent arrows above the gel image represent, respectively, the divergent and convergent PCR primers used in RNA (cDNA) and DNA (gDNA), and the white arrows in the gel image point to circRNAs. Numbers on the left ranging from 100 to 1kb+ represent sizes in DNA ladder. *GAPDH* was used as an internal control. **(B)** Validation of *CDR1as*, *hsa_skin_052271*, and *hsa_skin_088763* by Sanger sequencing. **(C)** Box and whisker plot for RNase R validation of differential expression of *CDR1as*, *hsa_skin_052271*, and *hsa_skin_088763* when PP vs. PN skin was compared. **(D)** Quantitative real-time PCR (qRT-PCR) validation of differentially expressed circRNAs *CDR1as*, *hsa_skin_052271*, and *hsa_skin_088763* in PP vs. PN skin. **(E)** RNA-seq (*hsa_skin2* of validation set) validation of differentially expressed circRNAs *CDR1as*, *hsa_skin_052271*, and *hsa_skin_088763* in PP vs. NN skin.

From these circRNAs, we selected two canonical circRNAs, *hsa_skin_194345* (*CDR1as*) and *hsa_skin_088763* (*hsa_circ_0109327*), and three i-circRNAs, *hsa_skin_100269*, *hsa_skin_143837*, and *hsa_skin_052271* for experimental validation using psoriatic skin biopsy samples (PP and PN) and HaCaT keratinocyte cells. Using sequencing reads across BF points as surrogates to circRNAs, the five circRNAs had 1440, 405, 41, 24, and 22 reads in the discovery

dataset, respectively. Note that these five circRNAs were all detected in the validation datasets, providing the first validation of these circRNAs in psoriasis. For experimental validation, divergent and convergent PCR primers were applied separately to the RNA and DNA of these circRNAs. Three circRNAs, *CDR1as*, *hsa_skin_052271*, and *hsa_skin_088763*, were experimentally validated (Figure 11A) and their BF points were confirmed with Sanger sequencing (Figure 11B). Sanger sequencing successfully recovered the full-length of *hsa_skin_052271*. Furthermore, the three validated circRNAs were subjected to RNase R treatment, and the changes in their expression in PP vs. PN were confirmed. The fold changes from the RNase R experiments were consistent with and even more pronounced than that of the untreated ones (Figure 11C). qRT-PCR was also applied to validate their RNA-seq based DE in PP vs. PN skin and consistent fold change directions were confirmed (Figure 11D). We also determined that these circRNAs were DE in PP vs. NN skin by an analysis of sequencing data from the *hsa_skin2* validation dataset (Figure 11E).

4.1.3 Putative circRNA functions in PS

The three experimentally validated circRNAs may potentially function as ce-circRNAs in psoriatic skin. *CDR1as* may regulate 149 out of 655 targets of *miR-7-5p* via 67 binding sites and 242 out of 962 targets of *miR-135b-5p* via 4 binding sites (Figure 12A). *CDR1as* was sharply down-regulated 3.7 fold between PP vs. NN, whereas both *miR-7-5p* and *miR-135b-5p* were up-regulated in psoriasis [91, 94, 111]. The target genes that were down-regulated in PP skin and positively correlated with the expression of *CDR1as* by Spearman's rank-order correlation analysis were called circRNA-associated genes of *CDR1as* (Figure 12A). Among the associated genes were *EGR3*, *GATA6*, *GATA3*, and *FOXN3*, which play important roles in psoriasis. *EGR3* can regulate late epidermal differentiation and contribute to the keratinocyte differentiation-related module in

a skin-specific network [112]. *GATA6* has significantly lower expression in psoriatic dermal MSCs [113]. *FOXN3*, regulating cell differentiation and cell cycle, is down-regulated along with *GATA3* in psoriasis [114].

CircRNA *hsa_skin_088763* from pseudogene *RP11-255H23.2* may regulate 109 out of 560 targets of *miR-338-3p* via 4 binding sites and 340 out of 1508 targets of *miR-23a/b-3p* via 10/11 binding sites (Figure 12B). *hsa_skin_088763* was down-regulated 2.6 fold when PP vs. NN was compared, and *miR-338-3p* and *miR-23a/b-3p* were up-regulated in psoriasis, consistent with this trend in their putative regulatory circRNA [115]. Several psoriasis-related genes were among the associated genes of *hsa_skin_088763*, including *GATA6*, *SIK2* (Figure 12B), *IL17RD* [116], *EGR3*, *FAS*, *LRIG1*, and *PPARGC1A* [117]. *LRIG1* negatively regulates growth factor signaling to regulate epidermal stem cell quiescence [118]; *SIK2* modulates cytokine responses during innate immune activation [119]; *FAS* signaling is essential for inducing key inflammatory cytokines in psoriasis [120]. All these results indicated that *hsa_skin_088763* functions as a ceRNA in psoriasis.

hsa_skin_052271, an i-circRNA from the third exon of gene *FLG2*, may mediate three target genes of *miR-135b-5p*, two target genes of *miR-205-5p*, and nine target genes of *miR-27a-3p* (Figure 12C). *hsa_skin_052271* was down-regulated 2.8 fold between PP and NN, and *miR-135b-5p* [91, 94, 111], *miR-205-5p* [91] and *miR-27a-3p* [115] were up-regulated in psoriasis. Psoriasis-related gene *GATA6* was an associated gene of *hsa_skin_052271* (Figure 12C).

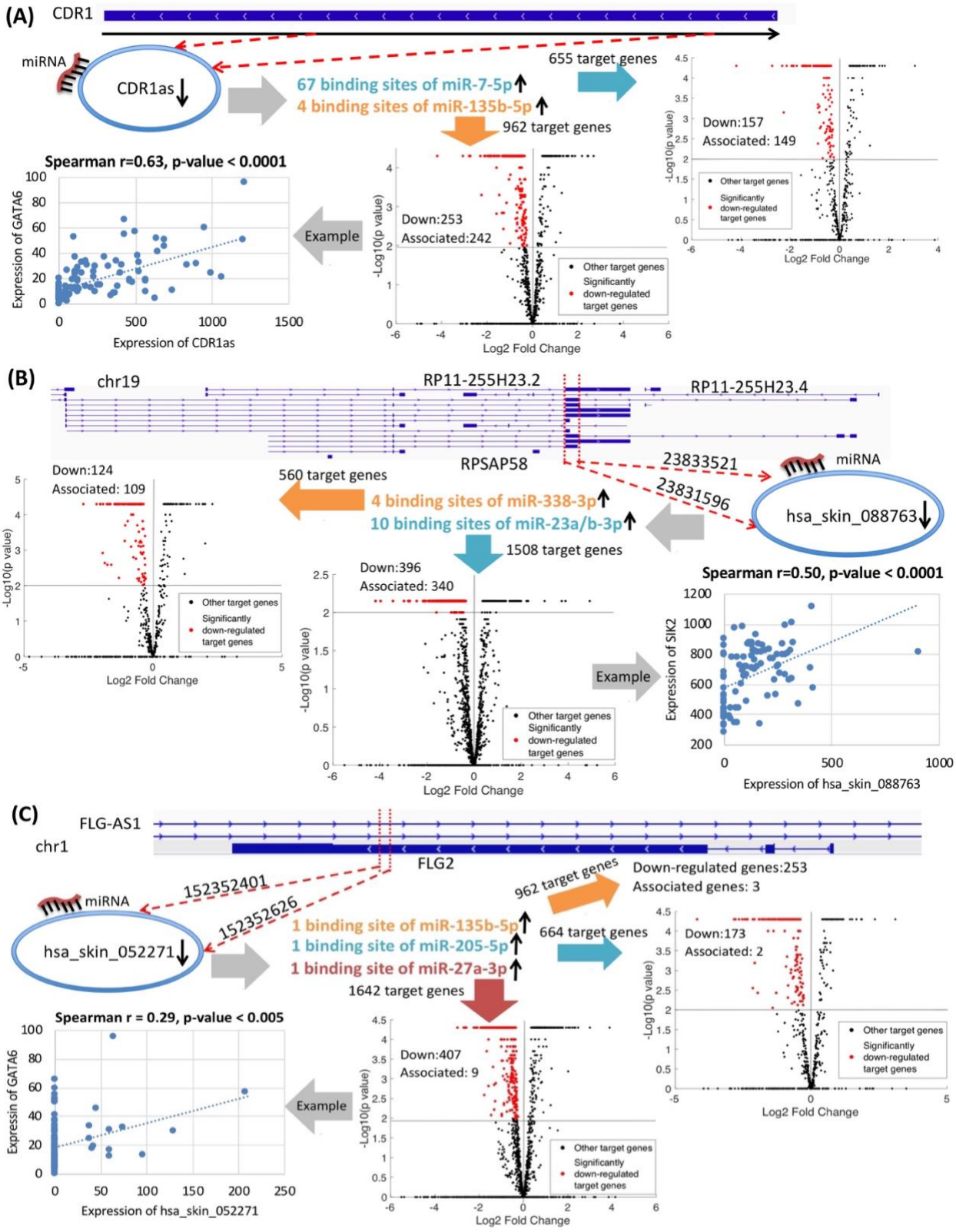


Figure 12. Putative functions of three circRNAs as ceRNAs. The genomic origin of **(A)** intergenic i-circRNA *CDR1as*, **(B)** exon boundary circRNA *hsa_skin_088763*, and **(C)** exon i-circRNA *hsa_skin_052271*, their psoriasis-related binding miRNAs predicted by miRanda, and their circRNA-

associated genes. Volcano plots visualize the differential expression of target genes between PP and NN skin. The red and black dots in the plot represent significantly down-regulated (p-value < 0.01) and other target genes, respectively. Among the significantly down-regulated target genes, those that are positively correlated with the expression of circRNAs are defined as circRNA-associated genes. Scatter plots show Spearman's rank-order correlation between three circRNAs and one of their associated genes, where r is the Spearman correlation coefficient.

4.2 Potential functions of chimeric RNAs in prostate cancer and psoriasis

Some chimeric RNAs and their encoded proteins have been taken as biomarkers and therapeutic targets of many cancers over the past decade [55, 63], which have often focused on cancer cells but failed to investigate the corresponding non-diseased cells. Recent research has also shown the existence of chimeric RNAs in non-cancerous cells and tissues [64, 121]. Here, we applied CAT to identify chimeric RNAs from four datasets of stranded RNA-seq in cancerous and normal prostate tissues, and psoriatic and normal skins (Table 3).

In total, we detected 2585, 2509, 1299, and 1577 chimeric RNAs in the tumor, normal, PP, and NN dataset, respectively. The host genes, i.e., genes overlapping with chimeric RNAs, were analyzed by Metascape [122] and the top 20 enriched pathways were listed for each tissue (Figure 13). In the prostate, cancer and normal cells shared several tissue-specific pathways, such as "lipid oxidation" and "membrane trafficking". Moreover, the host genes of cancer cells were enriched for some cancer-related pathways, for example, "modification-dependent protein catabolic process", "histone modification", "autophagy", "signal transduction by p53 class mediator" and "signaling by NOTCH1 PEST domain mutants in cancer" (Figure 13A). In psoriasis, the common enriched pathways of lesional and non-lesional tissues were skin-specific, such as "epidermis development" and "regulation of epidermis development". Furthermore, several psoriasis-related pathways were enriched by host genes of lesional skins, e.g., "deubiquitination", "regulation of water loss via skin", "activation of the innate immune response" and "myeloid leukocyte mediated immunity" (Figure 13B).

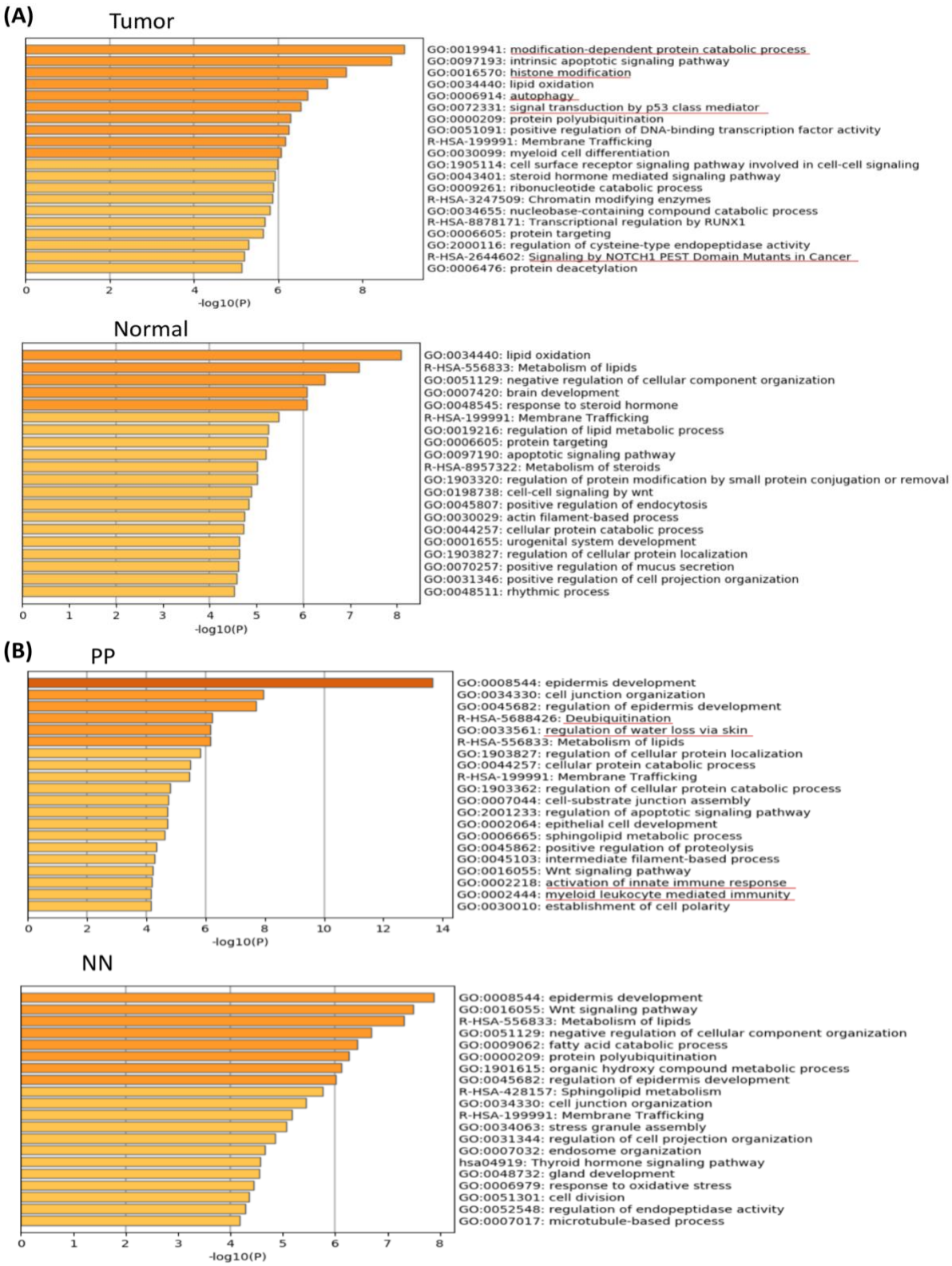


Figure 13. Pathways enriched by host genes of chimeric RNAs. Top 20 GO terms and pathways enriched by host genes of chimeric RNAs identified in (**A**) tumor and normal tissues of prostate and (**B**) in PP and NN samples of skin. Disease-associated pathways are underlined in red.

The total numbers of chimeric RNAs detected in tumor and normal tissues were comparable, but a good portion of chimeric RNAs was state specific, i.e., 1239 (47.9% of 2585) and 1163 (46.4% of 2509) were identified only in the tumor and normal tissues, respectively (Figure 14A). Similar results were also observed in psoriasis, namely, 645 (49.7% of 1299) and 923 (58.5% of 1577) chimeric RNAs were specific to PP and NN, respectively (Figure 14B).

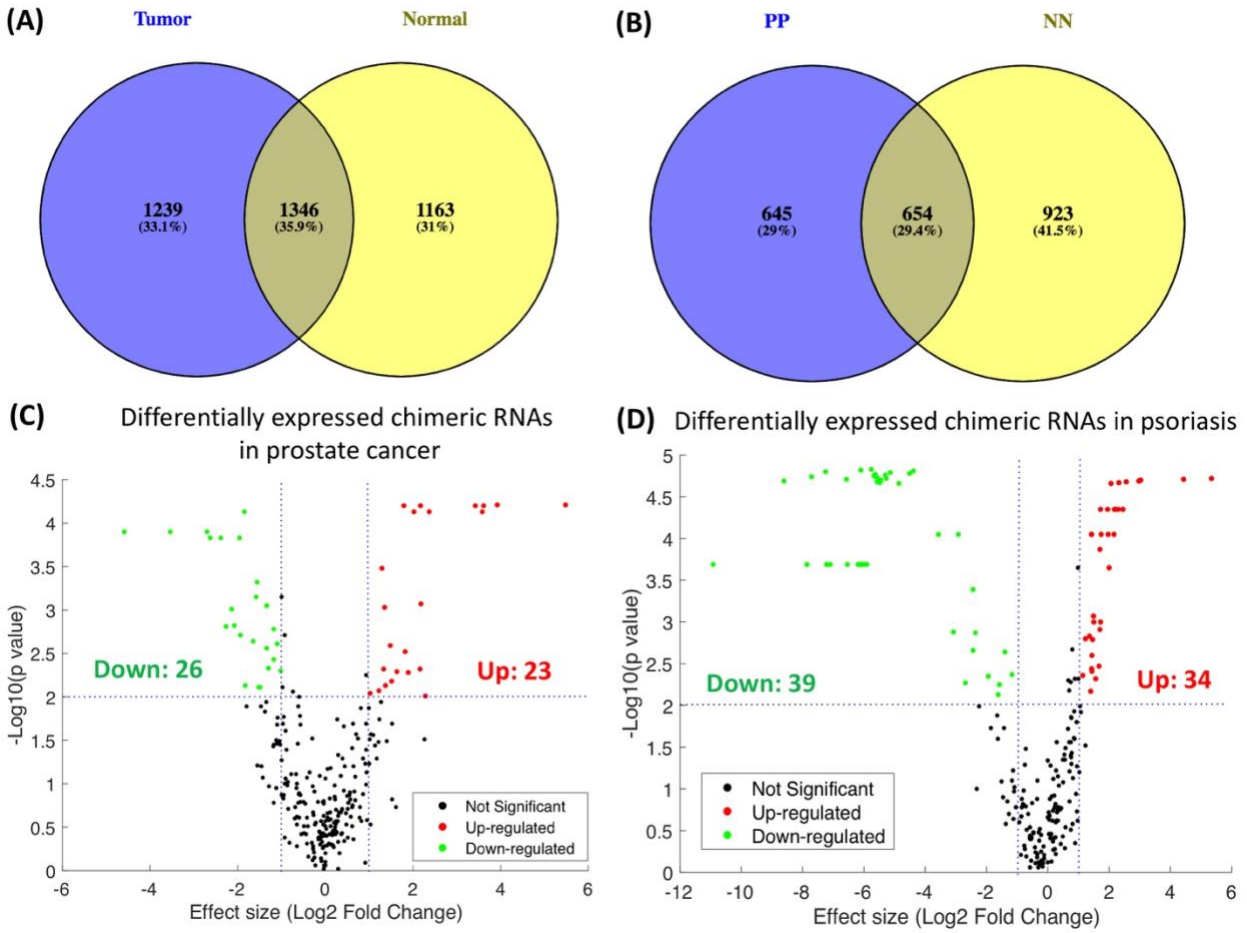


Figure 14. Differentially expressed chimeric RNAs. Venn diagrams show the number of chimeric RNAs identified in **(A)** tumor vs. normal, and **(B)** PP vs. NN. Differentially expressed chimeric RNAs **(C)** between tumor and normal tissues in prostate and **(D)** between PP and NN in skin. Chimeric RNAs with fold change > 2 and *p*-value < 0.01 were defined as significantly differentially expressed chimeric RNAs.

To better understand the functions of chimeric RNAs in prostate cancer and psoriasis, differential expression of chimeric RNAs that were expressed in at least 30 samples was analyzed by the Wilcoxon Rank Sum test [106]. Chimeric RNAs with differential expression greater than 2 folds and *p*-value < 0.01 were defined as significantly differentially expressed chimeric RNAs. Not

surprisingly, 23 chimeric RNAs were up-regulated between prostate cancer and normal tissues, whereas 26 were down-regulated (Figure 14C, see Appendix), and 34 chimeric RNAs were up-regulated between lesional and non-lesional skin, whereas 39 were down-regulated (Figure 14D, see Appendix), indicating potential functions of chimeric RNAs in prostate cancer and psoriasis.

In the prostate, the most significantly up-regulated chimeric RNA was an isoform of *TMPRSS2-ERG* (Figure 9C1) that concatenated the first exon of *TMPRSS2* with the second exon of *ERG*. It was sharply up-regulated 44.7 fold between tumor and normal tissues with p-value < 0.0001. Pathways of "nucleobase-containing compound transport" and "protein processing" were enriched by the host genes of up-regulated chimeric RNAs, while cancer-related pathways "ER to Golgi Anterograde Transport" and "Transcriptional misregulation in cancer" were enriched by the host genes of down-regulated chimeric RNAs (Figure 15A). This indicated that expression of chimeric RNAs may compete with the expression of their host genes, i.e., chimeric RNAs were lowly expressed in cancer when their cancer-related host genes were highly abundant in tumor tissues.

In psoriasis, analysis of the host genes of these DE chimeric RNAs revealed two and three enriched pathways for up-regulated and down-regulated chimeric RNAs, respectively (Figure 15B). Note that psoriasis-related pathways "skin development" and "regulation of Wnt signaling pathway" were enriched by the host genes of down-regulated chimeric RNAs. Moreover, among the 45 host genes of down-regulated chimeric RNAs, 14 genes (31.3%) were significantly up-regualted in PP vs. NN, confirming our hypothesis that the expression of these chimeric RNAs may compete with their host genes. The most significant down-regulated chimeric RNA was a novel interchromosomal chimeric RNA, *PDE6A-RMRP*, which joined the antisense of the interior region of intron 17/19 of *PDE6A* on chromosome 5 to the interior of exon 1 of *RMRP* on chromosome 9 (Figure 15C). More importantly, it was not observed in PP but expressed in all 38 NN samples

with 6,883 supporting reads and 250 unique supporting reads in total, and a 2-nt short homologous sequence was adjacent to the fusion point. Furthermore, one of the host genes, *RMRP* [123], is a long non-coding RNA (lncRNA) that regulates Th17 cells, which are important in the pathogenesis of psoriasis [124]. All these results suggest that this novel interior chimeric RNA, *PDE6A-RMRP*, may function in psoriasis and may be taken as a potential biomarker in psoriasis.

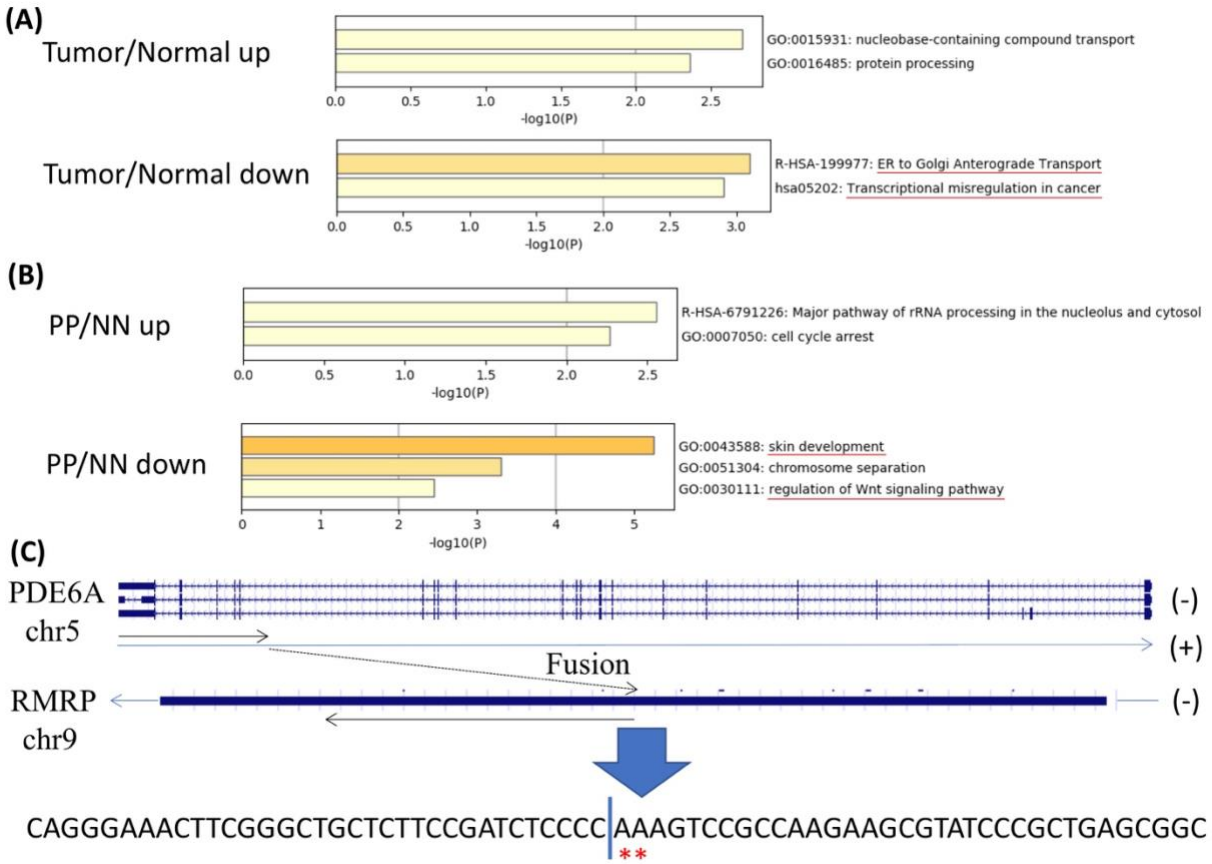


Figure 15. Pathways enriched by host genes of DE chimeric RNAs. Pathways enriched by host genes of up-regulated and down-regulated chimeric RNAs between **(A)** tumor and normal, and between **(B)** PP and NN. Disease related pathways were underlined in red. **(C)** A novel interchromosomal chimeric RNA identified in NN, *PDE6A-RMRP*, which joined the antisense of the interior region of intron 17/19 of gene *PDE6A* on chromosome 5 to the interior of exon 1 of gene *RMRP* on chromosome 9. A short homologous sequence of "AA" was observed at the fusion point, which was indicated with red asterisks.

Chapter 5. Biogenesis models and summary

Splicing is known to be a model of biogenesis for the production of messenger RNAs (mRNA), which removes introns and joins exons in the process. For circular and chimeric RNAs, back-splicing and trans-splicing are also established models of biogenesis. However, the discovery of the interior circular and chimeric RNAs was intriguing and shed new light on the biogenesis of these two types of non-co-linear RNAs. The short homologous sequences (SHSs) that we discovered in the fusion conjunctions of circular and chimeric RNAs may be a critical element of the underlying biogenesis models for circular and chimeric RNAs.

5.1 A splicesome-independent mechanism for circular and chimeric RNA production

Not replying on splicing signals was our new CAT method able to detect candidate circRNAs and chimeric RNAs that exist in all regions of the genome, especially interior regions of introns, exons, and intergenic transcripts (Figures 4D and 5A). To understand how splicing influences the formation of circular and chimeric RNAs, we compared circular and chimeric RNAs with mRNAs detected in the dataset of prostate tumors (Table 3). The percentages of RNAs without splicing signals were quite different between these three types of RNAs, i.e., on average 0.14% of mRNAs were not associated with splicing signals, while the average percentages of RNAs without splicing signals for circular and chimeric RNAs were 49% and 84%, respectively (Figure 16A). Moreover, the splicesome-based biogenesis model cannot explain all canonical circRNAs. In particular, 19,444 (21.05%) of the 92,369 circRNAs in the circBase [125] have no splicing signal adjacent to their back splicing points. This strongly suggested that a new mechanism other than splicing may exist and be responsible for the production of circular and chimeric RNAs that are not associated with splicing signals.

It was a big surprise to discover short homologous sequences (SHSs) at the fusion points of most circular and chimeric RNAs (Figures 7B, 9A, and 16B). This result was thought-provoking regarding the currently elusive mechanisms of their biogenesis. It has been well established that canonical circular and chimeric RNAs can be derived from canonical splicing sites [37, 38, 40]. On the other hand, there are pieces of evidence suggesting the existence of novel circRNA biogenesis pathways independent of the canonical splicing apparatus – blocking components of spliceosome in *Drosophila melanogaster* results in a higher ratio of circular RNAs to linear RNAs [126, 127]. The SHSs that we discovered may provide a missing piece to this puzzle. It is known that SHSs are involved in the production of some chimeric transcripts through template switching in RNA processing [128, 129]. It is then viable to hypothesize that template switching is involved in or responsible for the biogenesis of many circRNAs, particularly i-circRNAs. The complete i-circRNA *F-Circ2* from an oncogenic chimeric transcript in human discovered in an early study [130] provides direct evidence supporting this hypothesis. *F-Circ2* arises from a chimeric transcript that fuses two genes *PML* and *RARα* [130]. Remarkably, our close examination of the sequence of *F-Circ2* revealed that one of the fusion points of *F-Circ2* resides within an SHS (GCTGCCAG) inside of two exons, which starts at the 32-nt from the 5'-end of the 4-th exon of *PML* and at the 88-nt from the 5'-end of the 4-th exon of *RARα*. Besides, the early results and our analysis also suggested that RNA fold-back structures due to inverse complementary sequences or RNA-binding proteins at the two ends of the fusion points of a circRNA can bring the two ends into a close neighborhood; subsequently, the SHSs at the back fusion points of the circRNA can trigger template switching, resulting in RNA circulation. In short, the results that we presented suggested one novel circRNA and chimeric RNA biogenesis pathway independent of the canonical splicing mechanism (Figure 16C).

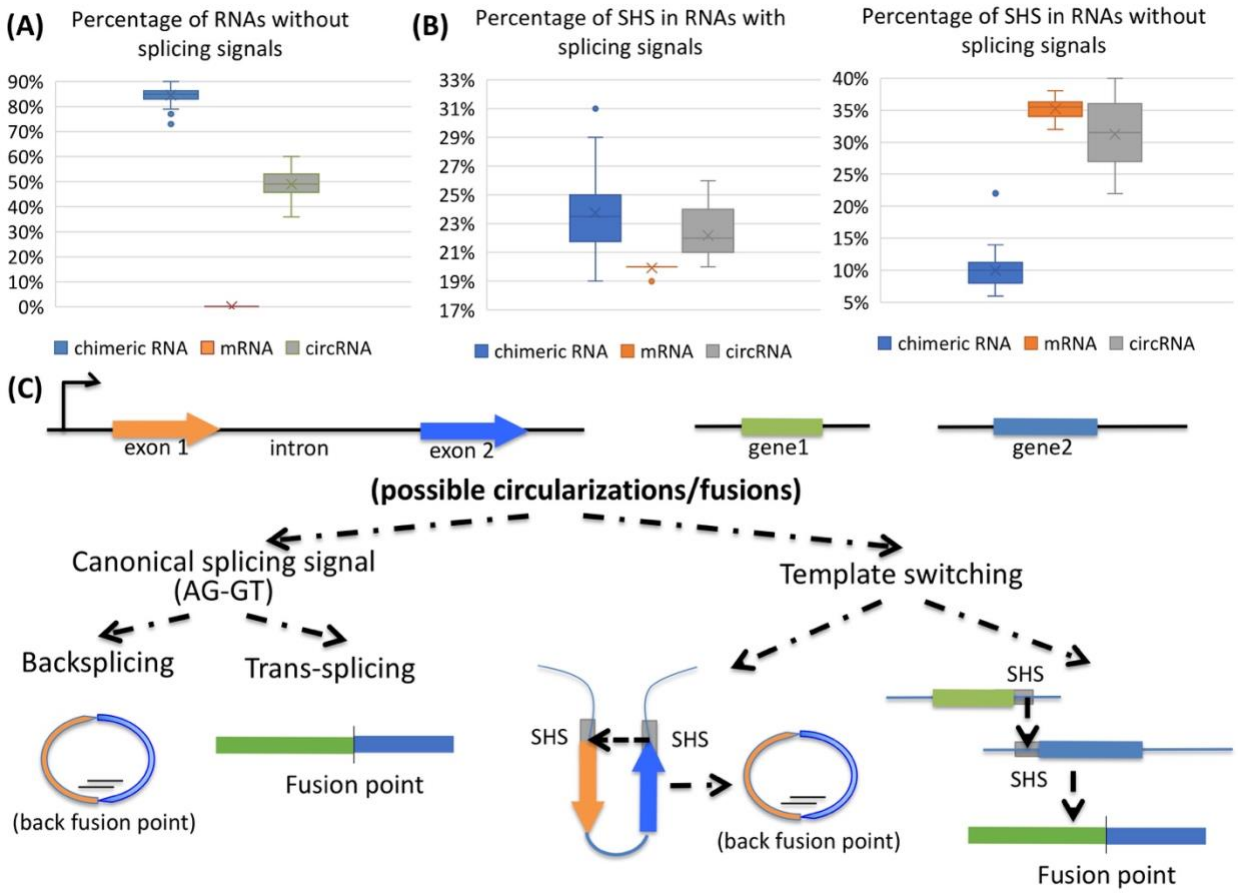


Figure 16. Mechanisms of circular and chimeric RNAs. **(A)** Box plot for percentages of RNAs without splicing signals in chimeric RNAs, mRNAs, and circRNAs in prostate cancer. (**B**) Box plots showing percentages of SHS in chimeric RNAs, mRNAs, and circRNAs with or without splicing signals in prostate cancer. (**C**) Two possible circular RNA and chimeric RNA mechanisms, (left) splicing apparatus that uses splicing signal AG/GT for production of canonical circRNAs and chimeric RNAs, and (right) RNA template switching due to SHS to generate circular or chimeric transcripts.

Although SHS is a very promising mechanism to generate circular and chimeric RNAs other than canonical splicing, there is a substantial portion of circular and chimeric RNAs whose production could not be explained by the existing models and the model that we suggested. For example, in the prostate tumor dataset, on average 24%, 20%, and 22% of chimeric RNAs, mRNAs, and circRNAs with splicing signals were associated with SHSs, respectively, and the averaged percentages were 10%, 35%, and 31% when there were no splicing signals (Figure 16B). This means that for chimeric RNAs without splicing signals, only ~10% could be generated by SHSs and the production for the remaining 90% could not be explained; for circRNAs without splicing

signals, ~30% may be formed by SHSs and the biogenesis for the other 70% circRNAs were still unknown. Therefore, multiple biogenesis mechanisms may co-exist for both circular and chimeric RNAs, and more future studies and experiments are needed to discover and validate novel mechanisms.

5.2 Summary

Although a few cases of circRNAs were reported decades ago, the recent genome-wide discovery of circRNAs as a ubiquitous form of RNA in eukaryotic organisms was surprising. The most important result that we discovered – interior circRNAs from interior regions of introns, exons, and intergenic transcripts – added a new member to the circRNA family. Utilizing our novel developed circRNA identification method CAT, we were able to detect candidate circRNAs that exist in all regions of the genome. Furthermore, in supporting and extending the early observation that circRNA may emerge from intergenic regions [7, 9, 49], we further recognized that circRNAs may arise from interior regions of intergenic transcripts. In short, our results showed that circRNAs could be produced from the interior of transcripts that have diverse genomic origins.

It was unexpected to discover more i-circRNAs than canonical circRNAs residing over intron and exon junctions. It was also surprising to detect more than one i-circRNA from one genomic locus. It is interesting to note that additional multiple i-circRNAs can be identified by PCR followed by Sanger sequencing, which were initially missed by RNA-seq profiling. This must be primarily because RNA-seq profiling was often not sufficiently deep. It is critical to highlight that multiple i-circRNAs that we identified are different from multiple circRNAs from an ORF [131] because the latter are products of alternative splicing that contains different exons or introns.

The success rate that *complete i-circRNAs* can be validated by PCR in HeLa cells is not high, at 35.7% (5 out of 14), comparing to the 100% success rate on canonical circRNAs (2 out of 2).

Despite this relatively low success rate, the validation of the 5 *complete i-circRNAs* is sufficient to demonstrate that they are bona fide circRNAs, the central theme of the current study. Moreover, a *complete i-circRNA* has also been previously discovered by PCR from an oncogenic chimeric transcript fusing two genes *PML* and *RARα* [130]. In addition, we also identified i-circRNAs in human normal and psoriatic skin, and experimentally validated three of them using disease-related cell lines and PCR followed by Sanger sequencing. Combined, all these results showed that i-circRNAs are not only genuine circRNAs, but also functional, particularly in diseases.

Not surprisingly, using CAT we identified chimeric RNAs originating from interior regions of introns, exons, and intergenic transcripts, which constitute a great proportion of the population of chimeric RNAs in the cell. Besides, chimeric RNAs also exhibit the two features of circRNAs, i.e., short homologous sequences at the fusion points and more than one chimeric transcript from the same genomic locus. These similarities between chimeric RNAs and circRNAs may due to the similar non-co-linear structures and the possibility that many of them, particularly the interior circular and chimeric RNAs, are generated by the same or similar biogenesis machinery.

We confirmed the expression and function of *TMPRSS2-ERG* in prostate cancer, which is the highest recurrent chimeric RNA [54]. More importantly, we discovered a novel interior chimeric RNA, *PDE6A-RMRP*, in human skins, which was associated with a 2-nt short homologous sequence but not any splicing signal. It was the most down-regulated chimeric RNA between psoriatic skin and normal skin – it was exclusively expressed in the normal skin. Furthermore, one of the host genes, *RMRP* [123], is a long non-coding RNA (lncRNA) that plays an important role in regulating the pathogenesis of psoriasis [124]. All these results suggest that the disappearance of this novel interior chimeric RNA, *PDE6A-RMRP*, in the psoriatic skin may serve as a potential

biomarker for psoriasis. Therefore, in addition to cancer, chimeric RNAs are also expressed in normal tissues and may be functional in immune-mediated diseases such as psoriasis.

References

1. Nigro, J.M., et al., *Scrambled exons.* Cell, 1991. **64**(3): p. 607-13.
2. Hsu, M.-T. and M. Coca-Prados, *Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells.* Nature, 1979. **280**(5720): p. 339-340.
3. Sanger, H.L., et al., *Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures.* Proceedings of the National Academy of Sciences, 1976. **73**(11): p. 3852-3856.
4. Jeck, W.R., et al., *Circular RNAs are abundant, conserved, and associated with ALU repeats.* Rna, 2013. **19**(2): p. 141-157.
5. Salzman, J., et al., *Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types.* PloS one, 2012. **7**(2): p. e30733.
6. Hansen, T.B., et al., *Natural RNA circles function as efficient microRNA sponges.* Nature, 2013. **495**(7441): p. 384-8.
7. Memczak, S., et al., *Circular RNAs are a large class of animal RNAs with regulatory potency.* Nature, 2013. **495**(7441): p. 333-338.
8. Lu, T., et al., *Transcriptome-wide investigation of circular RNAs in rice.* Rna, 2015. **21**(12): p. 2076-2087.
9. Ye, C.Y., et al., *Widespread noncoding circular RNAs in plants.* New Phytologist, 2015. **208**(1): p. 88-95.
10. Wang, P.L., et al., *Circular RNA is expressed across the eukaryotic tree of life.* PloS one, 2014. **9**(3): p. e90859.
11. Li, Z., et al., *Exon-intron circular RNAs regulate transcription in the nucleus.* Nature structural & molecular biology, 2015. **22**(3): p. 256.
12. Ashwal-Fluss, R., et al., *circRNA biogenesis competes with pre-mRNA splicing.* Molecular cell, 2014. **56**(1): p. 55-66.
13. Hansen, T.B., et al., *miRNA‐dependent gene silencing involving Ago2‐mediated cleavage of a circular antisense RNA.* The EMBO journal, 2011. **30**(21): p. 4414-4422.
14. Pamudurti, N.R., et al., *Translation of circRNAs.* Molecular cell, 2017. **66**(1): p. 9-21. e7.
15. Zhang, M., et al., *A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma.* Nature communications, 2018. **9**(1): p. 1-17.
16. Salzman, J., et al., *Cell-type specific features of circular RNA expression.* PLoS genetics, 2013. **9**(9): p. e1003777.
17. You, X., et al., *Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity.* Nature neuroscience, 2015. **18**(4): p. 603.
18. Filippenkov, I.B., et al., *Circular RNA of the human sphingomyelin synthase 1 gene: Multiple splice variants, evolutionary conservatism and expression in different tissues.* RNA biology, 2015. **12**(9): p. 1030-1042.
19. Rybak-Wolf, A., et al., *Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed.* Molecular cell, 2015. **58**(5): p. 870-885.
20. Memczak, S., et al., *Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood.* PloS one, 2015. **10**(10): p. e0141214.
21. Abe, N., et al., *Rolling circle translation of circular RNA in living human cells.* Scientific reports, 2015. **5**: p. 16435.
22. Venø, M.T., et al., *Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development.* Genome biology, 2015. **16**(1): p. 245.

23. Hansen, T.B., J. Kjems, and C.K. Damgaard, *Circular RNA and miR-7 in cancer.* Cancer research, 2013. **73**(18): p. 5609-5612.
24. Li, Y., et al., *Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis.* Cell research, 2015. **25**(8): p. 981-984.
25. Zhao, Z.-J. and J. Shen, *Circular RNA participates in the carcinogenesis and the malignant behavior of cancer.* RNA biology, 2017. **14**(5): p. 514-521.
26. Liu, Q., et al., *Emerging roles of circRNA related to the mechanical stress in human cartilage degradation of osteoarthritis.* Molecular Therapy-Nucleic Acids, 2017. **7**: p. 223-230.
27. Zhou, M.-y., J.-M. Yang, and X.-d. Xiong, *The emerging landscape of circular RNA in cardiovascular diseases.* Journal of Molecular and Cellular Cardiology, 2018. **122**: p. 134-139.
28. Bayoumi, A.S., et al., *Circular noncoding RNAs as potential therapies and circulating biomarkers for cardiovascular diseases.* Acta Pharmacologica Sinica, 2018. **39**(7): p. 1100-1109.
29. Lukiw, W.J.F.i.g., *Circular RNA (circRNA) in Alzheimer's disease (AD).* Frontiers in genetics 2013. **4**: p. 307.
30. Horsham, J., et al., *Clinical potential of microRNA-7 in cancer.* Journal of clinical medicine, 2015. **4**(9): p. 1668-1687.
31. Liu, Y.-C., et al., *CircNet: a database of circular RNAs derived from transcriptome sequencing data.* Nucleic acids research, 2016. **44**(D1): p. D209-D215.
32. Li, J.-H., et al., *starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data.* Nucleic acids research, 2014. **42**(D1): p. D92-D97.
33. Ghosal, S., et al., *Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits.* Frontiers in genetics, 2013. **4**: p. 283.
34. Lasda, E. and R. Parker, *Circular RNAs: diversity of form and function.* Rna, 2014. **20**(12): p. 1829-1842.
35. Jeck, W.R. and N.E. Sharpless, *Detecting and characterizing circular RNAs.* Nature biotechnology, 2014. **32**(5): p. 453.
36. Chen, L.-L. and L. Yang, *Regulation of circRNA biogenesis.* RNA biology, 2015. **12**(4): p. 381-388.
37. Petkovic, S. and S. Müller, *RNA circularization strategies in vivo and in vitro.* Nucleic acids research, 2015. **43**(4): p. 2454-2465.
38. Ebbesen, K.K., J. Kjems, and T.B. Hansen, *Circular RNAs: identification, biogenesis and function.* Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2016. **1859**(1): p. 163-168.
39. Chen, I., C.Y. Chen, and T.J. Chuang, *Biogenesis, identification, and function of exonic circular RNAs.* Wiley Interdisciplinary Reviews: RNA, 2015. **6**(5): p. 563-579.
40. Starke, S., et al., *Exon circularization requires canonical splice signals.* Cell reports, 2015. **10**(1): p. 103-111.
41. Zaphiropoulos, P.G., *Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis.* Molecular and cellular biology, 1997. **17**(6): p. 2985-2993.
42. Kelly, S., et al., *Exon skipping is correlated with exon circularization.* Journal of molecular biology, 2015. **427**(15): p. 2414-2417.

43. Liang, D. and J.E. Wilusz, *Short intronic repeat sequences facilitate circular RNA production.* Genes & development, 2014. **28**(20): p. 2233-2247.
44. Wilusz, J.E., *Repetitive elements regulate circular RNA biogenesis.* Mobile genetic elements, 2015. **5**(3): p. 39-45.
45. Zhang, X.O., et al., *Complementary sequence-mediated exon circularization.* Cell, 2014. **159**(1): p. 134-47.
46. Ivanov, A., et al., *Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals.* Cell reports, 2015. **10**(2): p. 170-177.
47. Conn, S.J., et al., *The RNA binding protein quaking regulates formation of circRNAs.* Cell, 2015. **160**(6): p. 1125-34.
48. Guo, J.U., et al., *Expanded identification and characterization of mammalian circular RNAs.* Genome biology, 2014. **15**(7): p. 409.
49. Gao, Y., J. Wang, and F. Zhao, *CIRI: an efficient and unbiased algorithm for de novo circular RNA identification.* Genome biology, 2015. **16**(1): p. 4.
50. Szabo, L., et al., *Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development.* Genome biology, 2015. **16**(1): p. 126.
51. Mitelman, F., B. Johansson, and F. Mertens, *The impact of translocations and gene fusions on cancer causation.* Nature Reviews Cancer, 2007. **7**(4): p. 233-245.
52. Michaeli, S., *Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome.* Future microbiology, 2011. **6**(4): p. 459-474.
53. Nilsen, T.W., *Trans-splicing of nematode premessenger RNA.* Annual review of microbiology, 1993. **47**(1): p. 413-440.
54. Jang, Y.E., et al., *ChimerDB 4.0: an updated and expanded database of fusion genes.* Nucleic acids research, 2020. **48**(D1): p. D817-D824.
55. Frenkel-Morgenstern, M., et al., *Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts.* Genome research, 2012. **22**(7): p. 1231-1242.
56. Leyten, G.H., et al., *Prospective multicentre evaluation of PCA3 and TMPRSS2-ERG gene fusions as diagnostic and prognostic urinary biomarkers for prostate cancer.* European urology, 2014. **65**(3): p. 534-542.
57. Panagopoulos, I., et al., *Genomic characterization of MOZ/CBP and CBP/MOZ chimeras in acute myeloid leukemia suggests the involvement of a damage‐repair mechanism in the origin of the t (8; 16)(p11; p13).* Genes, Chromosomes and Cancer, 2003. **36**(1): p. 90-98.
58. Nambiar, M., V. Kari, and S.C. Raghavan, *Chromosomal translocations in cancer.* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2008. **1786**(2): p. 139-152.
59. Soda, M., et al., *Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer.* Nature, 2007. **448**(7153): p. 561-566.
60. Xiao, X., et al., *Familial C3 glomerulonephritis caused by a novel CFHR5-CFHR2 fusion gene.* Molecular immunology, 2016. **77**: p. 89-96.
61. Kuno, Y., et al., *Constitutive kinase activation of the TEL-Syk fusion gene in myelodysplastic syndrome with t (9; 12)(q22; p12).* Blood, 2001. **97**(4): p. 1050-1055.
62. Loi, E., et al., *ELMOD3‐SH2D6 gene fusion as a possible co‐star actor in autism spectrum disorder scenario.* Journal of cellular and molecular medicine, 2020. **24**(2): p. 2064-2069.

63. Latysheva, N.S. and M.M. Babu, *Discovering and understanding oncogenic gene fusions through data intensive computational approaches.* Nucleic acids research, 2016. **44**(10): p. 4487-4503.
64. Babiceanu, M., et al., *Recurrent chimeric fusion RNAs in non-cancer tissues and cells.* Nucleic acids research, 2016. **44**(6): p. 2859-2872.
65. Li, P., et al., *CircRNA-Cdr1as exerts anti-oncogenic functions in bladder Cancer by sponging MicroRNA-135a.* Cellular Physiology and Biochemistry, 2018. **46**(4): p. 1606-1616.
66. Elfman, J. and H. Li, *Chimeric RNA in Cancer and stem cell differentiation.* Stem cells international, 2018. **2018**.
67. Zhang, X.-O., et al., *Diverse alternative back-splicing and alternative splicing landscape of circular RNAs.* Genome research, 2016. **26**(9): p. 1277-1287.
68. You, X. and T.O. Conrad, *Acfs: accurate circRNA identification and quantification from RNA-Seq data.* Scientific reports, 2016. **6**(1): p. 1-11.
69. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome biology, 2009. **10**(3): p. R25.
70. Gao, Y., J. Zhang, and F. Zhao, *Circular RNA identification based on multiple seed matching.* Briefings in bioinformatics, 2018. **19**(5): p. 803-810.
71. Cheng, J., F. Metge, and C. Dieterich, *Specific identification and quantification of circular RNAs from sequencing data.* Bioinformatics, 2016. **32**(7): p. 1094-1096.
72. Hansen, T.B., et al., *Comparison of circular RNA prediction tools.* Nucleic acids research, 2016. **44**(6): p. e58-e58.
73. Francis, R.W., et al., *FusionFinder: a software tool to identify expressed gene fusion candidates from RNA-Seq data.* PloS one, 2012. **7**(6): p. e39987.
74. Li, Y., et al., *FusionHunter: identifying fusion transcripts in cancer using paired-end RNA-seq.* Bioinformatics, 2011. **27**(12): p. 1708-1710.
75. Davidson, N.M., I.J. Majewski, and A. Oshlack, *JAFFA: High sensitivity transcriptome-focused fusion gene detection.* Genome medicine, 2015. **7**(1): p. 1-12.
76. Ge, H., et al., *FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution.* Bioinformatics, 2011. **27**(14): p. 1922-1928.
77. Zhang, Y., et al., *The biogenesis of nascent circular RNAs.* Cell reports, 2016. **15**(3): p. 611-624.
78. Ali, M.M., et al., *PAN-cancer analysis of S-phase enriched lncRNAs identifies oncogenic drivers and biomarkers.* Nature communications, 2018. **9**(1): p. 1-20.
79. Lorenz, R., et al., *ViennaRNA Package 2.0.* Algorithms for molecular biology 2011. **6**(1): p. 26.
80. Demichelis, F., et al., *TMPRSS2: ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort.* Oncogene, 2007. **26**(31): p. 4596-4599.
81. Tomlins, S.A., et al., *Role of the TMPRSS2-ERG gene fusion in prostate cancer.* Neoplasia, 2008. **10**(2): p. 177-IN9.
82. Harden, J.L., J.G. Krueger, and A.M.J.J.o.a. Bowcock, *The immunogenetics of psoriasis: a comprehensive review.* Journal of autoimmunity, 2015. **64**: p. 66-73.
83. Organization, W.H., *Global report on psoriasis.* World Health Organization, 2016.
84. Gelfand, J.M., et al., *The risk of stroke in patients with psoriasis.* Journal of Investigative Dermatology, 2009. **129**(10): p. 2411-2418.

85. Gelfand, J.M., et al., *Risk of myocardial infarction in patients with psoriasis.* Jama, 2006. **296**(14): p. 1735-1741.
86. Li, W., et al., *Psoriasis and risk of type 2 diabetes among women and men in the United States: a population-based cohort study.* Journal of Investigative Dermatology, 2012. **132**(2): p. 291-298.
87. Fuxench, Z.C.C., et al., *The risk of cancer in patients with psoriasis: a population-based cohort study in the health improvement network.* JAMA dermatology, 2016. **152**(3): p. 282-290.
88. Trafford, A.M., et al., *Association of Psoriasis With the Risk of Developing or Dying of Cancer: A Systematic Review and Meta-analysis.* JAMA dermatology, 2019. **155**(12): p. 1390-1403.
89. Strange, A., et al., *A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1.* Nature genetics, 2010. **42**(11): p. 985.
90. Zhang, X.-J., et al., *Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21.* Nature genetics, 2009. **41**(2): p. 205.
91. Xia, J. and W.J.P.g. Zhang, *MicroRNAs in normal and psoriatic skin.* Physiological genomics, 2013. **46**(4): p. 113-122.
92. Sonkoly, E., et al., *MicroRNAs: novel regulators involved in the pathogenesis of psoriasis?* PloS one, 2007. **2**(7): p. e610.
93. Xia, J., et al., *Noncanonical microRNAs and endogenous siRNAs in normal and psoriatic human skin.* Human molecular genetics, 2012. **22**(4): p. 737-748.
94. Joyce, C.E., et al., *Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome.* Human molecular genetics, 2011. **20**(20): p. 4025-4040.
95. Tsoi, L.C., et al., *Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin.* Genome biology, 2015. **16**(1): p. 24.
96. Meisgen, F., et al., *MiR‐21 is up‐regulated in psoriasis and suppresses T cell apoptosis.* Experimental dermatology, 2012. **21**(4): p. 312-314.
97. Xu, N., et al., *MicroRNA-31 is overexpressed in psoriasis and modulates inflammatory cytokine and chemokine production in keratinocytes via targeting serine/threonine kinase 40.* The Journal of Immunology, 2013. **190**(2): p. 678-688.
98. Sonkoly, E., et al., *Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene, PRINS.* Journal of Biological Chemistry, 2005. **280**(25): p. 24159-24167.
99. Széll, M., et al., *PRINS, a primate-specific long non-coding RNA, plays a role in the keratinocyte stress response and psoriasis pathogenesis.* Pflügers Archiv-European Journal of Physiology, 2016. **468**(6): p. 935-943.
100. Cui, H., et al., *The human long noncoding RNA lnc‐IL 7 R regulates the inflammatory response.* European journal of immunology, 2014. **44**(7): p. 2085-2095.
101. Hu, G., et al., *Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation.* Nature immunology, 2013. **14**(11): p. 1190.
102. Qiao, M., et al., *Circular RNA expression profile and analysis of their potential function in psoriasis.* Cellular Physiology and Biochemistry, 2018. **50**(1): p. 15-27.

103. Liu, R., et al., *Mesenchymal stem cells in psoriatic lesions affect the skin microenvironment through circular RNA.* Experimental dermatology, 2019. **28**(3): p. 292-299.
104. Moldovan, L.-I., et al., *High-throughput RNA sequencing from paired lesional-and non-lesional skin reveals major alterations in the psoriasis circRNAome.* BMC medical genomics, 2019. **12**(1): p. 1-17.
105. Liu, X., et al., *Interior circular RNA.* RNA biology, 2019: p. 1-11.
106. Laing, E. and C.P.J.B.r.n. Smith, *RankProdIt: A web-interactive Rank Products analysis tool.* BMC research notes, 2010. **3**(1): p. 221.
107. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nature protocols 2012. **7**(3): p. 562.
108. Salmena, L., et al., *A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?* Cell, 2011. **146**(3): p. 353-358.
109. Griffiths-Jones, S., et al., *miRBase: microRNA sequences, targets and gene nomenclature.* Nucleic acids research, 2006. **34**(suppl_1): p. D140-D144.
110. Enright, A.J., et al., *MicroRNA targets in Drosophila.* Genome biology, 2003. **5**(1): p. R1.
111. Choi, H.-R., et al., *Suppression of miR135b increases the proliferative potential of normal human keratinocytes.* The Journal of investigative dermatology, 2014. **134**(4): p. 1161.
112. Kim, K.-H., et al., *EGR3 Is a Late Epidermal Differentiation Regulator that Establishes the Skin-Specific Gene Network.* Journal of Investigative Dermatology, 2019. **139**(3): p. 615-625.
113. Campanati, A., et al., *Role of mesenchymal stem cells in the pathogenesis of psoriasis: current perspectives.* Psoriasis (Auckland, NZ), 2017. **7**: p. 73.
114. Rácz, E., et al., *GATA3 expression is decreased in psoriasis and during epidermal regeneration; induction by narrow-band UVB and IL-4.* PLoS One, 2011. **6**(5): p. e19806.
115. Løvendorf, M.B., et al., *MicroRNA-223 and miR-143 are important systemic biomarkers for disease activity in psoriasis.* Journal of dermatological science, 2014. **75**(2): p. 133-139.
116. Su, Y., et al., *Interleukin-17 receptor D constitutes an alternative receptor for interleukin-17A important in psoriasis-like skin inflammation.* Science immunology, 2019. **4**(36): p. eaau9657.
117. Sundarrajan, S., S. Lulu, and M. Arumugam, *Insights into protein interaction networks reveal non-receptor kinases as significant druggable targets for psoriasis.* Gene, 2015. **566**(2): p. 138-147.
118. Karlsson, T., et al., *Redistribution of LRIG proteins in psoriasis.* Journal of Investigative Dermatology, 2008. **128**(5): p. 1192-1195.
119. Sundberg, T.B., et al., *Development of chemical probes for investigation of salt-inducible kinase function in vivo.* ACS chemical biology, 2016. **11**(8): p. 2105-2111.
120. Gilhar, A., et al., *Fas pulls the trigger on psoriasis.* The American journal of pathology, 2006. **168**(1): p. 170-175.
121. Wu, C.-S., et al., *Integrative transcriptome sequencing identifies trans-splicing events with important roles in human embryonic stem cell pluripotency.* Genome research, 2014. **24**(1): p. 25-36.

122. Zhou, Y., et al., *Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.* Nature communications, 2019. **10**(1): p. 1523.
123. Xu, F., et al., *Long noncoding RNAs in autoimmune diseases.* Journal of biomedical materials research Part A, 2019. **107**(2): p. 468-475.
124. Fitch, E., et al., *Pathophysiology of psoriasis: recent advances on IL-23 and Th17 cytokines.* Current rheumatology reports, 2007. **9**(6): p. 461-467.
125. Glažar, P., P. Papavasileiou, and N. Rajewsky, *circBase: a database for circular RNAs.* Rna, 2014. **20**(11): p. 1666-1670.
126. Liang, D., et al., *The output of protein-coding genes shifts to circular RNAs when the pre-mRNA processing machinery is limiting.* Molecular cell, 2017. **68**(5): p. 940-954. e3.
127. Kramer, M.C., et al., *Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins.* Genes & development, 2015. **29**(20): p. 2168-2182.
128. Li, X., et al., *Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes.* Journal of molecular evolution, 2009. **68**(1): p. 56-65.
129. Yang, W., et al., *Possible formation of mitochondrial-RNA containing chimeric or trimeric RNA implies a post-transcriptional and post-splicing mechanism for RNA fusion.* PLoS One, 2013. **8**(10): p. e77016.
130. Guarnerio, J., et al., *Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations.* Cell, 2016. **165**(2): p. 289-302.
131. Chen, L.-L.J.N.r.M.c.b., *The biogenesis and emerging roles of circular RNAs.* Nature reviews Molecular cell biology, 2016. **17**(4): p. 205-211.

Appendix

CircRNAs differentially expressed between PP and NN skin in psoriasis

name	chrom	start	end	strand	type	PP	NN	log2(PP/NN)	p-value
hsa_skin_176158	chr11	36227082-36227084	36227428-36227430	+	3' B	20.63	59.07	-1.52	0.005
hsa_skin_017868	chr5	36982162-36982165	36986299-36986302	+	3' B	24.75	84.19	-1.77	0.005
hsa_skin_173421	chr15	72045723-72045725	72046634-72046636	-	5' B	15.71	42.25	-1.43	0.004
hsa_skin_102568	chr19	52383847-52383848	52385323-52385324	+	5' B	1.83	7.55	-2.05	0.005
hsa_skin_137810	chr20	3218514-3218518	3218992-3218996	+	5' B	2.98	21.99	-2.88	0.000
hsa_skin_028220	chr9	127444025-127444029	127445246-127445250	+	5' B	4.05	23.72	-2.55	0.002
hsa_skin_050166	chr9	4860124-4860125	4860901-4860902	+	5' B	15.69	42.72	-1.44	0.006
hsa_skin_039030	chr1	224189575-224189576	224190318-224190319	+	B	13.12	103.15	-2.97	0.000
hsa_skin_032293	chr10	45625950-45625951	45627105-45627106	-	B	9.80	14.89	-0.60	0.000
hsa_skin_006010	chr10	68644696-68644698	68647004-68647006	+	B	7.48	30.19	-2.01	0.001
hsa_skin_112375	chr11	92352094-92352096	92355403-92355405	+	B	3.54	16.56	-2.22	0.001
hsa_skin_150608	chr13	30630832-30630834	30631472-30631474	+	B	3.26	14.96	-2.20	0.000
hsa_skin_189139	chr17	73235473-73235476	73236993-73236996	+	B	4.05	17.35	-2.10	0.005
hsa_skin_192510	chr2	40428472	40430304	-	B	92.49	275.52	-1.57	0.005
hsa_skin_223866	chr22	32478979-32478980	32479274-32479275	+	B	3.17	34.37	-3.44	0.000
hsa_skin_059513	chr4	109462899-109462902	109463643-109463646	+	B	4.13	12.13	-1.56	0.003
hsa_skin_013327	chr5	73840477-73840481	73840758-73840762	+	B	17.45	44.27	-1.34	0.005
hsa_skin_130995	chr5	83537003-83537008	83542265-83542270	+	B	16.07	54.42	-1.76	0.008
hsa_skin_194732	chr8	38429681-38429684	38429948-38429951	-	B	2.47	20.78	-3.07	0.000
hsa_skin_194345	chrX	140783169-140783175	140784654-140784660	+	I	23.57	310.97	-3.72	0.000
hsa_skin_052271	chr1	152352401-152352423	152352626-152352648	-	I	1.43	9.92	-2.79	0.000
hsa_skin_194228	chr1	152354735-152354791	152354966-152355022	-	I	6.83	26.73	-1.97	0.000
hsa_skin_093700	chr1	152356347-152356353	152356497-152356503	-	I	1.43	11.77	-3.04	0.000
hsa_skin_098241	chr10	92450725-92450733	92451246-92451254	-	I	15.60	25.64	-0.72	0.000
hsa_skin_017093	chr17	39254341-39254346	39255167-39255172	-	I	2.85	14.16	-2.31	0.004
hsa_skin_088763	chr19	23831593-23831599	23833518-23833524	+	B	21.83	129.55	-2.57	0.000
hsa_skin_143837	chr2	113707872-113707876	113708437-113708441	-	I	1.46	14.89	-3.35	0.001
hsa_skin_142248	chr10	124942455-124942456	124943306-124943307	+	3' B	100.02	50.76	0.98	0.000
hsa_skin_169953	chr17	7576810-7576812	7576950-7576952	+	3' B	30.30	37.90	-0.32	0.010
hsa_skin_146560	chr16	16141172-16141173	16142942-16142943	+	5' B	27.17	12.48	1.12	0.000
hsa_skin_214394	chr17	40818151-40818157	40818475-40818481	-	5' B	29.05	20.63	0.49	0.003
hsa_skin_232176	chr7	139715928-139715933	139717012-139717017	-	5' B	36.79	14.87	1.31	0.000
hsa_skin_026246	chr1	151315572-151315574	151316509-151316511	-	B	40.16	30.11	0.42	0.001
hsa_skin_153897	chr10	68959803-68959805	68960247-68960249	+	B	77.69	33.06	1.23	0.000
hsa_skin_183503	chr12	122340750-122340754	122341695-122341699	-	B	181.70	141.33	0.36	0.005
hsa_skin_216799	chr13	52397231-52397234	52398194-52398197	-	B	37.60	6.77	2.47	0.000
hsa_skin_199788	chr14	71587569-71587570	71589369-71589370	+	B	53.33	14.32	1.90	0.000
hsa_skin_213436	chr16	346146-346150	347105-347109	-	B	25.78	12.18	1.08	0.000
hsa_skin_192421	chr17	20204329-20204333	20205909-20205913	+	B	71.48	28.28	1.34	0.000
hsa_skin_203265	chr17	83084936-83084937	83085322-83085323	+	B	61.75	24.55	1.33	0.000
hsa_skin_083378	chr2	32414764-32414769	32416159-32416164	+	B	27.21	17.12	0.67	0.000
hsa_skin_050529	chr2	37316235-37316238	37317178-37317181	-	B	204.16	144.49	0.50	0.006
hsa_skin_083065	chr4	36228580-36228583	36229644-36229647	-	B	680.04	415.55	0.71	0.001
hsa_skin_189369	chr6	134028379-134028380	134029720-134029721	-	B	18.41	11.27	0.71	0.002
hsa_skin_171562	chr8	127890586-127890589	127890996-127890999	+	B	37.36	17.79	1.07	0.000
hsa_skin_156445	chr9	34241183-34241184	34242107-34242108	+	B	8.79	2.43	1.85	0.000
hsa_skin_228653	chr10	92450736-92450738	92451257-92451259	-	I	5.71	1.82	1.65	0.000
hsa_skin_006690	chr11	65499641-65499650	65499860-65499869	+	I	13.62	11.38	0.26	0.008
hsa_skin_228518	chr11	65504856-65504859	65505049-65505052	+	I	10.72	5.15	1.06	0.002
hsa_skin_060631	chr12	52645124-52645131	52645271-52645278	-	I	12.85	4.19	1.62	0.002
hsa_skin_130118	chr20	62276203-62276204	62276962-62276963	+	I	12.98	9.89	0.39	0.001

* type: Whether the candidate originated from annotation boundary. B:boundary, I: interior, 5' B: 5' boundary, 3' B: 3' boundary. PP: average expression level in PP. NN: average expression level in NN. log2(PP/NN): log2 fold change of PP vs. NN.

Chimeric RNAs differentially expressed between tumor and normal tissues in prostate cancer

name	PCa	NN	log2(PCa/NN)	p-value
chr1:100017813-100017815_+|chr1:100049906-100049908_+	1.55	0.50	1.63	0.01
chr1:156286631-156286634_-|chr1:156279033-156279036_-	16.50	5.90	1.48	0.00
chr1:168243067-168243069_+|chr1:168291157-168291159_+	0.65	1.60	-1.30	0.00
chr1:187418001-187418003_+|chr1:187443625-187443627_+	0.85	2.15	-1.34	0.00
chr1:205659275-205659277_-|chr1:205646805-205646807_-	0.50	3.25	-2.70	0.00
chr1:205659488_-|chr1:205623891_-	8.90	34.25	-1.94	0.00
chr1:222615822-222615825_+|chr1:222621155-222621158_+	3.60	15.85	-2.14	0.00
chr1:925516-925520_-|chr1:923460-923464_-	1.15	3.40	-1.56	0.00
chr11:114450166-114450167_+|chr11:114466856-114466857_+	3.85	0.95	2.02	0.00
chr11:114517711-114517713_+|chr11:114526621-114526623_+	6.20	1.20	2.37	0.00
chr11:61582204-61582206_-|chr11:61556206-61556208_-	1.00	11.65	-3.54	0.00
chr11:92225173-92225175_+|chr11:92352094-92352096_+	2.60	8.15	-1.65	0.00
chr11:92309823-92309826_+|chr11:92352093-92352096_+	1.00	3.55	-1.83	0.01
chr12:50456716-50456724_+|chr4:110523085-110523093_-	1.90	0.55	1.79	0.00
chr13:31946178-31946181_+|chr13:32078830-32078833_+	9.95	3.50	1.51	0.01
chr13:95302064-95302066_+|chr13:95310833-95310835_+	7.05	2.00	1.82	0.00
chr15:69570902-69570904_+|chr15:69628736-69628738_+	5.45	1.20	2.18	0.00
chr15:69570902-69570904_+|chr15:69671723-69671725_+	5.35	1.10	2.28	0.01
chr15:81362035-81362036_+|chr15:81380207-81380208_+	7.30	2.90	1.33	0.00
chr17:3664678-3664680_-|chr17:3660326-3660328_-	0.95	2.40	-1.34	0.00
chr17:58324624-58324627_+|chr17:58337444-58337447_+	0.65	3.40	-2.39	0.00
chr17:58518932-58518933_+|chr17:58525699-58525700_+	0.25	1.55	-2.63	0.00
chr19:50851337-50851340_+|chr19:50856237-50856240_+	6.05	0.40	3.92	0.00
chr2:20651488-20651490_-|chr2:20645504-20645506_-	1.50	3.40	-1.18	0.00
chr2:210192801-210192803_-|chr2:210154609-210154611_-	1.30	3.20	-1.30	0.00
chr2:218033819-218033822_-|chr2:217991056-217991059_-	2.40	9.35	-1.96	0.00
chr20:59942509-59942513_+|chr20:59949250-59949254_+	0.85	1.70	-1.00	0.00
chr21:36698908-36698909_-|chr21:36643010-36643011_-	2.55	1.00	1.35	0.00
chr21:41507948-41507952_-|chr21:38445620-38445624_-	13.40	0.30	5.48	0.00
chr22:23527560-23527564_+|chr22:23536967-23536971_+	1.55	3.15	-1.02	0.01
chr3:52291108-52291111_+|chr3:52316451-52316454_+	4.05	2.00	1.02	0.01
chr3:53494417-53494419_+|chr3:53495044-53495046_+	16.05	1.50	3.42	0.00
chr4:164801513-164801514_+|chr4:164805605-164805606_+	5.95	2.30	1.37	0.01
chr4:169271607_-|chr4:169269307_-	3.35	0.75	2.16	0.00
chr5:115634485-115634488_-|chr5:115625980-115625983_-	0.50	1.40	-1.49	0.01
chr6:116060998-116060999_-|chr6:116060594-116060595_-	2.25	0.50	2.17	0.00
chr6:44731367-44731368_+|chr6:44829586-44829587_+	1.15	27.65	-4.59	0.00
chr6:99175100-99175111_+|chr4:103086529-103086540_-	1.30	5.50	-2.08	0.00
chr7:12094650-12094653_-|chr7:11636959-11636962_-	0.80	2.40	-1.58	0.00
chr7:155003584-155003588_-|chr7:154998782-154998786_-	2.45	1.00	1.29	0.00
chr7:156471152-156471154_-|chr7:156441378-156441380_-	7.20	0.60	3.58	0.00
chr7:89882352-89882355_-|chr7:89694367-89694370_-	2.60	0.70	1.89	0.01
chr7:99971745-99971746_-|chr7:99923603-99923604_-	6.75	15.30	-1.18	0.00
chr9:111038649-111038654_-|chr9:110973556-110973561_-	0.80	2.30	-1.52	0.01
chrX:119150934-119150936_-|chrX:119128059-119128061_-	0.60	2.90	-2.27	0.00
chrX:119150936_-|chrX:119123652_-	1.45	3.10	-1.10	0.00
chrX:125185641-125185644_-|chrX:124963816-124963819_-	1.25	4.50	-1.85	0.00
chrX:71212920-71212926_+|chrX:71223688-71223694_+	5.50	0.45	3.61	0.00
chrY:12662525-12662528_-|chrY:12662124-12662127_-	2.80	1.20	1.22	0.01

***** name: name of chimeric RNAs are formed by their genomic loci, i.e., 5'-end junction|3'-end junction. PCa: average expression level in tumor tissue of prostate cancer (PCa). NN: average expression level in normal tiisues (NN). log2(PCa/NN): log2 fold change of PCa vs. NN.

Chimeric RNAs significantly up-regulated between PP and NN skin in psoriasis

name	PP	NN	log2(PP/NN)	p-value
chr1:114585534-114585536_-|chr1:114581390-114581392_-	1.90	0.57	1.74	0.00
chr1:156705436-156705439_-|chr19:30548305-30548308_+	5.95	0.76	2.97	0.00
chr1:187284711-187284713_+|chr3:100555262-100555264_+	2.43	0.76	1.67	0.00
chr1:221146521-221146523_+|chr12:52680386-52680388_-	4.19	0.90	2.21	0.00
chr1:46683297-46683301_-|chr1:46665362-46665366_-	1.62	0.57	1.50	0.00
chr10:117290183-117290187_-|chr10:117270441-117270445_-	2.05	0.76	1.43	0.00
chr10:130588015-130588018_-|chr1:153615019-153615022_-	3.00	0.67	2.17	0.00
chr11:119018284-119018299_-|chr5:116052284-116052299_-	7.14	2.71	1.40	0.01
chr11:18588993-18588995_+|chr11:18601089-18601091_+	2.33	0.86	1.44	0.00
chr11:6603403-6603405_-|chr11:6592823-6592825_-	2.57	1.10	1.23	0.00
chr12:12338419-12338421_-|chr12:12322554-12322556_-	2.86	0.57	2.32	0.00
chr12:52208873-52208875_+|chr12:52223443-52223445_+	3.05	0.76	2.00	0.00
chr13:31946179-31946181_+|chr13:32038427-32038429_+	3.05	1.14	1.42	0.00
chr15:79560455_-|chr13:113585965_+	3.67	1.43	1.36	0.00
chr16:82170598_-|chr16:82164194_-	5.76	1.38	2.06	0.00
chr17:75131183-75131184_+|chr17:75148435-75148436_+	1.57	0.57	1.46	0.00
chr17:82496865-82496871_+|chr17:82563350-82563356_+	1.81	0.90	1.00	0.00
chr18:63842217-63842218_+|chr18:63891434-63891435_+	21.81	6.71	1.70	0.00
chr2:83058905-83058909_-|chr10:5499269-5499273_-	8.48	3.86	1.14	0.00
chr20:3822289-3822293_+|chr20:3854553-3854557_+	2.81	0.95	1.56	0.00
chr22:41680243-41680244_-|chr22:41667009-41667010_-	2.19	0.67	1.72	0.00
chr22:50544785-50544788_-|chr22:50542898-50542901_-	12.29	0.57	4.43	0.00
chr5:141651534-141651540_-|chr5:141651115-141651121_-	15.48	0.38	5.34	0.00
chr5:160929220-160929223_-|chr5:160785802-160785805_-	6.57	2.00	1.72	0.00
chr5:75047896-75047898_-|chr5:75030038-75030040_-	28.71	5.24	2.45	0.00
chr5:75053513-75053514_-|chr5:75048036-75048037_-	11.48	2.33	2.30	0.00
chr5:75053514_-|chr5:75030038_-	4.67	0.57	3.03	0.00
chr5:75053947-75053948_-|chr5:75048036-75048037_-	6.29	2.24	1.49	0.00
chr5:95838441-95838444_+|chr5:95856875-95856878_+	11.05	2.86	1.95	0.00
chr6:28080968-28080970_+|chr6:28085478-28085480_+	5.67	2.10	1.44	0.00
chr6:53929576-53929579_+|chr6:53982755-53982758_+	2.48	0.76	1.70	0.00
chr6:7052092-7052097_+|chr6:7176652-7176657_+	1.67	0.62	1.43	0.00
chr9:135938219-135938224_-|chr9:135925185-135925190_-	2.76	0.62	2.16	0.00
chrX:153446008-153446010_-|chrX:153441548-153441550_-	2.62	0.67	1.97	0.00

* PP: average expression level in PP. NN: average expression level in NN. log2(PP/NN): log2 fold change of PP vs. NN.

Chimeric RNAs significantly down-regulated between PP and NN skin in psoriasis

name	PP	NN	log2(PP/NN)	p-value
chr1:109445925-109445929_+|chr14:20343349-20343353_-	0.00	14.86	-7.22	0.00
chr1:152314598-152314600_-|chrX:121147118-121147120_-	0.00	9.29	-6.54	0.00
chr1:152355667-152355670_-|chr9:74121604-74121607_-	2.48	9.48	-1.94	0.00
chr1:153262111-153262114_+|chr13:56182412-56182415_+	0.00	4.95	-5.63	0.00
chr1:160406170-160406175_-|chr1:160340293-160340298_-	1.95	12.62	-2.69	0.01
chr1:224175417-224175420_-|chr8:101178386-101178389_-	0.67	1.76	-1.40	0.00
chr1:9871186-9871187_-|chr1:9773394-9773395_-	0.38	1.95	-2.36	0.00
chr11:62518405-62518407_-|chr1:22373684-22373686_+	0.57	1.76	-1.62	0.01
chr11:62665550-62665554_+|chrX:9808117-9808121_-	0.00	6.95	-6.12	0.00
chr12:109879105-109879108_+|chr12:109903023-109903026_+	1.00	3.00	-1.58	0.01
chr12:24562822-24562825_-|chr12:24407565-24407568_-	0.43	2.33	-2.44	0.00
chr13:104718251-104718253_-|chr12:6510358-6510360_+	0.00	23.24	-7.86	0.00
chr13:51827459-51827462_-|chr1:153614593-153614596_-	0.00	4.38	-5.45	0.00
chr13:97976327-97976333_+|chr19:35497440-35497446_-	0.00	2.29	-4.51	0.00
chr14:20343245-20343247_-|chr12:83269668-83269670_+	0.00	4.76	-5.57	0.00
chr15:52934997-52934999_+|chr1:152350840-152350842_-	0.00	6.00	-5.91	0.00
chr17:18040714-18040718_-|chr17:18039048-18039052_-	1.33	3.00	-1.17	0.00
chr17:40822043-40822045_-|chr3:23480087-23480089_-	0.00	4.48	-5.48	0.00
chr17:40822471-40822476_-|chr10:31636001-31636006_+	0.00	6.43	-6.01	0.00
chr17:40822478-40822480_-|chr12:117849353-117849355_+	0.00	3.95	-5.30	0.00
chr18:31355790-31355793_+|chr5:82146722-82146725_-	0.00	7.29	-6.19	0.00
chr19:48298676-48298680_+|chr18:31127978-31127982_-	0.00	3.86	-5.27	0.00
chr2:224535863-224535865_+|chr1:152357404-152357406_-	0.00	3.52	-5.14	0.00
chr2:27672859-27672861_+|chr12:52674976-52674978_-	0.00	6.86	-6.10	0.00
chr2:64488838-64488844_+|chr2:64551438-64551444_+	0.24	5.00	-4.39	0.00
chr2:64500920-64500925_+|chr2:64551440-64551445_+	0.48	5.67	-3.57	0.00
chr20:13221143-13221144_+|chr20:13270502-13270503_+	0.33	1.81	-2.44	0.00
chr21:45021667-45021669_+|chr12:52677457-52677459_-	0.00	6.57	-6.04	0.00
chr22:20464262-20464267_-|chr20:17962828-17962833_-	0.00	15.19	-7.25	0.00
chr3:156490801-156490804_-|chr11:65501034-65501037_+	0.00	5.43	-5.76	0.00
chr4:168324837-168324841_-|chr4:168311364-168311368_-	2.67	22.48	-3.08	0.00
chr5:149867232-149867234_+|chr9:35657877-35657879_-	0.00	193.95	-10.92	0.00
chr5:45824570-45824573_+|chr14:20343254-20343257_-	0.00	2.90	-4.86	0.00
chr5:95055283-95055285_+|chr17:40818455-40818457_-	0.00	9.52	-6.57	0.00
chr7:94369097-94369100_+|chr9:35657988-35657991_-	0.00	39.19	-8.61	0.00
chr7:94369098-94369100_+|chr9:35657988-35657990_-	0.00	4.76	-5.57	0.00
chr8:15900413-15900418_+|chr17:40819701-40819706_-	0.00	20.90	-7.71	0.00
chr9:35657762-35657764_-|chr7:50628592-50628594_+	0.00	13.76	-7.10	0.00
chrX:120315742-120315745_-|chrX:120304490-120304493_-	0.33	2.52	-2.92	0.00

* PP: average expression level in PP. NN: average expression level in NN. log2(PP/NN): log2 fold change of PP vs. NN.