Review

Transcriptional-Readthrough RNAs Reflect the Phenomenon of "A Gene Contains Gene(s)" or "Gene(s) within a Gene" in the Human Genome, and Thus Are Not Chimeric RNAs

Yan He [1],*, Chengfu Yuan [2],*, Lichan Chen [3], Mingjuan Lei [3], Lucas Zellmer [4], Hai Huang [5],* and Dezhong Joshua Liao [1,6],*

[1]   Key Lab of Endemic and Ethnic Diseases of the Ministry of Education of China in Guizhou Medical University, Guiyang 550004, Guizhou, China
[2]   Department of Biochemistry, China Three Gorges University, Yichang City 443002, Hubei, China
[3]   Hormel Institute, University of Minnesota, Austin, MN 55912, USA; lchen@hi.umn.edu (L.C.); mlei@hi.umn.edu (M.L.)
[4]   Masonic Cancer Center, University of Minnesota, 435 E. River Road, Minneapolis, MN 55455, USA; lzellmer@umn.edu
[5]   School of Clinical Laboratory Science, Guizhou Medical University, Guiyang 550004, Guizhou, China
[6]   Department of Pathology, Guizhou Medical University Hospital, Guiyang 550004, Guizhou, China
[*]   Correspondence: annieheyan@gmc.edu.cn (Y.H.); yuancf46@ctgu.edu.cn (C.Y.); huanghai828@gmc.edu.cn (H.H.); djliao@gmc.edu.cn (D.J.L.)

Abstract: Tens of thousands of chimeric RNAs, i.e., RNAs with sequences of two genes, have been identified in human cells. Most of them are formed by two neighboring genes on the same chromosome and are considered to be derived via transcriptional readthrough, but a true readthrough event still awaits more evidence and *trans*-splicing that joins two transcripts together remains as a possible mechanism. We regard those genomic loci that are transcriptionally read through as unannotated genes, because their transcriptional and posttranscriptional regulations are the same as those of already-annotated genes, including fusion genes formed due to genetic alterations. Therefore, readthrough RNAs and fusion-gene-derived RNAs are not chimeras. Only those two-gene RNAs formed at the RNA level, likely via *trans*-splicing, without corresponding genes as genomic parents, should be regarded as authentic chimeric RNAs. However, since in human cells, procedural and mechanistic details of *trans*-splicing have never been disclosed, we doubt the existence of *trans*-splicing. Therefore, there are probably no authentic chimeras in humans, after readthrough and fusion-gene derived RNAs are all put back into the group of ordinary RNAs. Therefore, it should be further determined whether in human cells all two-neighboring-gene RNAs are derived from transcriptional readthrough and whether *trans*-splicing truly exists.

Keywords: chimeric RNA; fusion gene; transcriptional readthrough; *cis*-splicing; *trans*-splicing; reverse transcription; polymerase chain reactions

1. Introduction

In 2007, the ENCODE (The Encyclopedia of DNA Elements) pilot project reported its identification and analysis of functional elements in 1% of the human genome [1,2]. In this report, it was estimated that RNAs from 65% of human genes are fused to another gene's RNA to form a new RNA that contains sequences of two genes and is called "chimeric RNA" or chimera. Interestingly, most of these chimeras are formed by RNAs from two neighboring genes on the same chromosome [1,2]. Since this ENCODE

report, high-throughput RNA sequencing technology has swiftly spread over all biomedical research and has led to the identification of tens of thousands of chimeric RNAs and other forms of noncolinear RNAs [3,4], as summarized by us previously [5,6]. This number is astonishing, considering that the human genome contains only about 20,000 protein-coding genes [7–14], although the number of genes may be much larger if noncoding genes are included and if readthrough genomic loci are considered as newly-identified genes and are included, as we have suggested before [6]. Many fusion RNAs derived from fusion genes formed due to genetic alterations [15–18], seen mainly in genetic diseases and tumors [19–24], have also been identified and are, peculiarly, renamed as chimeras, as they also contain sequences of two genes [5,6]. This reclassification of fusion RNAs to tout their novelty and importance seems unnecessary, as they belong to an ancient research sphere the importance of which has already been recognized for roughly six decades, since 1959, when the Philadelphia chromosome and its-encoded fusion genes were identified [25–30].

Despite the sheer number of chimeras identified, unfortunately neither the ENCODE's report nor most other relevant studies have disclosed the procedural and mechanistic details on how most chimeras might be formed. Since it is well known that transcription of genes occasionally does not stop at the canonical termination site but instead reads into the downstream gene [31–33], it is speculated that transcriptional readthrough may be a major mechanism for those two-neighboring-gene RNAs, albeit other mechanisms remain possible, such as *trans*-splicing that splices two RNAs into one. While the "readthrough" assumption is very reasonable and has been widely accepted, there is little irrefutable experimental proof to validate that a readthrough has indeed happened during formation of most two-neighboring-gene RNAs, and only a few such RNAs have received tenable evidence, because detection of a not-yet-spliced precursor transcript is difficult [34,35]. This in turn is because transcription is a transient procedure and splicing of the resulting transcript to a mature RNA ensues nearly at the same time as the start of transcription and is terminated almost as transcription is finished, making it difficult for researchers to determine what events transpired during this short spell [35]. Some relevant questions, such as why the transcription does not end as it should at the upstream gene, still remain inscrutable as well for most such RNAs. Moreover, although "chimeric RNA" means that an RNA consists of sequences of two different genes, the reality is that it has never been lucidly defined, and a variety of noncolinear RNAs have all been called "chimeric RNA" [6]. This is, in turn, because "what is a gene" remains an unanswered question, and many researchers, including us, consider that "gene" should be redefined at the RNA level in the post ENCODE era [5,6,36–40], and thus consider two-RNA RNAs as chimeras as well. All these problems have made many researchers befuddled and have gutted not only research on authentic chimeric RNAs per se but also research into ordinary colinear RNAs.

In this perspective article, we elaborate on our contemplation and reflection on the designation and classification of noncolinear RNAs, including those RNAs that contain sequences of two genes or contain transcripts from both DNA strands of a gene, for our contemporaries in the RNA research bailiwick to consider and debate. Only messenger RNAs (mRNAs) and long, i.e., larger than 200 nucleotides [41–43], noncoding RNAs are of concern, while those short RNAs that are generally esteemed to function as regulatory elements of genes are left out, in part because we are not aware of any chimeric short regulatory RNAs, which are typically 20 nucleotides in length. We refer to the DNA strand that harbors the gene as the Watson strand and its opposite strand of the DNA double helix as the Crick strand, to avoid confusion, since in the literature different researchers define Watson and Crick strands differently, whereas the RNAs transcribed from the Watson and Crick strands of a gene are referred to as sense and antisense, respectively.

2. There Are Different Types of Long Noncolinear RNAs

While probably less than 5% of the human genes (including all mitochondrial genes) contain only a single exon, and, thus, their transcripts do not need to undergo *cis*-splicing to produce mature RNA, transcripts from over 95% of human genes need to be *cis*-spliced to remove intron(s) and

to join exons together for the formation of mature RNAs [34,35,38,44]. Because of the removal of intron sequence(s), mature RNAs are no longer as continuous as the parental genes' sequences, but they still have the same 5'-to-3' orientation and, thus, are colinear, which in our opinion includes circular RNAs a well. However, there remains a large number of noncolinear RNAs, but mainly in evolutionarily-lower organisms, such as bacteria and other prokaryotic or unicellular eukaryotic organisms [45–50]. Nevertheless, noncolinear mature RNAs have also been reported in the cells of human, mouse, and rat origins, which to our knowledge include the following types:

1. RNAs with sequences of two different genes, which occur in two separate ways, i.e., (1) the two genes are adjacent to each other on the same chromosome; and (2) the two genes are located on two different chromosomes. Theoretically, there should also be many RNAs in which the two genes are on the same chromosome but are far away from each other, too far away for a transcriptional-readthrough to occur, but, unfathomably, there are few, if any, such RNAs reported in the literature, to our knowledge.
2. RNAs that contain repeats of one or more exons [51–53], such as some RNA variants of human estrogen receptor α (*ERα*) [54–56], rat *Cot* [57–59], rat *Sns* [60], and rat *Sa* [61].
3. RNAs that contain both sense and antisense sequences of the same gene, with the drosophila mdg4 mRNA variant being best studied [62,63].

A caveat needs to be given that many genetic alterations, as often seen in genetic diseases and tumors [19–24], can also lead to the formation of the abovementioned three categories of RNA in pathological situations. Indeed, some genetic alterations can cause fusion of two genes into one [15–18], and the fusion gene can be transcribed to two-gene RNAs in the same way as other genes [5,6]. Similarly, some genetic alterations can also result in RNAs with duplicated exons or with antisense sequences. However, the RNAs caused by these genetic alterations are still colinear and, thus, are excluded, because they have a corresponding gene as a genomic parent and are produced in the same way as all colinear RNAs from all genes.

3. *Trans*-Splicing Remains as a Possible Mechanism for Formation of Chimeric and Other Noncolinear RNAs

Besides *cis*-splicing that is a biochemical reaction using one single RNA molecule as the substrate and producing one single mature RNA as the product, there is also *trans*-splicing, which is another biochemical reaction that uses two RNA molecules as the substrates but produces only one single mature RNA as the product [5,6,64,65]. Although *trans*-splicing is a common event in some unicellular organisms, in some mitochondria of evolutionarily-lower eukaryotes, and in chloroplasts of some plants [45–50], it is also considered by many researchers to occur as a mechanism for the formation of some chimeric RNAs and other forms of noncolinear RNAs in evolutionarily-higher animals [15,66–74]. For example, a human *KLK4* RNA [75] was found to contain both sense and antisense sequences, and some RNA variants of *ERα* [54–56] and *Sp1* [76,77] were reported to bear duplicated exons. In normal human endometrium and in some human uterine tumors, a chimeric RNA involving a *JAZF1* sequence from 7p15 and a *JJAZ1* sequence from 17q11 has been reported to be derived via a *trans*-splicing like mechanism [78–80], although it has been known that these uterine tumors bear a *JAZF1-JJAZ1* fusion gene at high frequencies [78,81–85]. There are other reported chimeric RNAs in human cells that are not associated with a fusion gene, such as the *CCND1-Trop2* [25,86], *FAS-ERα* [87], *CYP3A43-CYP3A4* [88], *CYP3A43-CYP3A5* [88] and *Yq12-CDC2L2* [89] RNAs as well as an *ACTAT1* RNA that contains sequences from both chromosomes 1 and 7 [90–93]. A more complicated case is the seven mouse *Msh4* RNA variants, which together involve sequences from a total of four different chromosomes, and some of which involve both sense and antisense sequences of one of the genomic loci [94]. However, although these noncolinear RNAs were considered to be derived from a *trans*-splicing or a *trans*-splicing-like mechanism, unimpeachable evidence for a *trans*-splicing event in the formation of these RNAs, and the procedural and mechanistic details of the splicing, are still

lacking. After a decade since the initial reports on most of these RNAs, such as the *JAZF1-JJAZ1* RNA, we do not possess information about the procedural and mechanistic details of their *trans*-splicing to corroborate that they are really formed at the RNA level and are not technical artifacts or are not transcribed from a fusion gene. On the other hand, more publications continue emerging to report [95] or summarize [3,69,70] such *trans*-splicing related chimeras or other noncolinear RNAs. Moreover, many bioinformatic experts are establishing different algorithms to cull chimeras from different sets of high-throughput sequencing data [96–104], although all these data sets contain many spurious sequences, as we and others have pointed out [5,6,64,105–135]. This situation is worrisome to us.

Although it should be a requirement to show more-concrete evidence for the true existence of *trans*-splicing in evolutionarily-higher animal species, such as in the human, rat, and mouse, there are technical constraints hindering such studies [6,105]. For example, splicing is initiated and finished too quickly to study its detail, as aforementioned. In addition, the reported detection of the abovementioned RNAs all involved reverse transcription (RT) and polymerase chain reactions (PCR), which are techniques that easily create spurious results, as we and others have repeatedly described before, due to template-switching, mis-priming, self-priming, DNA or complementary DNA (cDNA) damage, and PCR-reconditioning, among other reasons [5,6,64,105–135]. Therefore, approaches without involvement of RT and PCR are needed to minimize technical artifacts for indisputable evidence and to obtain procedural and mechanistic details of the presumed *trans*-splicing. RNA protection assay [136–138], or the cDNA protection assay established by us [105], is currently the best approach for this purpose, to our knowledge.

## 4. Some Human Genomic Loci Are Crowded Gene Habitats

In the human genome, genes are not evenly dispersed over chromosomal DNA. Some genomic loci are very crowded gene habitats, such as the 14q23.3-24.1 and 2q21.1 chromosomal regions (Figure 1), while other genomic regions harbor very few genes. In those crowded loci, "a gene contains gene(s)" or "gene(s) within a gene" is a common phenomenon [38]. For example, both the Watson and Crick strands of the *GPNH* gene or the *POTEI* gene encode many other genes, making the *GPNH* or *POTEI* a readthrough gene whose precursor transcript contains many other genes; thus, both are examples of "a gene contains gene(s)" or "gene(s) within a gene" (Figure 1). The genes within the *GPNH* or *POTEI* include not only protein-coding ones but also noncoding ones and pseudogenes, and some of them have until now not yet been characterized and, thus, are temporarily annotated with "LOC" (stands for Locus) and a number (Figure 1). Therefore, the precursor transcript of the *GPNH* or *POTEI* gene can be considered as a readthrough one that spans over many genes, meaning that readthrough can occur to multiple, and not just two, consecutive genes in a genomic locus, although the sequences of the inside genes may be lopped off during *cis*-splicing and, thus, may not occur in a *GPNH* or *POTEI* RNA variant.

To our knowledge, the *CNTNAP2* (located at 7q35–36.1) and *PTPRD* (located at 9p24.1–9p23) genes, both being longer than 2.3 megabase-pairs, are among the largest genes in the human genome, while most other genes are smaller than one-tenth of this size. This means that a single transcription can read through at least 2.3 mega-nucleotides. Therefore, theoretically, transcription can also go through a genomic locus that contains several genes as long as it, for some reason, does not stop at a canonical transcription-termination site and as long as the transcription-distance is within 2.3 mega-nucleotides. Actually, there hitherto has been no evidence showing that a transcription cannot go beyond 2.3 mega-nucleotides. However, what is still inexplicable to us is that, to our knowledge, there has not been any mature RNA found known to possess sequences of three or more chromosomal genes, although we have found RNAs with sequences from three or four mitochondrial genes in some databases of expression sequence tags [65]. For instance, the NCBI (National Center for Bioinformation of the United States) database shows that on the minus strand of the human 6p24.3 region, the *BLOC1S5* gene and its downstream gene *TXNDC5* together produce a *BLOC1S5-TXNDC5* RNA, while the *BLOC1S5* and its upstream gene *EEF1E1* together produce a *EEF1E1-BLOC1S5* RNA (Figure 2 and

Table 1). However, no RNA containing sequences of all three genes, i.e., no *EEF1E1-BLOC1S5-TXNDC5* RNA, has been reported so far. This conundrum, i.e., why there has not been a three-gene RNA reported, is bewitching and awaits exploration.
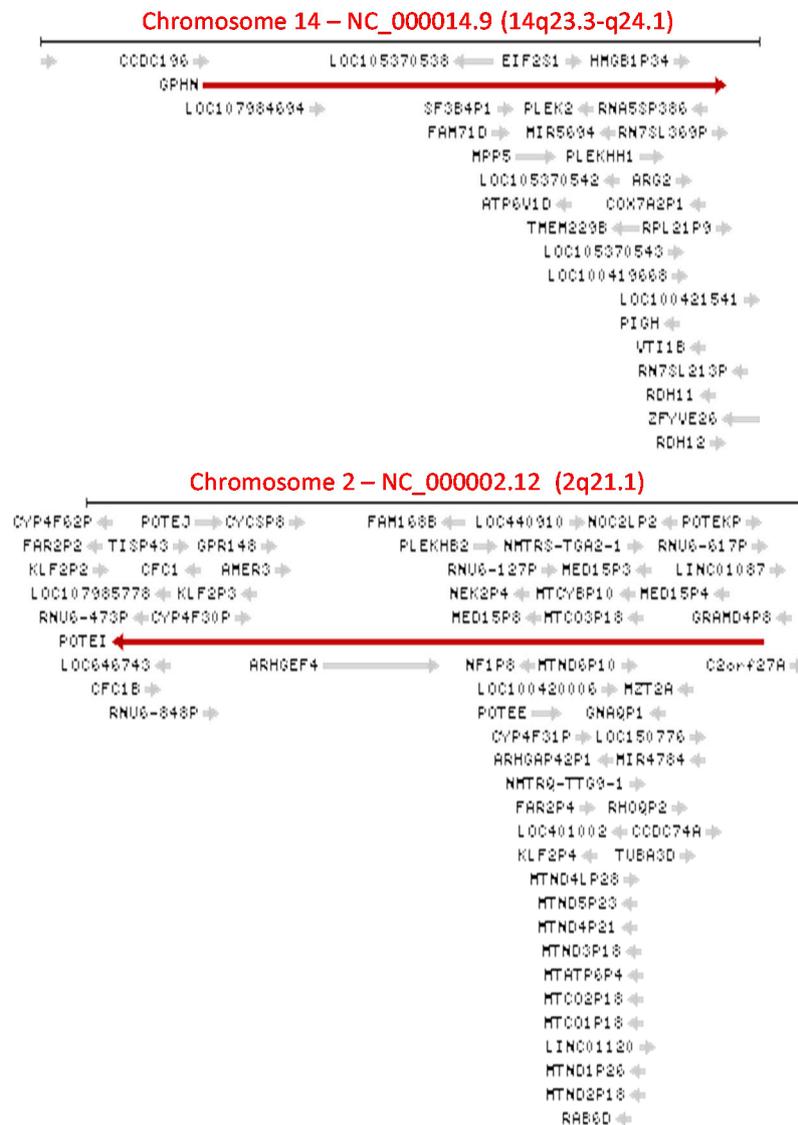
Figure 1. Images copied from the National Center for Bioinformation of the United States (NCBI) database illustrating that some human genomic loci are crowded gene habitats. Within the *GPHN* gene (the long red arrow in the top image) on the plus DNA strand (arrow to the right) of the 14q23.3–24.1 region, there are also many other genes encoded not only by the same plus strand (short grey arrows to the right) but also by the minus strand (short grey arrows to the left). Similarly, within the *POTEI* gene (the long red arrow in the bottom image) on the minus DNA strand (arrow to the left) of the 2q21.1 region, there are also many other genes encoded not only by the same minus strand (short grey arrows to the left) but also by the plus strand (short grey arrows to the right). Some of these genes are temporarily annotated with "LOC" (locus) and a number, since they have not yet been characterized.
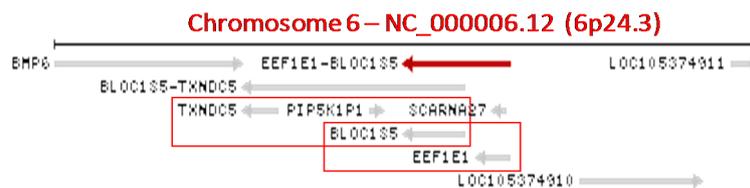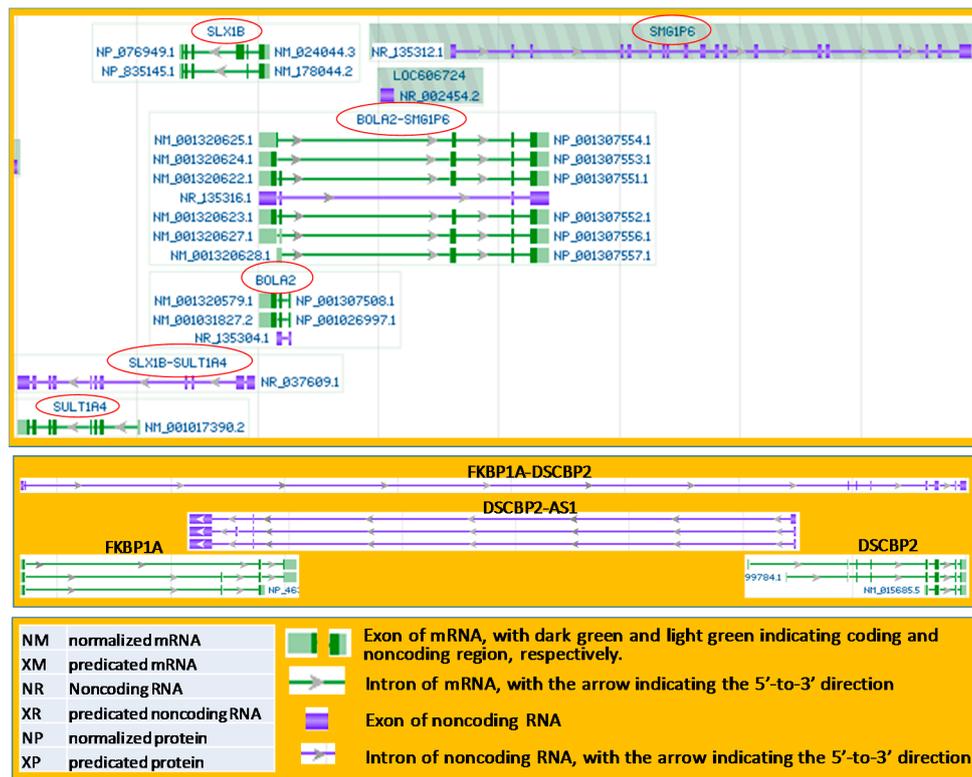
Figure 2. An image copied from the NCBI database illustrating that the *BLOC1S5* gene and its upstream gene *EEF1E1* on the minus strand of the 6p24.3 region together produce an *EEF1E1-BLOC1S5* RNA (red arrow to the left), while it and its downstream gene *TXNDC5* together produce a *BLOC1S5-TXNDC5* RNA (the long grey arrow to the left). Note that there is no *EEF1E1-BLOC1S5-TXNDC5* RNA shown in the image.

Table 1. Some two-neighboring-gene RNAs of the human origin documented in the NCBI database.

Name	Gene ID *	Location	Coding or not	Name	Gene ID	Location	Coding or not
MROH7-TTC4	100527960	1p32.3	noncoding	DNAAF4-CCPG1	100533483	15q21.3	noncoding
GJA9-MYCBP	100527950	1p34.3	noncoding	ST20-MTHFS	100528021	15q25.1	coding
CENPS-CORT	100526739	1p36.22	both **	C15orf38-AP3S2	100526783	15q26.1	coding
PMF1-BGLAP	100527963	1q22	coding	SLX1A-SULT1A3	100526830	16p11.2	noncoding
TSNAX-DISC1	100303453	1q42.2	noncoding	SLX1B-SULT1A4	100526831	16p11.2	noncoding
HSPE1-MOB4	100529241	2q33.1	coding	BOLA2-SMG1P6	107282092	16p11.2	both
ABHD14A-ACY1	100526760	3p21.2	coding	PKD1P6-NPIPP1	105369154	16p13.11	noncoding
ARPC4-TTLL3	100526693	3p25.3	coding	CORO7-PAM16	100529144	16p13.3	coding
FAM47E-STBD1	100631383	4q21.1	coding	CKLF-CMTM1	100529251	16q21	coding
TMED7-TICAM2	100302736	5q22.3	coding	TVP23C-CDRT4	100533496	17p12	both
CNPY3-GNMT	107080644	6p21.1	both	RNASEK-C17orf49	100529209	17p13.1	noncoding
RPS10-NUDT3	100529239	6p21.31	coding	TNFSF12-TNFSF13	407977	17p13.1	coding
PPT2-EGFL8	100532746	6p21.32	noncoding	SENP3-EIF4A1	100533955	17p13.1	noncoding
ATP6V1G2-DDX39B	100532737	6p21.33	noncoding	RAD51L3-RFFL	100529207	17q12	noncoding
MSH5-SAPCD1	100532732	6p21.33	noncoding	PTGES3L-AARSD1	100885850	17q21.31	coding
BLOC1S5-TXNDC5	100526836	6p24.3	noncoding	NME1-NME2	654364	17q21.33	both
EEF1E1-BLOC1S5	100526837	6p24.3	noncoding	TBC1D3P1-DHX40P1	653645	17q23.1	noncoding
URGCP-MRPS24	100534592	7p13	coding	TEN1-CDK3	100529145	17q25.1	noncoding
ATP5J2-PTCD1	100526740	7q22.1	coding	RPL17-C18orf32	100526842	18q21.1	coding
C7orf55-LUC7L2	100996928	7q34	coding	PPAN-P2RY11	692312	19p13.2	coding
C10orf32-AS3MT	100528007	10q24.32	noncoding	RAB4B-EGLN2	100529264	19q13.2	noncoding
TMX2-CTNND1	100528016	11q12.1	noncoding	MIA-RAB4B	100529262	19q13.2	noncoding
KCNK4-TEX40	106780802	11q13.1	noncoding	FKBP1A-SDCBP2	100528031	20p13	noncoding
RBM14-RBM4	100526737	11q13.2	coding	SYS1-DBNDD2	767557	20q13.12	noncoding
HSPB2-C11orf52	100528019	11q23.1	noncoding	SLMO2-ATP5E	100533975	20q13.32	noncoding
BLOC1S1-RDH5	100528022	12q13.2	noncoding	STX16-NPEPL1	100534593	20q13.32	noncoding
ZNF664-RFLNA	100533183	12q24.31	coding	SPECC1L-ADORA2A	101730217	22q11.23	noncoding
BCL2L2-PABPN1	100529063	14q11.2	coding	PIR-FIGF	100532742	Xp22.2	noncoding
CHURC1-FNTB	100529261	14q23.3	coding	RPL36A-HNRNPH2	100529097	Xq22.1	coding
SERF2-C15orf63	100529067	15q15.3	noncoding				

*: "Gene ID" means gene identification number. **: "Both" means that some RNA variant(s) are protein-coding while some other(s) are noncoding.

As another situation of the crowdedness of genomic loci, occasionally, both the plus and minus strands of the same genomic locus can produce RNAs that contain two genes' sequences. For example, the plus strand of the human 16p11.2 region produces the *BOLA2-SMG1P6* RNA while the minus strand produces the *SLX1B-SULT1A4* RNA, as shown in the NCBI database (Figure 3, top panel). Moreover, when the opposite DNA strand does not encode gene(s), it may still produce antisense RNA(s) (Figure 3, middle panel).
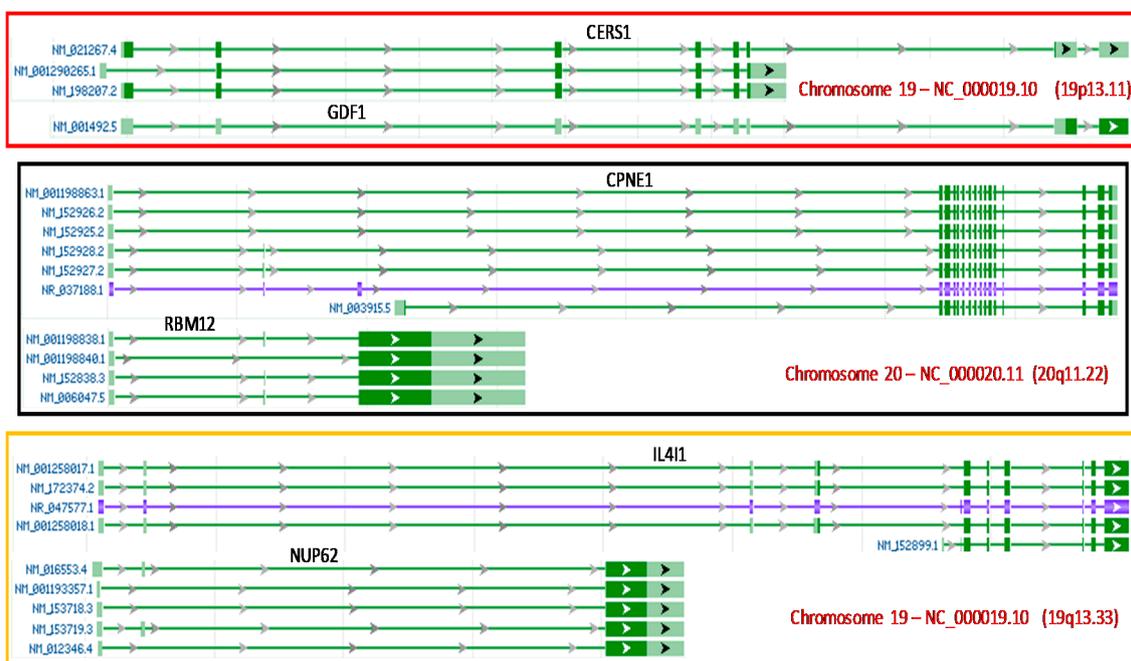
Figure 3. Illustrations copied and modified from the NCBI database showing transcripts from both strands of the DNA double helix. **Top panel**: the protein-coding *BOLA2* gene and the noncoding *SHG1P6* gene on the plus DNA strand together produce six *BOLA2-SMG1P6* messenger RNAs (mRNAs) and one noncoding *BOLA2-SMG1P6* RNA, while the protein-coding *SLX1B* and *SULT1A4* genes on the minus strand together produce a *SLX1B-SULT1A4* noncoding RNA. All genes or RNAs mentioned are highlighted with red circles. **Middle panel**: The protein coding *FKBP1A* and *SDCBP2* genes on the plus strand of the human 20p13 region together produce a *FKBP1A-SDCBP2* noncoding RNA, while the minus DNA strand of this region is also transcribed to three antisense (AS) RNAs that overlap, in a reverse-complementary manner, with an end of the *FKBP1A* and *SDCBP2* mRNAs. The overlaps can easily lead to creation of an artificial *FKBP1A-SDCBP2* cDNA during reverse transcription (RT) or PCR, as we described before [5,6,64,105]. **Bottom panel**: The NCBI database uses NM, XM, NR, XR, NP, and XP to indicate normalized mRNA, predicated mRNA, noncoding RNA, predicated noncoding RNA, normalized protein, and predicated protein, respectively, while it uses green and blue colors to indicate mRNA and noncoding RNA, respectively. The NCBI also uses boxes and lines to indicate exons and introns, respectively, with their lengths in proportion to the lengths of the exons or introns in the number of nucleotides (RNA) or base-pairs (DNA).

5. Some Genes Are Encoded by the Same Genomic Locus with Their RNAs Sharing Exons

As aforementioned [5,6,36–40], "what is a gene" has become an unanswered question in the post ENCODE era. In our opinion, a long mature RNA should be regarded as a gene, regardless of whether it is protein-coding or noncoding and whether it is produced from a linear DNA or is produced solely at the RNA level without a corresponding genomic base [6,38]. Short noncoding RNAs should not be considered as genes because each of them, such as a microRNA, often is not unique and has many repeats in the 3.2–3.5 billion-base-pair sequence of the human genome [7,8,11].

A protein, after it has been translated from an mRNA but before it is posttranslationally modified to different protein forms, should also be regarded as a gene, partly because in some special situations, one single mRNA sequence may be annotated as different genes in the NCBI, which is a special case of the "a gene contains gene(s)" situation or a special case of the crowdedness of some genomic loci. This

situation can be reflected by the so-called "alternative reading frame (ARF)" of mRNAs, as seen in the mRNAs that are encoded by a single genomic locus called *INK4* and are translated to the p15, p16, and p19 tumor suppressor proteins in human and rodent cells [139,140]. As a better example, the *GDF1* mRNA (NM_001492.5) is identical to the longest mRNA (NM_021267.4) of the *CES1* gene, although it encodes different open reading frames (ORFs) when it is the *GDF1* mRNA than when it is one of the *CES1* mRNAs. This is because both *GDF1* and *CES1* genes reside at the same genomic locus (19p13.11) and are transcribed from the same initiation site, as illustrated in Figure 4. If we do not regard different proteins as different genes, the same mRNA-encoded *GDF1* and *CES1* can only be considered as the same gene. There are other similar cases in which two genes not only reside at the same genomic locus but are also transcribed from the same initiation site, with the RNAs of the two different genes sharing some exons. For example, the *RBM12* gene is within the *CPNE1* gene in the human 20q11.22 region, with the two genes sharing the same transcription initiation site and with two of the *RMB12*'s three exons also appearing in some *CPNE1* RNA variants (Figure 4). The relationships between the *IL4I1* and *NUP62* genes, and between their RNAs, are the same as those between the *CPNE1* and *RBM12* genes, and between their RNAs (Figure 4).

Figure 4. Images copied and modified from the NCBI database illustrating that one human genomic locus harbors two genes whose RNAs not only are transcribed from the same initiation site but also share exons. **Top panel**: The *CERS1* and *GDF1* genes are encoded by the same human genomic locus at the 19p13.11 region, and the *GDF1* mRNA is identical to the largest *CERS1* mRNA, but the same mRNA codes for different open reading frames (ORFs) for the *GDF1* and the *CERS1* genes. **Middle panel**: The *CPNE1* and *RBM12* genes are encoded by the same genomic locus at the human 20q11.22 region and are transcribed from the same initiation site. While the *CPNE1* transcripts may be *cis*-spliced to six mRNAs and one noncoding RNA, the *RBM12* transcripts may be *cis*-spliced to four mRNAs. The *CPNE1* RNAs share some exons with the *RBM12* RNAs. **Bottom panel**: The three mRNAs and one noncoding RNA of the *IL4I1* gene share some exons with the five mRNAs of the *NUP62* gene, and both genes locate at the same genomic locus in the human 19q13.33 region, with some RNAs of these two genes sharing the same transcription initiation site.

Many long mature RNAs that encompass sequences of two neighboring genes can be protein-coding or noncoding, regardless of whether their 5' or 3' partner gene encodes mRNA(s) or noncoding RNA(s). For instance, the *CNPY3* gene and its downstream gene, *GNMT*, encode

both mRNAs and noncoding RNAs, and several *CNPY3-GNMT* RNAs are also protein-coding and noncoding (Figure 5), although we do not know whether the two-gene RNAs are derived from a readthrough, a *trans*-splicing, an unknown mechanism, or even a combination of different mechanisms. To many RNA experts, it may not be necessary to point out that a given cell or tissue type in a given situation may not express all the RNA variants, such as all the *CNPY3*, *GNMT* or *CNPY3-GNMT* variants. However, it is worth noting that, currently, there is no pellucid definition for noncoding RNA. Many researchers arbitrarily consider those RNAs whose largest ORF is smaller than 100 codons, i.e., 300 nucleotides, as noncoding [141–145], and further arbitrarily regard those RNAs with 200 or more nucleotides as long noncoding ones while those smaller than 200 nucleotides (which may encode more than 60 amino acids) as short ones, while some others only consider those RNAs encoding less than 30 amino acids as noncoding [42,43]. Obviously, this definition of "noncoding" ignores ample evidence proving that peptides much smaller than 99 amino acids may have biological functions [145–155], as has been described by us [38]. Since peptides as short as 11 amino acids still have important biological functions [147–150], even some short noncoding RNAs may have effects by producing small proteins. Therefore, it is comprehensible that some RNAs are classified as noncoding in the NCBI database but as protein-coding in the Ensembl database. For instance, the *STX16-NPEPL1* RNA (Gene ID: 100534593; 20q13.32) is predicated to be noncoding in the NCBI database (NR_037945.1) but to be coding in the Ensembl database (ENSG00000254995).
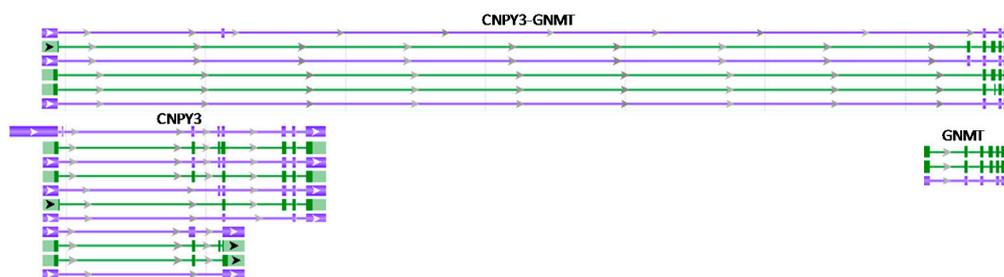

Figure 5. An illustration copied and modified from the NCBI database showing multiple mRNAs and noncoding RNAs of the *CNPY3*, *GNMT,* and *CNPY3-GNMT* genes in the human 6p21.1 region.

6. Two-Gene RNAs from Unknown Mechanism Make RNA Classification Difficult

Traditionally, RNAs are classified into the three categories of messenger RNA, transfer RNA, and ribosomal RNA. However, long mature RNAs can actually be categorized in different ways, such as using the RNA polymerase that synthesizes the RNA [38], but each classification method has its strengths and weaknesses. For example, based on whether or not an RNA has a corresponding parental gene in the nuclear or mitochondrial genome, RNAs can be dichotomized into two groups, i.e., (1) those that have a corresponding parental gene, i.e., have a genomic DNA parent; and (2) those that are produced at the RNA level without a genomic parent [6]. The former group includes not only all those RNAs that are clearly known to be derived from a readthrough mechanism as a subgroup, but also all RNAs that are transcribed from fusion genes formed due to genetic alterations [15–18], mostly discerned in genetic diseases and tumors [22–24], as another subgroup. It needs to be pointed out that for most two-neighboring-gene RNAs, examples being listed in Table 1, their derivation is unknown, in part because *trans*-splicing as a possible mechanism has not yet been ruled out and, therefore, cannot currently be sorted into the "readthrough" subgroup. The latter group lacks a genomic parent and is complex because it covers a variety of noncolinear RNAs, including those neighboring-gene RNAs from unknown mechanisms. Therefore, all methods of sorting that we can think of seem to become problematic once dealing with those RNAs containing sequences from two neighboring genes resulting from unknown mechanisms. Actually, it is even more problematic when dealing with mitochondrial RNAs that may form trimeras or even tetrameras, i.e., those RNAs containing sequences of three or

four mitochondrial genes, as we once reported [65], because how these trimeras or tetrameras are yielded remains unknown.

If a two-gene RNA is detected at high abundance in a situation wherein one of the two partner genes is undetectable, either the upstream or the downstream one, it may be a hint that a readthrough mechanism may underlie the production of the two-gene RNA, because the lack of one of the two partner transcripts makes *trans*-splicing impossible. Moreover, some two-gene RNAs contain exons from the intergenic region, such as the human *ZNF664-FAM101A* RNA produced from the 12q24.31 region (Figure 6). The existence of the intergenic-sequence-derived exon(s) makes it unlikely that the RNAs are produced via a *trans*-splicing of two individually transcribed RNA molecules, thus, indirectly supporting that the RNAs are derived from a readthrough mechanism. Nevertheless, uncontested experimental proof showing a readthrough event, including the existence of the not-yet-spliced precursor transcript and the relevant procedure, is still required for the claim that a two-gene RNA is engendered via readthrough. We should not assume that all two-neighboring-gene RNAs are produced by transcriptional readthrough simply because readthrough is common, while arbitrarily ruling out the possible involvement of *trans*-splicing that is also considered by other researchers to be a common event [15,66–71]. A caveat probably needs to be given that convincing experimental proof should require a non-RT and non-PCR approach to avoid technical spuriousness that may be created by these techniques [5,6,64,105–135], by using the cDNA protection assay established by us [105], the less sensitive RNA protection assay [136–138], or other approaches [110,111] as alternatives.
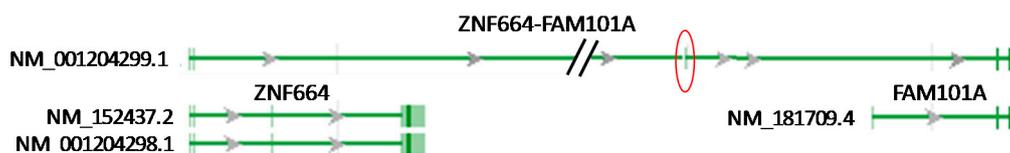
Figure 6. An image copied and modified from the NCBI showing that the *ZNF664-FAM101A* RNA contains one exon (in the red circle) derived from the very-long intergenic region, making this RNA more likely to be produced via a transcriptional-readthrough mechanism but not via a *trans*-splicing of a *ZNF664* transcript and a *FAM101A* transcript, although, theoretically, there may exist an unknown mechanism that can splice three transcripts (i.e., the *ZNF664*, the intergenic, and the *FAM101A* transcripts) into one mature RNA.

7. We Propose to Classify Long Mature RNAs into Four Types

In our opinion, long mature RNAs should be categorized based on the mechanism used to produce the RNA. There are two criteria for the mechanism, i.e., (1) whether or not the RNA has one single gene as the sole genomic parent and (2) whether or not the RNA is derived from *cis*-splicing of a single RNA transcript. By these criteria, all long mature RNAs that have been reported can be classified into four different types (Table 1). Those RNAs transcribed from already-annotated genes, which constitute the vast majority of long mature RNAs, are sorted into type I. It is essential to note that this type also includes those two-neighboring-gene RNAs that are clearly known to be derived from transcriptional readthrough or from fusion genes that are formed pathologically. This is because we regard each genomic locus encoding readthrough RNA as an unannotated, i.e., a newly-identified, gene, which in turn is because these unannotated genes do not show any difference from those already-annotated ones, pertaining to all transcriptional and posttranscriptional regulations. It goes without saying that these newly-identified genes should be annotated and assigned a name and a gene identification number (gene ID). Actually, the NCBI has already assigned a gene ID to each of those RNAs that contain sequences of two adjacent genes and named them simply by using a hyphen to link the names of the two genes, with examples shown in Table 1. We suggest to the RNA research fraternity to follow the NCBI's nomenclature to annotate all those, and only those, RNAs that are clearly known to be derived from a readthrough mechanism. However, most of those two-neighboring-gene RNAs that have been

reported in the literature or listed in the NCBI database have not yet been confirmed to be derived via this mechanism and, thus, should not be grouped into this category at the moment, in our opinion. Therefore, to accommodate those RNAs for which derivation is not yet known, we temporarily put them into type II. Here, "temporarily" means that they should eventually be recategorized into either type I if a readthrough is later confirmed, or into a new type if a *trans*-splicing event is confirmed or a new mechanism is identified. Those noncolinear RNAs that are not two-gene ones, such as the aforementioned *KLK4* RNA variant containing both sense and antisense [75] as well as the *ERα* RNA variants that contain duplicated exons [54–56], are all grouped into type III. It remains possible that a *trans*-splicing or a currently-unknown mechanism may account for the formations of this type of RNA. Those RNAs that contain sequences of two genes on different chromosomes and for which *trans*-splicing has been claimed as a source, such as the *JAZF1-JJAZ1* chimeric RNA that was reported to be derived via a mechanism mimicking *trans*-splicing [78,81], are authentic chimeric RNAs and are classified into type IV.

8. Do *Trans*-Splicing and Authentic Chimeric RNAs Really Exist in Human Cells?

Although we are aware of a handful of RNAs in human cells that have been reported to be chimeric RNAs formed via *trans*-splicing [25,54–56,75,76,78,87–91,137,156], and have grouped them into type IV in Table 2, we still doubt (1) whether *trans*-splicing really exists and, thus, (2) whether *trans*-splicing-derived authentic chimeric RNAs truly exist, in human cells. We have several lines of thought that lead us to these suspicions:

1. The number of *cis*-splicing events and *cis*-splicing derived RNAs in human cells are numerous, and *trans*-splicing is very common in evolutionarily-low organisms [157–160], whereas reported *trans*-splicing events in human cells have so far been very few. Therefore, it seems to us that *trans*-splicing may have undergone regression during evolution towards higher organisms, although we still need to determine whether *trans*-splicing has become defunct in healthy humans and whether it reappears during carcinogenesis, which would be considered an atavism, i.e., a reverse-evolutionary process.

2. Most, if not all, published studies that claim the observation of *trans*-splicing in human cells do not provide us with procedural and mechanistic details of the splicing. Therefore, we still know very little about it, although *cis*-splicing is well-characterized in human cells and *trans*-splicing is well characterized in evolutionarily-lower organisms. For example, although we do know that a large number of proteins are involved in *cis*-splicing, we do not know how many proteins are involved in *trans*-splicing and what these proteins are in human cells. After more than a decade since the initial publications on many chimeric RNAs and other noncolinear RNAs that are believed to be derived from *trans*-splicing, few follow-up studies, either by the initial reporters or by other researchers, have been published on the procedural and mechanistic details of the *trans*-splicing per se and of how the splicing leads to the formation of chimeras or other noncolinear RNAs in human cells.

3. If *trans*-splicing does exist in human cells as a mechanism for chimeric RNA formation, we should see more of those chimeras with sequences of two genes that are on the same chromosome but are farther away from each other, too far away for transcriptional readthrough to occur. However, the fact is that two-distant-gene chimeras, if they exist, are rare, which provides indirect evidence against the true existence of a *trans*-splicing mechanism.

4. Yu et al. once tried to validate many reported noncolinear RNAs and suggested that 50% of them are artifacts produced in vitro [161]. This high rate of spuriousness identified by a single study suggests to us that more stringent vindication is required for authentication of the remaining 50%.

Table 2. Classification of long mature RNAs.

#	Transcript Mechanism	Genetic Base	RNAs
I	Well characterized	With an annotated or unannotated (including reathrough) gene as a base	Classical mRNAs and noncoding RNAs Circular RNAs
		With a fusion gene as a DNA base	Fusion RNAs
II	Unknown	With or without a DNA base?	RNAs with neighboring-genes' sequences
III	Less known	Without a DNA base	RNAs with sense and antisense sequences RNAs with duplicated exons
IV	Unknown	With two genes as bases	Authentic chimeric RNAs

Note: "Transcript mechanism" indicates the regulatory mechanisms for the transcription and posttranscription, including *cis*-splicing. Readthrough RNAs are considered to be derived from unannotated genes and thus grouped into type I.

9. Concluding Remarks

Tens of thousands of so-called chimeric RNAs in human cells have been reported in the literature or deposited in different databases, but many of them may be technical artifacts produced during RT or PCR that is part of the high-throughput RNA sequencing technology [5,6,64,105–135]. Most of these chimeras contain sequences of two adjacent genes on the same chromosome and are generally considered to be derived via transcriptional readthrough, but for many of them this remains a reasonable assumption awaiting uncontentious evidence, in part because *trans*-splicing is still a possible mechanism. We agree on the readthrough assumption but regard those genomic loci that are transcriptionally read through as previously unidentified, or newly identified, genes waiting for annotation and characterization. To reiterate, we do not consider readthrough-derived RNAs as chimeras, because readthrough genomic loci reflect the phenomenon of "a gene contains gene(s)" or "gene(s) within a gene" seen in the human genome, and show no difference from the 20,000 human genes and from all fusion genes formed due to genetic alterations. Recapitulated more categorically, there is no difference among unannotated, already-annotated, and fusion genes appertaining to their transcriptional, posttranscriptional, translational, and posttranslational regulations. Therefore, we find no reason to call readthrough RNAs chimeras. We define authentic chimeric RNAs as those formed at the RNA level without one corresponding gene as the sole genomic parent. *Trans*-splicing is the only possible mechanism known so far to be accountable for the formation of such authentic chimeras and other forms of noncolinear RNAs, and probably for the formation of some two-neighboring-gene RNAs as well. However, we doubt the true existence of *trans*-splicing and, thus, the true existence of authentic chimeric RNAs, in human cells, in part because very few RNAs that might be derived from *trans*-splicing have been reported so far, and, for these RNAs, there is a lack of procedural and mechanistic details of the presumed *trans*-splicing. Although we sort long mature RNAs into four different types to accommodate all reported ones, there probably is only one single type, i.e., type I in Table 2, because those in our type II will eventually be regrouped into type I while those in our types III and IV may not really exist, by our speculation. In our opinion, partly because readthrough-derived RNAs are commonly considered as chimeras in the RNA research province, characterization of their parental genes has largely been forgotten, which in turn impedes our understanding of these newly-identified genes. Therefore, it is imperative to stop considering these RNAs as chimeras and, instead, to characterize, as we have for many other genes, their parental genes at all transcriptional, posttranscriptional, translational, and posttranslational levels, with emphasis on their alternative *cis*-splicing. Moreover, it is imperative to determine whether *trans*-splicing really occurs in human cells. If it does not exist, then those two-neighboring-gene RNAs cannot be derived from it and, thus, are more likely to come from a transcriptional readthrough. On the other hand, if it really exists, those RNAs thought to be derived from *trans*-splicing are likely authentic chimeras and many more authentic ones may be awaiting our discovery.

Author Contributions: Y.H., C.Y., H.H. and D.J.L. outlined the concepts and conclusions. Y.H. and D.J.L. prepared the manuscript. L.C. and M.L. prepared figures and figure legends. L.Z. edited and revised the manuscript. L.C., M.L. and L.Z. also participated in discussions of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Birney, E.; Stamatoyannopoulos, J.A.; Dutta, A.; Guigo, R.; Gingeras, T.R.; Margulies, E.H.; Weng, Z.; Snyder, M.; Dermitzakis, E.T.; Thurman, R.E.; et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **2007**, *447*, 799–816. [CrossRef] [PubMed]

2. Gingeras, T.R. Implications of chimaeric non-co-linear transcripts. *Nature* **2009**, *461*, 206–211. [CrossRef] [PubMed]

3. Chwalenia, K.; Facemire, L.; Li, H. Chimeric RNAs in cancer and normal physiology. *Wiley Interdiscip. Rev. RNA* **2017**, *8*. [CrossRef] [PubMed]

4. Sokol, M.; Jessen, KM.; Pedersen, F.S. Utility of next-generation RNA-sequencing in identifying chimeric transcription involving human endogenous retroviruses. *APMIS* **2016**, *124*, 127–139. [CrossRef] [PubMed]

5. Peng, Z.; Yuan, C.; Zellmer, L.; Liu, S.; Xu, N.; Liao, D.J. Hypothesis: Artifacts, Including Spurious Chimeric RNAs with a Short Homologous Sequence, Caused by Consecutive Reverse Transcriptions and Endogenous Random Primers. *J. Cancer* **2015**, *6*, 555–567. [CrossRef] [PubMed]

6. Yuan, C.; Han, Y.; Zellmer, L.; Yang, W.; Guan, Z.; Yu, W.; Huang, H.; Liao, D.J. It Is Imperative to Establish a Pellucid Definition of Chimeric RNA and to Clear Up a Lot of Confusion in the Relevant Research. *Int. J. Mol. Sci.* **2017**, *18*, 714. [CrossRef] [PubMed]

7. Bamshad, M.J.; Ng, S.B.; Bigham, A.W.; Tabor, H.K.; Emond, M.J.; Nickerson, D.A.; Shendure, J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **2011**, *12*, 745–755. [CrossRef] [PubMed]

8. Belizario, J.E. The humankind genome: From genetic diversity to the origin of human diseases. *Genome* **2013**, *56*, 705–716. [CrossRef] [PubMed]

9. Clark, M.B.; Amaral, P.P.; Schlesinger, F.J.; Dinger, M.E.; Taft, R.J.; Rinn, J.L.; Ponting, C.P.; Stadler, P.F.; Morris, K.V.; Morillon, A.; et al. The reality of pervasive transcription. *PLoS Biol.* **2011**, *9*, e1000625. [CrossRef] [PubMed]

10. Liu, X.; Wang, Y.; Yang, W.; Guan, Z.; Yu, W.; Liao, D.J. Protein multiplicity can lead to misconduct in western blotting and misinterpretation of immunohistochemical staining results, creating much conflicting data. *Prog. Histochem. Cytochem.* **2016**, *51*, 51–58. [CrossRef] [PubMed]

11. Lou, X.; Zhang, J.; Liu, S.; Xu, N.; Liao, D.J. The other side of the coin: The tumor-suppressive aspect of oncogenes and the oncogenic aspect of tumor-suppressive genes, such as those along the CCND-CDK4/6-RB axis. *Cell Cycle* **2014**, *13*, 1677–1693. [CrossRef] [PubMed]

12. Pennisi, E. ENCODE project writes eulogy for junk DNA. *Science* **2012**, *337*, 1159–1161. [CrossRef] [PubMed]

13. Skipper, M.; Dhand, R.; Campbell, P. Presenting ENCODE. *Nature* **2012**, *489*, 45. [CrossRef] [PubMed]

14. Zhang, J.; Lou, X.; Shen, H.; Zellmer, L.; Sun, Y.; Liu, S.; Xu, N.; Liao, D.J. Isoforms of wild type proteins often appear as low molecular weight bands on SDS-PAGE. *Biotechnol. J.* **2014**, *9*, 1044–1054. [CrossRef] [PubMed]

15. Luo, J.H.; Liu, S.; Zuo, Z.H.; Chen, R.; Tseng, G.C.; Yu, Y.P. Discovery and Classification of Fusion Transcripts in Prostate Cancer and Normal Prostate Tissue. *Am. J. Pathol.* **2015**, *185*, 1834–1845. [CrossRef] [PubMed]

16. Davare, M.A.; Tognon, C.E. Detecting and targetting oncogenic fusion proteins in the genomic era. *Biol. Cell* **2015**, *107*, 111–129. [CrossRef] [PubMed]

17. Mertens, F.; Tayebwa, J. Evolving techniques for gene fusion detection in soft tissue tumours. *Histopathology* **2014**, *64*, 151–162. [CrossRef] [PubMed]

18. Mertens, F.; Johansson, B.; Fioretos, T.; Mitelman, F. The emerging complexity of gene fusions in cancer. *Nat. Rev. Cancer* **2015**, *15*, 371–381. [CrossRef] [PubMed]

19. Kinali, M.; Arechavala-Gomeza, V.; Cirak, S.; Glover, A.; Guglieri, M.; Feng, L.; Hollingsworth, K.G.; Hunt, D.; Jungbluth, H.; Roper, H.P.; et al. Muscle histology vs. MRI in Duchenne muscular dystrophy. *Neurology* **2011**, *76*, 346–353. [CrossRef] [PubMed]

20. Aartsma-Rus, A.; Van Deutekom, J.C.; Fokkema, I.F.; Van Ommen, G.J.; Den Dunnen, J.T. Entries in the Leiden Duchenne muscular dystrophy mutation database: An overview of mutation types and paradoxical cases that confirm the reading-frame rule. *Muscle Nerve* **2006**, *34*, 135–144. [CrossRef] [PubMed]

21. Shlien, A.; Raine, K.; Fuligni, F.; Arnold, R.; Nik-Zainal, S.; Dronov, S.; Mamanova, L.; Rosic, A.; Ju, Y.S.; Cooke, S.L.; et al. Direct Transcriptional Consequences of Somatic Mutation in Breast Cancer. *Cell Rep.* **2016**, *16*, 2032–2046. [CrossRef] [PubMed]

22. Jia, Y.; Chen, L.; Jia, Q.; Dou, X.; Xu, N.; Liao, D.J. The well-accepted notion that gene amplification contributes to increased expression still remains, after all these years, a reasonable but unproven assumption. *J. Carcinog.* **2016**, *15*, 3. [CrossRef] [PubMed]

23. Wang, G.; Chen, L.; Yu, B.; Zellmer, L.; Xu, N.; Liao, D.J. Learning about the Importance of Mutation Prevention from Curable Cancers and Benign Tumors. *J. Cancer* **2016**, *7*, 436–445. [CrossRef] [PubMed]

24. Zhang, J.; Lou, X.; Zellmer, L.; Liu, S.; Xu, N.; Liao, D.J. Just like the rest of evolution in Mother Nature, the evolution of cancers may be driven by natural selection, and not by haphazard mutations. *Oncoscience* **2014**, *1*, 580–590. [CrossRef] [PubMed]

25. Guerra, E.; Trerotola, M.; Dell'Arciprete, R.; Bonasera, V.; Palombo, B.; El-Sewedy, T.; Ciccimarra, T.; Crescenzi, C.; Lorenzini, F.; Rossi, C.; et al. A bicistronic CYCLIN D1-TROP2 mRNA chimera demonstrates a novel oncogenic mechanism in human cancer. *Cancer Res.* **2008**, *68*, 8113–8121. [CrossRef] [PubMed]

26. Hungerford, D.A. The philadelphia chromosome and some others. *Ann. Intern. Med.* **1964**, *61*, 789–793. [CrossRef] [PubMed]

27. Koretzky, G.A. The legacy of the Philadelphia chromosome. *J. Clin. Investig.* **2007**, *117*, 2030–2032. [CrossRef] [PubMed]

28. Nowell, P.; Hungerford, D. A minute chromosome in human chronic granulocytic leukemia. *Science* **1960**, *132*, 1497.

29. Nowell, P.C.; Hungerford, D.A. Chromosome studies on normal and leukemic human leukocytes. *J. Natl. Cancer Inst.* **1960**, *25*, 85–109. [PubMed]

30. Nowell, P.C. The minute chromosome (Phl) in chronic granulocytic leukemia. *Blut* **1962**, *8*, 65–66. [CrossRef] [PubMed]

31. Vilborg, A.; Steitz, J.A. Readthrough transcription: How are DoGs made and what do they do? *RNA Biol.* **2017**, *14*, 632–636. [CrossRef] [PubMed]

32. Henkin, T.M. The T box riboswitch: A novel regulatory RNA that utilizes tRNA as its ligand. *Biochim. Biophys. Acta* **2014**, *1839*, 959–963. [CrossRef] [PubMed]

33. Vilborg, A.; Sabath, N.; Wiesel, Y.; Nathans, J.; Levy-Adam, F.; Yario, T.A.; Steitz, J.A.; Shalgi, R. Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E8362–E8371. [CrossRef] [PubMed]

34. Yang, M.; Sun, Y.; Ma, L.; Wang, C.; Wu, J.M.; Bi, A.; Liao, D.J. Complex alternative splicing of the Smarca2 gene suggests the importance of Smarca2-B variants. *J. Cancer* **2011**, *2*, 386–400. [CrossRef] [PubMed]

35. Yang, M.; Wu, J.; Wu, S.H.; Bi, A.D.; Liao, D.J. Splicing of mouse p53 pre-mRNA does not always follow the "first come, first served" principle and may be influenced by cisplatin treatment and serum starvation. *Mol. Biol. Rep.* **2012**, *39*, 9247–9256. [CrossRef] [PubMed]

36. Finta, C.; Warner, S.C.; Zaphiropoulos, P.G. Intergenic mRNAs. Minor gene products or tools of diversity? *Histol. Histopathol.* **2002**, *17*, 677–682. [PubMed]

37. Gerstein, M.B.; Bruce, C.; Rozowsky, J.S.; Zheng, D.; Du, J.; Korbel, J.O.; Emanuelsson, O.; Zhang, Z.D.; Weissman, S.; Snyder, M. What is a gene, post-ENCODE? History and updated definition. *Genome Res.* **2007**, *17*, 669–681. [CrossRef] [PubMed]

38. Jia, Y.; Chen, L.; Ma, Y.; Zhang, J.; Xu, N.; Liao, D.J. To Know How a Gene Works, We Need to Redefine It First but then, More Importantly, to Let the Cell Itself Decide How to Transcribe and Process Its RNAs. *Int. J. Biol. Sci.* **2015**, *11*, 1413–1423. [CrossRef] [PubMed]

39. Ponting, C.P.; Belgard, T.G. Transcribed dark matter: Meaning or myth? *Hum. Mol. Genet.* **2010**, *19*, R162–R168. [CrossRef] [PubMed]

40. Portin, P. The elusive concept of the gene. *Hereditas* **2009**, *146*, 112–117. [CrossRef] [PubMed]

41. Signal, B.; Gloss, B.S.; Dinger, M.E. Computational Approaches for Functional Prediction and Characterisation of Long Noncoding RNAs. *Trends Genet.* **2016**, *32*, 620–637. [CrossRef] [PubMed]

42. Jia, H.; Osak, M.; Bogu, G.K.; Stanton, L.W.; Johnson, R.; Lipovich, L. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA* **2010**, *16*, 1478–1487. [CrossRef] [PubMed]

43. Castelo-Branco, G.; Bonetti, A. Birth, coming of age and death: The intriguing life of long noncoding RNAs. *Semin. Cell Dev. Biol.* **2017**. [CrossRef]

44. Nilsen, T.W.; Graveley, B.R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **2010**, *463*, 457–463. [CrossRef] [PubMed]

45. Glanz, S.; Kuck, U. Trans-splicing of organelle introns—A detour to continuous RNAs. *Bioessays* **2009**, *31*, 921–934. [CrossRef] [PubMed]

46. Jacobs, J.; Glanz, S.; Bunse-Grassmann, A.; Kruse, O.; Kuck, U. RNA *trans*-splicing: Identification of components of a putative chloroplast spliceosome. *Eur. J. Cell Biol.* **2010**, *89*, 932–939. [CrossRef] [PubMed]

47. Lasda, E.L.; Blumenthal, T. *Trans*-splicing. *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 417–434. [CrossRef] [PubMed]

48. Borst, P. Maxi-circles, glycosomes, gene transposition, expression sites, transsplicing, transferrin receptors and base. *J. Mol. Biochem. Parasitol.* **2016**, *205*, 39–52. [CrossRef] [PubMed]

49. Berger, A.; Maire, S.; Gaillard, M.C.; Sahel, J.A.; Hantraye, P.; Bemelmans, A.P. mRNA *trans*-splicing in gene therapy for genetic diseases. *Wiley Interdiscip. Rev. RNA* **2016**, *7*, 487–498. [CrossRef] [PubMed]

50. Maniatis, T.; Tasic, B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **2002**, *418*, 236–243. [CrossRef] [PubMed]

51. Dixon, R.J.; Eperon, I.C.; Samani, N.J. Complementary intron sequence motifs associated with human exon repetition: A role for intragenic, inter-transcript interactions in gene expression. *Bioinformatics* **2007**, *23*, 150–155. [CrossRef] [PubMed]

52. Rigatti, R.; Jia, J.H.; Samani, N.J.; Eperon, I.C. Exon repetition: A major pathway for processing mRNA of some genes is allele-specific. *Nucleic Acids Res.* **2004**, *32*, 441–446. [CrossRef] [PubMed]

53. Shao, X.; Shepelev, V.; Fedorov, A. Bioinformatic analysis of exon repetition, exon scrambling and *trans*-splicing in humans. *Bioinformatics* **2006**, *22*, 692–698. [CrossRef] [PubMed]

54. Flouriot, G.; Brand, H.; Seraphin, B.; Gannon, F. Natural *trans*-spliced mRNAs are generated from the human estrogen receptor-α (hER α) gene. *J. Biol. Chem.* **2002**, *277*, 26244–26251. [CrossRef] [PubMed]

55. Pink, J.J.; Wu, S.Q.; Wolf, D.M.; Bilimoria, M.M.; Jordan, V.C. A novel 80 kDa human estrogen receptor containing a duplication of exons 6 and 7. *Nucleic Acids Res.* **1996**, *24*, 962–969. [CrossRef] [PubMed]

56. Pink, J.J.; Fritsch, M.; Bilimoria, M.M.; Assikis, V.J.; Jordan, V.C. Cloning and characterization of a 77-kDa oestrogen receptor isolated from a human breast cancer cell line. *Br. J. Cancer* **1997**, *75*, 17–27. [CrossRef] [PubMed]

57. Caudevilla, C.; Serra, D.; Miliar, A.; Codony, C.; Asins, G.; Bach, M.; Hegardt, F.G. Natural *trans*-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12185–12190. [CrossRef] [PubMed]

58. Caudevilla, C.; Serra, D.; Miliar, A.; Codony, C.; Asins, G.; Bach, M.; Hegardt, F.G. Processing of carnitine octanoyltransferase pre-mRNAs by *cis* and *trans*-splicing. *Adv. Exp. Med. Biol.* **1999**, *466*, 95–102. [PubMed]

59. Caudevilla, C.; Codony, C.; Serra, D.; Plasencia, G.; Roman, R.; Graessmann, A.; Asins, G.; Bach-Elias, M.; Hegardt, F.G. Localization of an exonic splicing enhancer responsible for mammalian natural *trans*-splicing. *Nucleic Acids Res.* **2001**, *29*, 3108–3115. [CrossRef] [PubMed]

60. Akopian, A.N.; Okuse, K.; Souslova, V.; England, S.; Ogata, N.; Wood, J.N. *Trans*-splicing of a voltage-gated sodium channel is regulated by nerve growth factor. *FEBS Lett.* **1999**, *445*, 177–182. [CrossRef]

61. Frantz, S.A.; Thiara, A.S.; Lodwick, D.; Ng, L.L.; Eperon, I.C.; Samani, N.J. Exon repetition in mRNA. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5400–5405. [CrossRef] [PubMed]

62. Yu, S.; Waldholm, J.; Bohm, S.; Visa, N. Brahma regulates a specific *trans*-splicing event at the *mod(mdg4)* locus of *Drosophila melanogaster*. *RNA Biol.* **2014**, *11*, 134–145. [CrossRef] [PubMed]

63. Labrador, M.; Mongelard, F.; Plata-Rengifo, P.; Baxter, E.M.; Corces, V.G.; Gerasimova, T.I. Protein encoding by both DNA strands. *Nature* **2001**, *409*, 1000. [CrossRef] [PubMed]

64. Xie, B.; Yang, W.; Ouyang, Y.; Chen, L.; Jiang, H.; Liao, Y.; Liao, D.J. Two RNAs or DNAs May Artificially Fuse Together at a Short Homologous Sequence (SHS) during Reverse Transcription or Polymerase Chain Reactions, and Thus Reporting an SHS-Containing Chimeric RNA Requires Extra Caution. *PLoS ONE* **2016**, *11*, e0154855. [CrossRef] [PubMed]

65. Yang, W.; Wu, J.M.; Bi, A.D.; Ou-Yang, Y.C.; Shen, H.H.; Chirn, G.W.; Zhou, J.H.; Weiss, E.; Holman, E.P.; Liao, D.J. 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*, e77016. [CrossRef] [PubMed]

66. Burgess, D.J. Gene expression: Controls and roles for trans-splicing. *Nat. Rev. Genet.* **2013**, *14*, 822. [CrossRef] [PubMed]

67. Wu, C.S.; Yu, C.Y.; Chuang, C.Y.; Hsiao, M.; Kao, C.F.; Kuo, H.C.; Chuang, T.J. Integrative transcriptome sequencing identifies *trans*-splicing events with important roles in human embryonic stem cell pluripotency. *Genome Res.* **2014**, *24*, 25–36. [CrossRef] [PubMed]

68. Kowarz, E.; Merkens, J.; Karas, M.; Dingermann, T.; Marschalek, R. Premature transcript termination, trans-splicing and DNA repair: A vicious path to cancer. *Am. J. Blood Res.* **2011**, *1*, 1–12. [PubMed]

69. Lei, Q.; Li, C.; Zuo, Z.; Huang, C.; Cheng, H.; Zhou, R. Evolutionary Insights into RNA trans-Splicing in Vertebrates. *Genome Biol. Evol.* **2016**, *8*, 562–577. [CrossRef] [PubMed]

70. Babiceanu, M.; Qin, F.; Xie, Z.; Jia, Y.; Lopez, K.; Janus, N.; Facemire, L.; Kumar, S.; Pang, Y.; Qi, Y.; et al. Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res.* **2016**, *44*, 2859–2872. [CrossRef] [PubMed]

71. Jia, Y.; Xie, Z.; Li, H. Intergenically Spliced Chimeric RNAs in Cancer. *Trends Cancer* **2016**, *2*, 475–484. [CrossRef] [PubMed]

72. Zaphiropoulos, P.G. *Trans*-splicing in Higher Eukaryotes: Implications for Cancer Development? *Front. Genet.* **2011**, *2*, 92. [CrossRef] [PubMed]

73. Horiuchi, T.; Aigaki, T. Alternative *trans*-splicing: A novel mode of pre-mRNA processing. *Biol. Cell* **2006**, *98*, 135–140. [CrossRef] [PubMed]

74. Zhang, L.; Lu, H.; Xin, D.; Cheng, H.; Zhou, R. A novel ncRNA gene from mouse chromosome 5 trans-splices with *Dmrt1* on chromosome 19. *Biochem. Biophys. Res. Commun.* **2010**, *400*, 696–700. [CrossRef] [PubMed]

75. Lai, J.; Lehman, M.L.; Dinger, M.E.; Hendy, S.C.; Mercer, T.R.; Seim, I.; Lawrence, M.G.; Mattick, J.S.; Clements, J.A.; Nelson, C.C. A variant of the *KLK4* gene is expressed as a *cis* sense-antisense chimeric transcript in prostate cancer cells. *RNA* **2010**, *16*, 1156–1166. [CrossRef] [PubMed]

76. Takahara, T.; Kanazu, S.I.; Yanagisawa, S.; Akanuma, H. Heterogeneous Sp1 mRNAs in human HepG2 cells include a product of homotypic *trans*-splicing. *J. Biol. Chem.* **2000**, *275*, 38067–38072. [CrossRef] [PubMed]

77. Takahara, T.; Tasic, B.; Maniatis, T.; Akanuma, H.; Yanagisawa, S. Delay in synthesis of the 3′ splice site promotes *trans*-splicing of the preceding 5′ splice site. *Mol. Cell* **2005**, *18*, 245–251. [CrossRef] [PubMed]

78. Li, H.; Wang, J.; Mor, G.; Sklar, J. A neoplastic gene fusion mimics *trans*-splicing of RNAs in normal human cells. *Science* **2008**, *321*, 1357–1361. [CrossRef] [PubMed]

79. Rowley, J.D.; Blumenthal, T. The cart before the horse. *Science* **2008**, *321*, 1302–1304. [CrossRef] [PubMed]

80. Li, H.; Wang, J.; Mor, G.; Sklar, J. Erratum for the Report "A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells". *Science* **2015**, *350*, aad3463.

81. Li, H.; Wang, J.; Ma, X.; Sklar, J. Gene fusions and RNA trans-splicing in normal and neoplastic human cells. *Cell Cycle* **2009**, *8*, 218–222. [CrossRef] [PubMed]

82. Conklin, C.M.; Longacre, T.A. Endometrial stromal tumors: The new WHO classification. *Adv. Anat. Pathol.* **2014**, *21*, 383–393. [CrossRef] [PubMed]

83. Choi, Y.J.; Jung, S.H.; Kim, M.S.; Baek, I.P.; Rhee, J.K.; Lee, S.H.; Hur, S.Y.; Kim, T.M.; Chung, Y.J.; Lee, S.H. Genomic landscape of endometrial stromal sarcoma of uterus. *Oncotarget* **2015**, *6*, 33319–33328. [CrossRef] [PubMed]

84. Amador-Ortiz, C.; Roma, A.A.; Huettner, P.C.; Becker, N.; Pfeifer, J.D. *JAZF1* and *JJAZ1* gene fusion in primary extrauterine endometrial stromal sarcoma. *Hum. Pathol.* **2011**, *42*, 939–946. [CrossRef] [PubMed]

85. Oliva, E.; de Leval, L.; Soslow, R.A.; Herens, C. High frequency of JAZF1-JJAZ1 gene fusion in endometrial stromal tumors with smooth muscle differentiation by interphase FISH detection. *Am. J. Surg. Pathol.* **2007**, *31*, 1277–1284. [CrossRef] [PubMed]

86.　Terrinoni, A.; Dell'Arciprete, R.; Fornaro, M.; Stella, M.; Alberti, S. Cyclin D1 gene contains a cryptic promoter that is functional in human cancer cells. *Genes Chromosomes Cancer* **2001**, *31*, 209–220. [CrossRef] [PubMed]

87.　Ye, Q.; Chung, L.W.; Li, S.; Zhau, H.E. Identification of a novel FAS/ER-α fusion transcript expressed in human cancer cells. *Biochim. Biophys. Acta* **2000**, *1493*, 373–377. [CrossRef]

88.　Finta, C.; Zaphiropoulos, P.G. Intergenic mRNA molecules resulting from *trans*-splicing. *J. Biol. Chem.* **2002**, *277*, 5882–5890. [CrossRef] [PubMed]

89.　Jehan, Z.; Vallinayagam, S.; Tiwari, S.; Pradhan, S.; Singh, L.; Suresh, A.; Reddy, H.M.; Ahuja, Y.R.; Jesudasan, R.A. Novel noncoding RNA from human Y distal heterochromatic block (Yq12) generates testis-specific chimeric *CDC2L2*. *Genome Res.* **2007**, *17*, 433–440. [CrossRef] [PubMed]

90.　Chen, J.; Zhao, X.N.; Yang, L.; Hu, G.J.; Lu, M.; Xiong, Y.; Yang, X.Y.; Chang, C.C.; Song, B.L.; Chang, T.Y.; et al. RNA secondary structures located in the interchromosomal region of human ACAT1 chimeric mRNA are required to produce the 56-kDa isoform. *Cell Res.* **2008**, *18*, 921–936. [CrossRef] [PubMed]

91.　Hu, G.J.; Chen, J.; Zhao, X.N.; Xu, J.J.; Guo, D.Q.; Lu, M.; Zhu, M.; Xiong, Y.; Li, Q.; Chang, C.C.; et al. Production of ACAT1 56-kDa isoform in human cells via *trans*-splicing involving the ampicillin resistance gene. *Cell Res.* **2013**, *23*, 1007–1024. [CrossRef] [PubMed]

92.　Li, B.L.; Li, X.L.; Duan, Z.J.; Lee, O.; Lin, S.; Ma, Z.M.; Chang, C.C.; Yang, X.Y.; Park, J.P.; Mohandas, T.K.; et al. Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes. *J. Biol. Chem.* **1999**, *274*, 11060–11071. [CrossRef] [PubMed]

93.　Yang, L.; Lee, O.; Chen, J.; Chen, J.; Chang, C.C.; Zhou, P.; Wang, Z.Z.; Ma, H.H.; Sha, H.F.; Feng, J.X.; et al. Human acyl-coenzyme A:cholesterol acyltransferase 1 (*acat1*) sequences located in two different chromosomes (7 and 1) are required to produce a novel ACAT1 isoenzyme with additional sequence at the N terminus. *J. Biol. Chem.* **2004**, *279*, 46253–46262. [CrossRef] [PubMed]

94.　Hirano, M.; Noda, T. Genomic organization of the mouse *Msh4* gene producing bicistronic, chimeric and antisense mRNA. *Gene* **2004**, *342*, 165–177. [CrossRef] [PubMed]

95.　Fang, W.; Wei, Y.; Kang, Y.; Landweber, L.F. Detection of a common chimeric transcript between human chromosomes 7 and 16. *Biol. Direct* **2012**, *7*, 49. [CrossRef] [PubMed]

96.　Alexiou, P.; Maragkakis, M.; Mourelatos, Z.; Vourekas, A. cCLIP-Seq: Retrieval of Chimeric Reads from HITS-CLIP (CLIP-Seq) Libraries. *Methods Mol. Biol.* **2018**, *1680*, 87–100. [PubMed]

97.　Pinson, M.E.; Pogorelcnik, R.; Court, F.; Arnaud, P.; Vaurs-Barriere, C. CLIFinder: Identification of LINE-1 Chimeric Transcripts in RNA-seq data. *Bioinformatics* **2017**. [CrossRef] [PubMed]

98.　Lagstad, S.; Zhao, S.; Hoff, A.M.; Johannessen, B.; Lingjaerde, O.C.; Skotheim, R.I. *chimeraviz*: A tool for visualizing chimeric RNA. *Bioinformatics* **2017**, *33*, 2954–2956. [CrossRef] [PubMed]

99.　Li, Y.; Heavican, T.B.; Vellichirammal, N.N.; Iqbal, J.; Guda, C. ChimeRScope: A novel alignment-free algorithm for fusion transcript prediction using paired-end RNA-Seq data. *Nucleic Acids Res.* **2017**, *45*, e120. [CrossRef] [PubMed]

100.　Paciello, G.; Ficarra, E. FuGePrior: A novel gene fusion prioritization algorithm based on accurate fusion structure analysis in cancer RNA-seq samples. *BMC Bioinform.* **2017**, *18*, 58. [CrossRef] [PubMed]

101.　Gorohovski, A.; Tagore, S.; Palande, V.; Malka, A.; Raviv-Shay, D.; Frenkel-Morgenstern, M. ChiTaRS-3.1-the enhanced chimeric transcripts and RNA-seq database matched with protein-protein interactions. *Nucleic Acids Res.* **2017**, *45*, D790–D795. [CrossRef] [PubMed]

102.　Rodriguez-Martin, B.; Palumbo, E.; Marco-Sola, S.; Griebel, T.; Ribeca, P.; Alonso, G.; Rastrojo, A.; Aguado, B.; Guigo, R.; Djebali, S. ChimPipe: Accurate detection of fusion genes and transcription-induced chimeras from RNA-seq data. *BMC Genom.* **2017**, *18*, 7. [CrossRef] [PubMed]

103.　Okonechnikov, K.; Imai-Matsushima, A.; Paul, L.; Seitz, A.; Meyer, T.F.; Garcia-Alcalde, F. InFusion: Advancing Discovery of Fusion Genes and Chimeric Transcripts from Deep RNA-Sequencing Data. *PLoS ONE* **2016**, *11*, e0167417. [CrossRef] [PubMed]

104.　Kumar, S.; Razzaq, S.K.; Vo, A.D.; Gautam, M.; Li, H. Identifying fusion transcripts using next generation sequencing. *Wiley Interdiscip. Rev. RNA* **2016**, *7*, 811–823. [CrossRef] [PubMed]

105.　Yuan, C.; Liu, Y.; Yang, M.; Liao, D.J. New methods as alternative or corrective measures for the pitfalls and artifacts of reverse transcription and polymerase chain reactions (RT-PCR) in cloning chimeric or antisense-accompanied RNA. *RNA Biol.* **2013**, *10*, 958–967. [CrossRef] [PubMed]

106. Kim, M.J.; Cho, S.I.; Chae, J.H.; Lim, B.C.; Lee, J.S.; Lee, S.J.; Seo, S.H.; Park, H.; Cho, A.; Kim, S.Y.; et al. Pitfalls of Multiple Ligation-Dependent Probe Amplifications in Detecting DMD Exon Deletions or Duplications. *J. Mol. Diagn.* **2016**, *18*, 253–259. [CrossRef] [PubMed]

107. Labaj, P.P.; Kreil, D.P. Sensitivity, specificity, and reproducibility of RNA-Seq differential expression calls. *Biol. Direct* **2016**, *11*, 66. [CrossRef] [PubMed]

108. Bustin, S.; Nolan, T. Talking the talk, but not walking the walk: RT-qPCR as a paradigm for the lack of reproducibility in molecular research. *Eur. J. Clin. Investig.* **2017**. [CrossRef] [PubMed]

109. Bustin, S.A. The reproducibility of biomedical research: Sleepers awake! *Biomol. Detect. Quantif.* **2014**, *2*, 35–42. [CrossRef] [PubMed]

110. Koo, K.M.; Carrascosa, L.G.; Shiddiky, M.J.; Trau, M. Amplification-Free Detection of Gene Fusions in Prostate Cancer Urinary Samples Using mRNA-Gold Affinity Interactions. *Anal. Chem.* **2016**, *88*, 6781–6788. [CrossRef] [PubMed]

111. Gillen, A.E.; Yamamoto, T.M.; Kline, E.; Hesselberth, J.R.; Kabos, P. Improvements to the HITS-CLIP protocol eliminate widespread mispriming artifacts. *BMC Genom.* **2016**, *17*, 338. [CrossRef] [PubMed]

112. Lecanda, A.; Nilges, B.S.; Sharma, P.; Nedialkova, D.D.; Schwarz, J.; Vaquerizas, J.M.; Leidel, S.A. Dual randomization of oligonucleotides to reduce the bias in ribosome-profiling libraries. *Methods* **2016**, *107*, 89–97. [CrossRef] [PubMed]

113. Waugh, C.; Cromer, D.; Grimm, A.; Chopra, A.; Mallal, S.; Davenport, M.; Mak, J. A general method to eliminate laboratory induced recombinants during massive, parallel sequencing of cDNA library. *Virol. J.* **2015**, *12*, 55. [CrossRef] [PubMed]

114. Thompson, J.R.; Marcelino, L.A.; Polz, M.F. Heteroduplexes in mixed-template amplifications: Formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res.* **2002**, *30*, 2083–2088. [CrossRef] [PubMed]

115. Shao, W.; Boltz, V.F.; Spindler, J.E.; Kearney, M.F.; Maldarelli, F.; Mellors, J.W.; Stewart, C.; Volfovsky, N.; Levitsky, A.; Stephens, R.M.; et al. Analysis of 454 sequencing error rate, error sources, and artifact recombination for detection of Low-frequency drug resistance mutations in HIV-1 DNA. *Retrovirology* **2013**, *10*, 18. [CrossRef] [PubMed]

116. Houseley, J.; Tollervey, D. Apparent non-canonical *trans*-splicing is generated by reverse transcriptase in vitro. *PLoS ONE* **2010**, *5*, e12271. [CrossRef] [PubMed]

117. Beaumeunier, S.; Audoux, J.; Boureux, A.; Ruffle, F.; Commes, T.; Philippe, N.; Alves, R. On the evaluation of the fidelity of supervised classifiers in the prediction of chimeric RNAs. *BioData Min.* **2016**, *9*, 34. [CrossRef] [PubMed]

118. Brakenhoff, R.H.; Schoenmakers, J.G.; Lubsen, N.H. Chimeric cDNA clones: A novel PCR artifact. *Nucleic Acids Res.* **1991**, *19*, 1949. [CrossRef] [PubMed]

119. Cocquet, J.; Chong, A.; Zhang, G.; Veitia, R.A. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **2006**, *88*, 127–131. [CrossRef] [PubMed]

120. Haas, B.J.; Gevers, D.; Earl, A.M.; Feldgarden, M.; Ward, D.V.; Giannoukos, G.; Ciulla, D.; Tabbaa, D.; Highlander, S.K.; Sodergren, E.; et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* **2011**, *21*, 494–504. [CrossRef] [PubMed]

121. Lerat, H.; Berby, F.; Trabaud, M.A.; Vidalin, O.; Major, M.; Trepo, C.; Inchauspe, G. Specific detection of hepatitis C virus minus strand RNA in hematopoietic cells. *J. Clin. Investig.* **1996**, *97*, 845–851. [CrossRef] [PubMed]

122. Mader, R.M.; Schmidt, W.M.; Sedivy, R.; Rizovski, B.; Braun, J.; Kalipciyan, M.; Exner, M.; Steger, G.G.; Mueller, M.W. Reverse transcriptase template switching during reverse transcriptase-polymerase chain reaction: Artificial generation of deletions in ribonucleotide reductase mRNA. *J. Lab. Clin. Med.* **2001**, *137*, 422–428. [CrossRef] [PubMed]

123. McManus, C.J.; Duff, M.O.; Eipper-Mains, J.; Graveley, B.R. Global analysis of *trans*-splicing in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 12975–12979. [CrossRef] [PubMed]

124. Ozsolak, F.; Milos, P.M. RNA sequencing: Advances, challenges and opportunities. *Nat. Rev. Genet.* **2011**, *12*, 87–98. [CrossRef] [PubMed]

125. Paabo, S.; Irwin, D.M.; Wilson, A.C. DNA damage promotes jumping between templates during enzymatic amplification. *J. Biol. Chem.* **1990**, *265*, 4718–4721. [PubMed]

126. Qiu, X.; Wu, L.; Huang, H.; McDonel, P.E.; Palumbo, A.V.; Tiedje, J.M.; Zhou, J. Evaluation of PCR-Generated Chimeras, Mutations, and Heteroduplexes with 16S rRNA Gene-Based Cloning. *Appl. Environ. Microbiol.* **2001**, *67*, 880–887. [CrossRef] [PubMed]

127. Quail, M.A.; Kozarewa, I.; Smith, F.; Scally, A.; Stephens, P.J.; Durbin, R.; Swerdlow, H.; Turner, D.J. A large genome center's improvements to the Illumina sequencing system. *Nat. Methods* **2008**, *5*, 1005–1010. [CrossRef] [PubMed]

128. Ro, S.; Kang, S.H.; Farrelly, A.M.; Ordog, T.; Partain, R.; Fleming, N.; Sanders, K.M.; Kenyon, J.L.; Keef, K.D. Template switching within exons 3 and 4 of KV11.1 (HERG) gives rise to a 5′ truncated cDNA. *Biochem. Biophys. Res. Commun.* **2006**, *345*, 1342–1349. [CrossRef] [PubMed]

129. Roy, S.W.; Irimia, M. When good transcripts go bad: Artifactual RT-PCR 'splicing' and genome analysis. *Bioessays* **2008**, *30*, 601–605. [CrossRef] [PubMed]

130. Shammas, F.V.; Heikkila, R.; Osland, A. Fluorescence-based method for measuring and determining the mechanisms of recombination in quantitative PCR. *Clin. Chim. Acta* **2001**, *304*, 19–28. [CrossRef]

131. Tuiskunen, A.; Leparc-Goffart, I.; Boubis, L.; Monteil, V.; Klingstrom, J.; Tolou, H.J.; Lundkvist, A.; Plumet, S. Self-priming of reverse transcriptase impairs strand-specific detection of dengue virus RNA. *J. Gen. Virol.* **2010**, *91*, 1019–1027. [CrossRef] [PubMed]

132. Zaphiropoulos, P.G. Template switching generated during reverse transcription? *FEBS Lett.* **2002**, *527*, 326. [CrossRef]

133. Gao, R.; Zhao, A.H.; Du, Y.; Ho, W.T.; Fu, X.; Zhao, Z.J. PCR artifacts can explain the reported biallelic JAK2 mutations. *Blood Cancer J.* **2012**, *2*, e56. [CrossRef] [PubMed]

134. Roy, S.W.; Irimia, M. Intron mis-splicing: No alternative? *Genome Biol.* **2008**, *9*, 208. [CrossRef] [PubMed]

135. Zheng, W.; Chung, L.M.; Zhao, H. Bias detection and correction in RNA-Sequencing data. *BMC Bioinform.* **2011**, *12*, 290. [CrossRef] [PubMed]

136. Djebali, S.; Lagarde, J.; Kapranov, P.; Lacroix, V.; Borel, C.; Mudge, J.M.; Howald, C.; Foissac, S.; Ucla, C.; Chrast, J.; et al. Evidence for transcript networks composed of chimeric RNAs in human cells. *PLoS ONE* **2012**, *7*, e28213. [CrossRef] [PubMed]

137. Zhang, C.; Xie, Y.; Martignetti, J.A.; Yeo, T.T.; Massa, S.M.; Longo, F.M. A candidate chimeric mammalian mRNA transcript is derived from distinct chromosomes and is associated with nonconsensus splice junction motifs. *DNA Cell Biol.* **2003**, *22*, 303–315. [CrossRef] [PubMed]

138. Maillet, P.; Delaunay, J.; Baklouti, F. Chimeric probe-mediated ribonuclease protection assay for molecular diagnosis of mRNA deficiencies. *Hum. Mutat.* **1996**, *7*, 61–64. [CrossRef]

139. Sherr, C.J. Divorcing ARF and p53: An unsettled case. *Nat. Rev. Cancer* **2006**, *6*, 663–673. [CrossRef] [PubMed]

140. Tian, X.; Azpurua, J.; Ke, Z.; Augereau, A.; Zhang, Z.D.; Vijg, J.; Gladyshev, V.N.; Gorbunova, V.; Seluanov, A. *INK4* locus of the tumor-resistant rodent, the naked mole rat, expresses a functional p15/p16 hybrid isoform. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 1053–1058. [CrossRef] [PubMed]

141. Bazzini, A.A.; Johnstone, T.G.; Christiano, R.; Mackowiak, S.D.; Obermayer, B.; Fleming, E.S.; Vejnar, C.E.; Lee, M.T.; Rajewsky, N.; Walther, T.C.; et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* **2014**, *33*, 981–993. [CrossRef] [PubMed]

142. Cheng, H.; Chan, W.S.; Li, Z.; Wang, D.; Liu, S.; Zhou, Y. Small open reading frames: Current prediction techniques and future prospect. *Curr. Protein Pept. Sci.* **2011**, *12*, 503–507. [CrossRef] [PubMed]

143. Kageyama, Y.; Kondo, T.; Hashimoto, Y. Coding vs non-coding: Translatability of short ORFs found in putative non-coding transcripts. *Biochimie* **2011**, *93*, 1981–1986. [CrossRef] [PubMed]

144. Landry, C.R.; Zhong, X.; Nielly-Thibault, L.; Roucou, X. Found in translation: Functions and evolution of a recently discovered alternative proteome. *Curr. Opin. Struct. Biol.* **2015**, *32*, 74–80. [CrossRef] [PubMed]

145. Pauli, A.; Valen, E.; Schier, A.F. Identifying (non-)coding RNAs and small peptides: Challenges and opportunities. *Bioessays* **2015**, *37*, 103–112. [CrossRef] [PubMed]

146. Andrews, S.J.; Rothnagel, J.A. Emerging evidence for functional peptides encoded by short open reading frames. *Nat. Rev. Genet.* **2014**, *15*, 193–204. [CrossRef] [PubMed]

147. Chu, Q.; Ma, J.; Saghatelian, A. Identification and characterization of sORF-encoded polypeptides. *Crit. Rev. Biochem. Mol. Biol.* **2015**, *50*, 134–141. [CrossRef] [PubMed]

148. Kondo, T.; Hashimoto, Y.; Kato, K.; Inagaki, S.; Hayashi, S.; Kageyama, Y. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat. Cell Biol.* **2007**, *9*, 660–665. [CrossRef] [PubMed]

149. Kondo, T.; Plaza, S.; Zanet, J.; Benrabah, E.; Valenti, P.; Hashimoto, Y.; Kobayashi, S.; Payre, F.; Kageyama, Y. Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science* **2010**, *329*, 336–339. [CrossRef] [PubMed]

150. Ladoukakis, E.; Pereira, V.; Magny, E.G.; Eyre-Walker, A.; Couso, J.P. Hundreds of putatively functional small open reading frames in *Drosophila*. *Genome Biol.* **2011**, *12*, R118. [CrossRef] [PubMed]

151. Magny, E.G.; Pueyo, J.I.; Pearl, F.M.; Cespedes, M.A.; Niven, J.E.; Bishop, S.A.; Couso, J.P. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **2013**, *341*, 1116–1120. [CrossRef] [PubMed]

152. Anderson, D.M.; Anderson, K.M.; Chang, C.L.; Makarewich, C.A.; Nelson, B.R.; McAnally, J.R.; Kasaragod, P.; Shelton, J.M.; Liou, J.; Bassel-Duby, R.; et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **2015**, *160*, 595–606. [CrossRef] [PubMed]

153. Hashimoto, Y.; Kondo, T.; Kageyama, Y. Lilliputians get into the limelight: Novel class of small peptide genes in morphogenesis. *Dev. Growth Differ.* **2008**, *50*, S269–S276. [CrossRef] [PubMed]

154. Pauli, A.; Norris, M.L.; Valen, E.; Chew, G.L.; Gagnon, J.A.; Zimmerman, S.; Mitchell, A.; Ma, J.; Dubrulle, J.; Reyon, D.; et al. Toddler: An embryonic signal that promotes cell movement via Apelin receptors. *Science* **2014**, *343*, 1248636. [CrossRef] [PubMed]

155. Zanet, J.; Benrabah, E.; Li, T.; Pelissier-Monier, A.; Chanut-Delalande, H.; Ronsin, B.; Bellen, H.J.; Payre, F.; Plaza, S. Pri sORF peptides induce selective proteasome-mediated protein processing. *Science* **2015**, *349*, 1356–1358. [CrossRef] [PubMed]

156. Zhang, J.S.; Longo, F.M. LAR tyrosine phosphatase receptor: Alternative splicing is preferential to the nervous system, coordinated with cell growth and generates novel isoforms containing extensive CAG repeats. *J. Cell Biol.* **1995**, *128*, 415–431. [CrossRef] [PubMed]

157. Allen, M.A.; Hillier, L.W.; Waterston, R.H.; Blumenthal, T. A global analysis of *C. elegans trans*-splicing. *Genome Res.* **2011**, *21*, 255–264. [CrossRef] [PubMed]

158. Denker, J.A.; Zuckerman, D.M.; Maroney, P.A.; Nilsen, T.W. New components of the spliced leader RNP required for nematode *trans*-splicing. *Nature* **2002**, *417*, 667–670. [CrossRef] [PubMed]

159. Hastings, K.E. SL trans-splicing: Easy come or easy go? *Trends Genet.* **2005**, *21*, 240–247. [CrossRef] [PubMed]

160. Nilsen, T.W. Evolutionary origin of SL-addition *trans*-splicing: Still an enigma. *Trends Genet.* **2001**, *17*, 678–680. [CrossRef]

161. Yu, C.Y.; Liu, H.J.; Hung, L.Y.; Kuo, H.C.; Chuang, T.J. Is an observed non-co-linear RNA product spliced in *trans*, in *cis* or just *in vitro*? *Nucleic Acids Res.* **2014**, *42*, 9410–9423. [CrossRef] [PubMed]