Available online freely at www.isisn.org

Animal Science Journal

Print ISSN: 2220-9549 Online ISSN: 2220-9557

Journal by Innovative Scientific Information & Services Network

REVIEW ARTICLE ANIMAL SCIENCE JOURNAL, 2021 12(1): 15-27. **OPEN ACCESS**

Comprehensive review on The tools: The approach to overcome mutations in RNA Splicing

Muhammad Sikandar[1], Umar Farooq[2], Zulfiqar Ahmad[1], Sumaira Shabbir[1], Hassan Usman[3], Shamsa kanwal[4], Reeta Akram[5] and Mahwish Aslam[1]

[1]Department of Zoology, Wildlife and fisheries, University of Agriculture Faisalabad**, Pakistan**
[2]Depatrment of Forestry, Wildlife and Fisheries Govt of Punjab, **Pakistan**
[3]Department of Zoology, Wildlife and fisheries Government College University Faisalabad, **Pakistan**
[4]Department: Department of Botany, The Women University Multan, **Pakistan**
[5]Department (Institute of Molecular Biology and biotechnology IMBB), The University of Lahore **Pakistan**

***Correspondence:** muhammadsikandar2929@yahoo.com **Accepted: 20-10** 2021 **Published online:** 28-10- 2021

Cancer is the main cause when the mutation in the genes occur which may be the lead to several type of mutations in our body. Usually cancer that is common by the mutation of genes is cervical cancer, breast cancer that may affect the body mechanism and directly effects on the targeting cells. By allowing one gene to produce numerous unique protein isoforms, alternative splicing contributes to the majority of protein diversity in higher eukaryotes. It adds an additional layer of control over gene expression. Alternative splicing is used to encode proteins with distinct functions in up to 95 percent of human multi-exon genes. Furthermore, alternative splicing is linked to about 15% of human genetic illnesses and malignancies. Mapping and quantifying alternative splicing events are important for downstream analysis, especially when disease is associated with them. Correct isoform expression from high-throughput RNA-seq data, on the other hand, remains a challenge. Alternative splicing is linked to a group of delicate machines that interact with one another to facilitate vital biological processes such as cell production and differentiation. We hope to demonstrate I) alternative splicing mechanisms and control, and II) alternative splicing-related human disease in this mini-review. III) computational approach used for measuring isoforms with alternative splicing which derived from RNA sequence.

Keywords: RNA, Sequencing, Cancer, Mutation, Splicing

Background on Mutation in sequencing

The disparity between the number of reported protein-coding genes as well as the numbers of observable human polypeptides shows considerable deviations from a "one gene–one polypeptide" notion after the Human Genome Project was completed in 2003. Alternative splicing is now widely acknowledged as a fundamental factor in higher eukaryotes producing protein diversity by allowing a single gene to produce many protein isoforms while also increasing the complexity of gene expression controls [Bodily 2011]. Alternative splicing occurs in up to 95 percent of human multi-exon genes, resulting in proteins that perform a variety of functions in various physiological processes

[Rositch 2013]. Furthermore, alternative splicing has been linked to around 15% of human genetic illnesses and malignancies [Dall 2008:Kreimer 2012]. Alternative splicing research will aid in the development of treatment methods for splicing-related disorders and will help us better understand mRNA complexity and control.

Mechanisms and regulation of alternative splicing

Constitutive splicing produces the same isoforms by splicing mRNA in the same way, whereas alternative splicing produces various isoforms by employing different sets of exons. Alternative splicing is divided into five categories [Arbyn 2017]

1) The most prevalent type of alternative splicing in mammalian cells is exon skipping (also known as cassette exons), which results in the entire loss of one or more exons [Forman 2017:Bruni 2010]. 2) Mutually exclusive exons are a rare sort of splicing event in which two or more splicing events become mutually exclusive. They're all turned on or off at the same moment. [Ferlay 2012]. 3) Alternative 50 splice sites (alternative donors): the use of a different 50 donor site to shift the upstream exon's 30 border. 4) Alternative 30 splice sites (alternative acceptors): In contrast to alternative 50 splice junction sites, each alternative 30 splice connection site causes a change in the subsequent exon's 50 border. 5) The mechanism of one or even more mRNA introns becoming unspliced is known as intron retention (IR). Those intron-retaining mRNA may have a distinct fate [Bitz 2013,Kelly 2018]. The nonsense-mediated decay mechanism degrades some of them, while others may produce new protein isoforms [11]. IR can often result in the production of malfunctioning proteins, which can lead to illnesses. In addition to the five major subtypes, alternative polyadenylation sites and promoters are daily examined. These two, however, also improve the genome's coding potential.

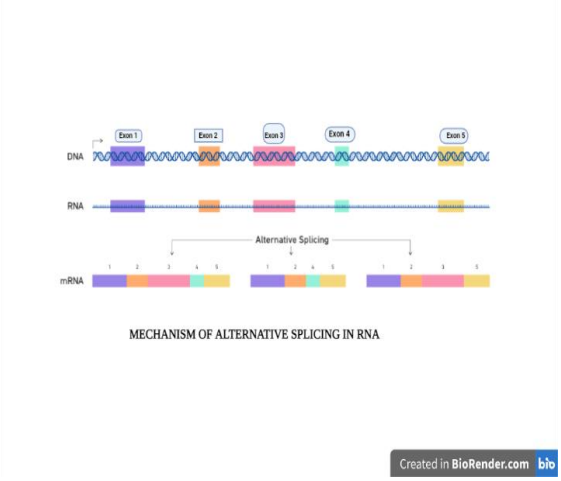

Figure 1: Mechanism of Alternative splicing

General mechanisms of pre-mRNA splicing

Pre-mRNA splicing takes place on the spliceosome, a massive ribonucleoprotein complex (RNP) [Sgadari 2003 Rositch:Toschi 2011]. The spliceosome (Fig. 2) is a dynamic complex made up mainly five nuclear ribonucleoproteins (snRNPs) (U1, U2, U4/U6, U5) which detect and assemble on each intron in transcribed pre-mRNA. U1 connects to the 50 splice site with the aid of a U2 auxiliary factor protein (U2AF) during the assembly process, forming base pairing between both the U1 snRNA and the splice site. Complex A is created when the first synthesized sample complex (complex E or commitment complex) binds with U2 (pre spliceosome). In the final stage, if complex A is synthesized, an intron would be spliced out and that the last exon will also be preserved. Complex B is made up of complex A and complex B. Then, in order to produce the complex C, a sequence of sophisticated rearrangement activities take place (catalytic step 1 spliceosome). The U1 and U4 depart complex B after the U6 snRNP replaces the U1 interaction at the donor site. This freshly generated complex C catalyzes two splicing transesterification steps. The phosphate just at 50 splice site is targeted mostly by 20 hydroxyl group there at branch point during the first transesterification, causing the 50 ending of the intron to also be separated from of the upstream exon but subsequently linked to the transition point through a phosphodiester link. The ADD references as per journal style phosphate just at intron's 30 splice site is targeted by a downstream exon's 30 hydroxyl in the second transesterification. The intron is eventually released, and the two exons create a

phosphodiester connection [6].

Alternative splicing mechanisms

The interplay of cisacting regulatory regions and matching trans-acting regulatory proteins regulates alternative splicing. Splicing activity at neighbouring splice sites can be enhanced (enhancers) or inhibited (silencers) by two distinct methods of cis-acting elements. Splicing enhancers include exonic splicing enhancer (ESEs) with intron splicing enhancers (ISEs) (intronic splicing enhancers, ISEs). They link to splicing activator proteins such as the serine/arginine-rich nuclear phosphoproteins family to increase the likelihood of a similar site being spliced (SR protein family). When splicing silencers are found in the surrounding, introns are termed called exonic splicing silencers (ESSs) as well as intronic splicing silencers (ISSs). They bind to splicing repressor proteins such heterogeneous nuclear ribonucleoproteins (hnRNPs) to inhibit or promote exon skip between adjacent exons.

A single regulator's presence or absence can occasionally be enough to alternative splice a pre-mRNA [Hampson 2016]. To generate an alternate splicing route, multiple cis-acting mainly sequences can collaborate with splicing activators and repressors to increase or inhibit spliceosome activity just at splice location [Kelly 2018,Giraudo 2004 Hampson 2006,Arbeit 1996]. Splicing activators, such as SR proteins, and splicing repressors, such as hnRNP proteins, are not always the same. In mammals, HnRNPs are the most well-known RNA-binding protein family [Arbeit 1996]. hnRNPA1 binds to ESS or ISS in humans to restrict exon addition by steric functions, or to the both side of a cassette exon to produce a loop that allows the exon to be skipped [27,28]. hnRNPA1 increases exon 6 inclusion whilst also interacting with the allocation of 50 splice donor area for exon 5 when it interacts with Fas pre-mRNA [Maione 2009]. hnRNPL, another member of the family, can both activate and repress exon 5 of the CD45 gene [Ragnum 2015]. Mostly, the propertise of splicing factors domenstrate to be position dependent in many aspects [Pan 2015]. Depending on whether it is coupled to exons or introns, a splicing factor can behave as a splicing suppressor or activator in various pre-mRNAs. One of these main splicing factors is Nova-1, a neuron-specific RNA-binding protein. Over 91% of Nova-dependent exon addition things happened along either alternative Fifty splice sites or conventional 30 splice sites, according to CLIP analysis, while 74% of Nova-dependent exon skip present near continuous 50 splice sites [Epstein 2007].

Splicing factors are important factor for the control of splicing . If we can pinpoint the binding locations of RNA and splicing factors, we can learn lots of about splicing control. Cross-linking followed by immunoprecipitation (CLIP) is the most used approach for studying RNA-protein binding sites [Monini 2009,Hampson 2014]. CLIP, in combination with the high sequencing, allows for transcriptome-wide analysis of RNAs that are also associated with a particular protein [34]. Methods based on CLIP have already been established in a variety of ways. The three most important which are slightly elevated sequenced CLIP (HITS-CLIP or CLIP-Seq) [Epstein 2014], photoactivatable-ribonucleoside enhanced CLIP (PAR-CLIP) [Debock 2011], and individual CLIP (iCLIP) [Carmeliet 2011]. The Nova protein family was the first to use HITSCLIP to find genome-wide protein-RNA interactions [32]. It is a well-established and successful approach, albeit because of the low cross-linking effectiveness, the false negative rate is considerable [Martin 2019]. Because photoreactive ribonucleoside analogues are integrated into PAR-CLIP, it has a greater efficiency than HITS-CLIP. As a result of this method, the signal-to-noise ratio and resolution both have enhanced [Debock 2011,Bourboulia 2010]. The treatment's potential drawback is it may be damaging [38,39], despite evidence to the contrary [Debock 2011, Bourboulia 2010]. In terms of convenience or efficiency, iCLIP exceeds the first two. The computational analysis or experiments, on the other hand, are both difficult. three have helped to expand our genomic sequence map of protein RNA associations. Experiments based on CLIPs, as well as various versions. CLIP-based technique choosing, experimental design and bioinformatic pipeline selection have all been substantially investigated [Jackson 2017,Passini 2011,Geib 2009]. Researchers can use these high-resolution mapping of protein-RNA interactions to learn more about how splicing is controlled and also how protein-RNA connection affect the human pathophysiological processes.. Furthermore, detecting RNA-protein interactions can provide biomarkers and, more critically, possible therapeutic targets [Martin 2019].

This protein-RNA interaction was not the only factor in alternative splicing control. RNA structure is also crucial [Bruno 2004]. They have the power to either encourage or prevent splicing. Because

RNA structures put splicing signals closer together, they improve splicing. A stemloop is created in the adenovirus ADML gene, for example, to put the 30 splice site closer to three alternative branch locations [Rodriguez 2009]. Another wonderful example of RNA structure facilitating alternative splicing is a docking and selection site in the Drosophila Melanogaster DSCAMgene. By mutually exclusive splicing of 95 different exons, DSCAM encodes over 38,000 unique mature mRNA isoforms [Quesada 2012].The majority of its variability is due to an exon 6 cluster, that has 48 alternative exons. A preserved selection sequence is complementary to a part of the docking site. Only one of the 48 alternative exons is employed to create mature mRNA due to this RNA-RNA base pairing between the docking site and selection region [Fica 2013]. This pattern reveals how competitive RNA secondary patterns are important for protein isoform formation and maintenance. Natural RNA structures commonly sequester key splicing regions, restricting the utilization of splice sites, which reduces splicing. In other situations, structures surrounding exons and splice sites may also have a negative effect on the assembly of U1 or U2 snRNPs [Long 2009]. To summarize, this sensitive system controls alternative splicing in a variety of biological processes.

Alternative splicing associated human disease

Given the importance of pre-mRNA splicing for protein diversity and organism performance, it's no surprise that disruption of regular splicing sequences can result in gene malfunction and even sickness. There are almost 150,000 transcript isoforms in the human genome, despite the fact that there are only about 20,000 protein-coding genes. As a conclusion, each human gene has seven transcript isoforms on average. Meanwhile, a recent study found that local splicing changes account for nearly 30% of tissue-dependent transcript variations [Quesada 2012].

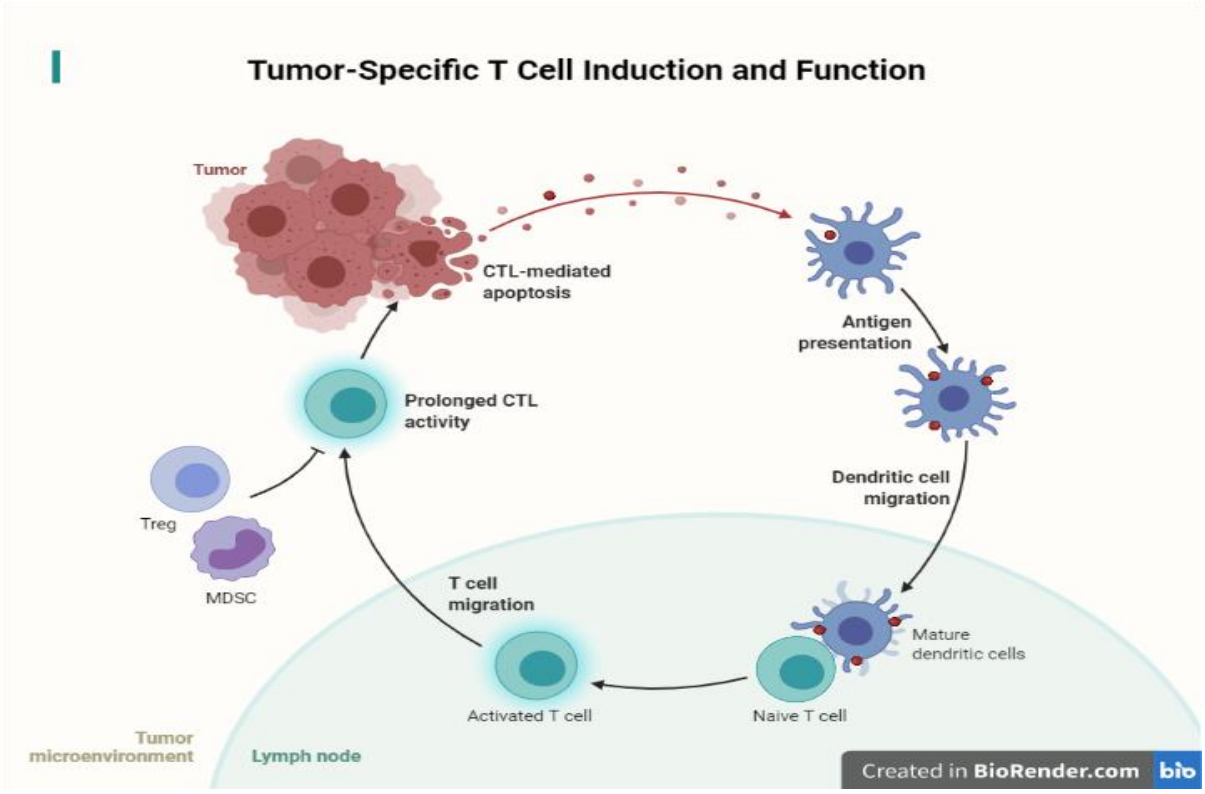

Figure:2 Tumor specific T Cell induction and function

Given that the flexibility of each gene with alternative splicing exemplifies the high level of human genome complexity, it's easy to think of such flexibility as a risk factor. Splicing errors had connected to several human disorders, cause mutations that impacts either trans acting proteins

involved and cis acting regulatory regions cause irregular splicing. Several kind of mutations are less prevalent than trans-acting mutations. It's most probable because mutations that affect a single gene's splicing in cis are generally more detrimental than mutations that affect a single gene's splicing in trans [Hua 2010]. We'll go over some well-known diseases resulting from mutations in trans-acting splicing factors as well as cis-acting regulatory areas, both of which induce abnormal alternative splicing in diverse ways.

Mutations in trans-splicing factors resulting in tumorigenesis

In humans, alternative splicing regulates gene expression in a pervasive and flexible manner. As a result, cancer cells frequently use this property in order to multiply and survive. Myelodysplastic syndromes (MDS) are bone marrow malignancies produced by immature, improperly produced blood cells [Bai 2014]. The most frequently changed gene in MDS victims is SF3B1, which encodes subunit 1 of both the splicing factors 3b binding protein, that appears to be an important factor of U2 snRNP [Adesso 2013;Das 2014]. Main point sequences from pre-mRNA bind to splicing factors 3b but also 3a. To engage and attach U2 snRNP to pre-mRNA, stable binding is required [65]. SF3B1 also plays a function in the commitment complex [Alamancos 2015] and is a component of the minor U12-type spliceosome [Kahles 2016]. SF3B1 knockdown inhibits development and causes dysregulation of a variety of genes and pathways [Adesso 2013]. Because SF3B1 is so vital as in spliceosome apparatus, this is no unknown that the RNA-seq analysis of tumour tissues containing SF3B1 mutations revealed widespread splicing problems caused by either a loss of branch section fidelity [Trapnell 2010]. Somatic mutations in splicing factors that understand 30 splice region , like U2AF1 as well as SRSF2, have already been linked to MDS [Katz 2010]. In those cited reviews [Bray 2016], the exact mechanism and routes of aberrant splicing in cancer growth have been published and explored thoroughly. More crucially, the discovery of mutations in those splicing components raises the possibility that the spliceosome machinery could be used to treat some tumours [Adesso 2013,Das 2014].

Spliceostatin A, a potent antitumor compound inhibiting splicing

Defects in the spliceosome machinery, as we've seen, cause substantial dysfunctions and diseases. Spliceostatin A (SSA), which inhibits pre-mRNA splicing in cancer cells, has been shown to decrease their growth [Pertea 2015:Stejjger 2013]. SSA is really a methylated precursor to FR901464, a naturally occurring chemical that suppresses pre-mRNA splicing in vitro and in vivo models by binding to SF3b, a subcomplex of the U2 snRNP protein that is required for premRNA branchpoint identification. SSA disrupts the binding of SF3b 155-kDa subunit with mRNA, resulting in non - productive stimulation of U2 snRNP to the a 50 branchpoint.

Furthermore, downregulation of genes needed for cell growth explained the anti-proliferative impacts of SSA [Stejjger 2013]. Because certain viruses require splicing to infect people, SSA is not only a promising cancer therapeutic target, but it can also reduce viral replication [Stejjger 2013:Roberts 2013].

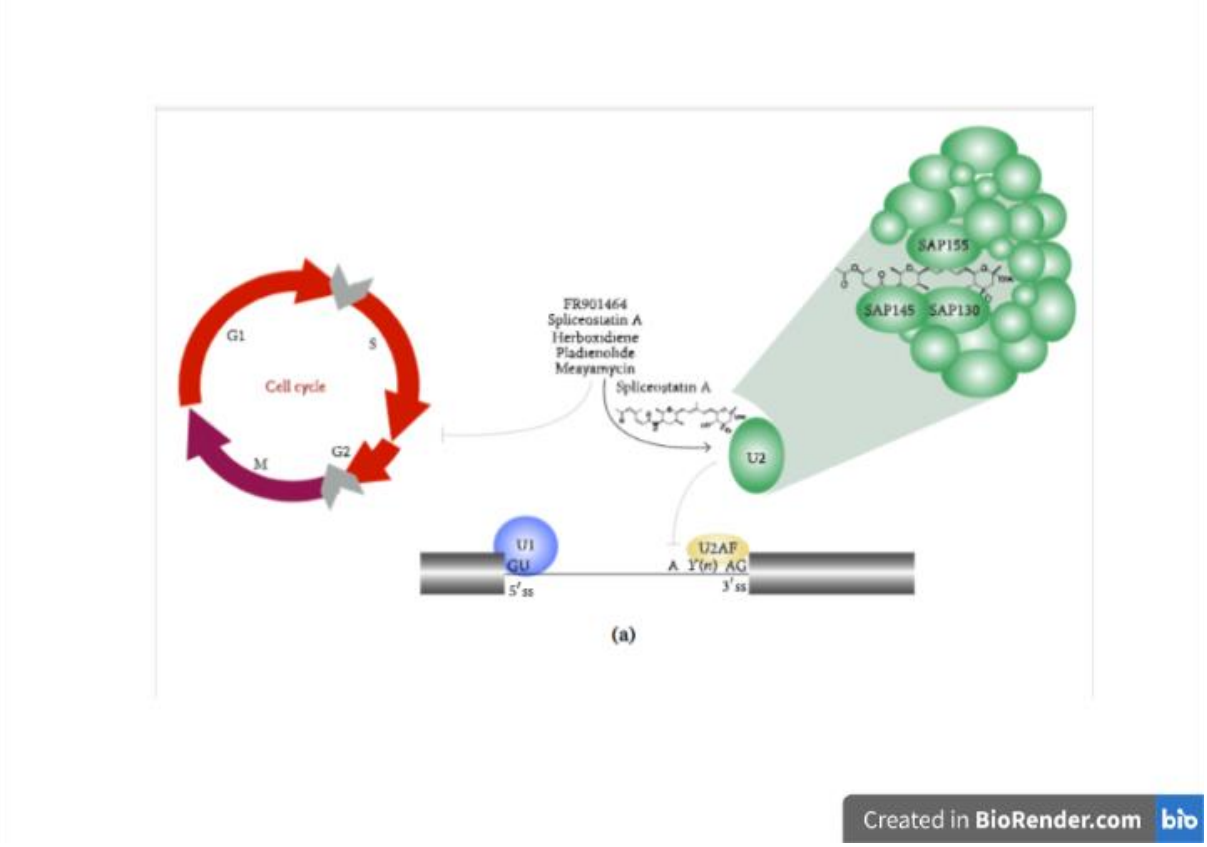

Figure 3 Molecular mechanism of Natural compounds that block splicing

Computational tools for isoform quantification from RNA-seq

Parallel computing Rna - seq (RNA-seq) has evolved as a potent method for thorough transcriptome analysis in recent years. Large-scale transcription and splicing experiments have also become possible relatively low cost of sequence analysis. More splicing-related disorders will be discovered faster as a result of these investigations, and the knowledge obtained will help develop preventive and therapeutic approaches for these diseases. However, technological limitations, primarily the restricted read duration, continue to obstruct precise quantification of alternative splicing. mRNA is taken from of the tissue, fragmented, then reverse transcribed into cDNA, that is then amplified and sequenced using greater, short-read sequencing technology during an RNA-seq experiment. Transcriptome arrangement might be used to rebuild transcribed genomic regions in an ideal environment. Given the typical read length of RNA-seq, which varies from 50 to 150 bp, and also the facts that transcript variants of the same gene may have varying record lengths, highly parallel RNA sequencing (RNA-seq) has emerged in last years as a powerful method for detailed transcriptome study. Large-scale expression and splicing investigations have also become possible relatively low cost of sequence analysis. More splicing-related disorders will be discovered faster as a result of these investigations, and the knowledge obtained will help develop preventive and therapeutic approaches for these diseases. However, technological limitations, primarily the restricted read duration, continue to obstruct precise quantification of alternative splicing mRNA is taken from tissue, fragmented, then reverse transcribed into cDNA, which then is amplified and sequenced using high-throughput, brief sequencing technology in an RNA-seq experiment. Transcriptome arrangement might be used to rebuild transcribed genomic regions in an ideal environment. However, because RNA-seq read lengths typically range from 50 to 150 bp, and because transcript isoforms of the same gene

are sometimes difficult to identify, quantifying transcript isoform expression remains problematic. Furthermore, one of the most difficult jobs in RNA-seq is the detection of new transcripts utilising short reads [Conesa 2016]. nferring full-length transcripts is more challenging with short reads that infrequently span splice junctions, especially for lowly released transcripts. Discovering transcription beginning and end locations is likewise a difficult task [Stejjger 2013]. As the sequence depth continues to increase, another calculating challenge has lately surfaced as a result of the affordable sequencing. The vast sequence data would not only tax hardware resource, but it may also confound heuristic algorithms which aren't designed to handle such large data sets [Roberts 2013].

In the recent decade, a variety of computational techniques have been created to address these major issues. both Isoform and exon-centric techniques are the two types of alternative splicing analyses. Using isoform-centric techniques to measure full-length transcript isoforms, an alternative splicing ratio could be calculated as the ratio of isoforms having and isoforms except for the target alternative exon. Approaches that are mainly concerned about estimating exon inclusion ratios are known as exon-centric methods. Transcript quantification can be classified as term referencing or de novo assembly based based on how a reference genome or reference transcriptome are used [80]. Reads are initially compared to the reference genome in reference-based approaches. The mapped reads are then computationally allocated to known annotated isoforms. Furthermore, the mapped reads could be used to build a splicing network that contains all viable splicing events, and then specific isoforms can be produced by traversing the graph. The reference-based method's great sensitivity also allows for the discovery of novel transcripts [Li 2011]. Several prominent techniques, such as StringTie [Pertea 2015] and Cufflink [Zhang 2014], for example, use genome-aligned reads and existing annotations to assemble transcripts, including novel ones. Quantification of transcript isoforms can be done immediately or afterwards. Although the majority of reference-based quantification methods link reading to a genome, alignment free approaches have lately become accessible. Pre-indexing reference transcripts, dividing sequence reads into k-mers, and performing "pseudo"-alignments on the pre-indexed transcripts are all alignment-free methods.

We will concentrate on term relating isoform-centric or exon-centric approaches in this review. There will be a discussion of relevant packages for identifying and assessing isoforms or splicing actions, even without novel isoforms or splicing events. We compiled a list of benchmark criteria for these sample tools based on simulated experiments, literature research [Zhang 2016:Li 2011], and programme documentation to make user selection easier. We will shortly mention de novo assembly based approaches afterwards.

Isoform-centric analysis tools

Cufflinks

Despite the fact that Cufflinks is undoubtedly one of the most popular tools for expression quantification, it was developed almost a decade ago, particularly for combined novel transcript discovery and excess calculation [Trapnell 2010]. PASA [Haas 2003], which compresses alignments to transcripts that depend on splicing compliance, and Dilworth's theorem [Eriksson 2008], which has been originally used to assemble a parsimonious collection of haplotypes using sequenced copies of a hybrid viral population, are both used in Cufflinks. The method simplifies the transcript assembly challenge by looking for the best fit in a balanced bipartite graph of fragment compatibilities. According to the results of their validation, Cufflinks is capable of not only enhancing transcriptome-based genome annotations but also discovering additional transcripts.

StringTie

Another highly effective transcript quantification tool that can also detect novel transcripts is StringTie [Pertea 2015]. It uses a special network flow approach and an additional de novo building stage to concurrently discover and quantify transcripts. StringTie acquired 53 percent additional transcripts than Cufflinks in such a benchmark with 90 million measurements from human blood. It is also faster and consumes less memory. The new Tuxedo protocol (HISAT, StringTie, Ballgown) is faster than previous Tuxedo protocol (Tophat, Cufflinks, Cuffdiff) [79], consumes less memory, and produces more accurate results overall. Cuffdiff, DESeq2, edgeR, and other downstream specialised tools are compatible with StringTie's output.

The RSEM approach is also extensively used to calculate relative abundance of genes and isoforms using single-end with paired end RNA-Sequence data [Zhang 2016,Kaanitz 2015,Compeau 2011]. Original FASTQ

sequences as well as BAM/SAM alignment files are accepted by RSEM. For read alignment, it uses Bowtie2, STAR, and HISAT binders on the inside. Users can also provide aligned BAM/SAM files to choose their chosen alternative aligner. RSEM is an Expectation-Maximization (EM) read pseudoalignment that employs fast k-mer hashing and the transcriptome de Bruijn graphBoth of these proteins have been shown to be important for DNA and RNA synthesis [78]. The equivalence classes allow for a simpler posterior distribution and speedier algorithm convergence by grouping pseudoalignments from same transcripts together. Because its pseudoalignments explicitly conserve the information offered by k-mers across reads, Kallisto's accuracy is comparable to that of alignment-based RNAseq quantification techniques, but its speed is two orders of magnitude faster.

Exon-centric analysis tools

MISO

As previously stated, event-based techniques to alternative splicing analysis are becoming more popular. MISO (Mixture-of-Isoforms) is just a statistical approach for detecting and measuring spliced exons and isoforms which are regulated differentially across datasets [Katz 2010]. Using a probabilistic technique, the MISO model evaluates the likelihood of reading coming from a certain isoform (Bayesian inference). For quantification and differential recognition across samples, MISO enhances the accuracy of the both exon-centric and isoform-centric analysis.

SUPPA

SUPPA [Alamancos 2015] is another useful tool for looking into alternative splicing occurrences. SUPPA uses an input annotation file to generate alternative splicing events. It uses quick transcript quantification to calculate the proportion called "percent spliced-in" (PSI) that determine the relative quantity of splicing mechanisms or transcript isoforms. Differential splicing is also measured using differences in the relative abundances (DPSI) between distinct samples. SUPPA outperforms most traditional methods based on analysis of experimentally proven events in terms of computational speed and accuracy. For splicing analysis of large datasets, SUPPA would be a viable and cost-effective option. SUPPA is limited to labeled splicing events on it's own and, but it becomes a powerful tool for detecting novel splicing events when coupled with alternative transcript reconstruction tools like StringTie.

SplAdder

SplAdder is a powerful tool for detecting unique splicing events, which conventional event-based techniques are incapable of performing. SplAdder avoids the computational burden of identifying whole transcripts by using a new method that analyses individual splicing processes as proxies to transcriptome properties [Kahles 2016]. Depending on the given annotation, SplAdder creates a splicing graph for all of a gene's transcripts. The splicing graph has been updated with additional data from RNA-seq (i.e. new introns or exon segments detected the alignment). The enriched annotation graph may be used to identify alternative splicing events, which can then be evaluated using RNA-seq data. SplAdder can detect all of the primary issues, however it is hampered by the computational accuracy issue caused by short reads. Longer reads are possible thanks to recent advances in 3rd sequencing (long-read technology), like Pacific Biosciences' single molecule real-time sequencing (SMRT) and Oxford's Nanopore sequencing. Because of the precision of isoform quantification, direct measurement of splicing processes with reads larger than a full-length transcript may soon be possible.

CONCLUSION

According to the experts suggested the concept of alternative splicing more than four decades ago which involves splicing together a unique combination of exons to create unique differentiation products. We're learning so much about alternative splicing thanks to molecular research, higher sequencing, and bioinformatic techniques, and there's still a lot more to understand on how it works so at cellular level.

The measurement of full-length transcripts with highly expressed genes is one of the two primary stumbling hurdles of alternative splicing study focused on RNA-seq, as previously mentioned. Advances in two types of technologies, we believe, may be able to resolve those concerns in the near future. 1) Long-read sequencing, like as SMRT and Nanopore sequencing, may sequence a transcript in its entirety without the requirement for reconstruction. Comprehensive transcripts including alternate splicing events have been discovered using these technologies. 2) Separate RNA-seq (scRNA-seq) uses next-generation sequence technology to measure gene expression at the cellular level. scRNA-seq offers

the ability to better understand highly expressed transcripts and splicing events with the right methods and bioinformatic pipeline.Both technologies, without a question, have limitations. comprehensive sequencing had higher error rates, detect de novo transcript and quantification challenging, whereas scRNA-seq provides few expressed genes due to the limited sequence depth or little difference between biological and technical noise. However, with the merging of modern technologies and traditional ones, as well as the development of increasingly powerful computing tools, In the distant future, a complete grasp of splicing will be obtained.

CONFLICT OF INTEREST

The authors declared that present study was performed in absence of any conflict of interest.

ACKNOWLEDGEMENT

Thanks to my worthy Masters Supervisor Kashif Jillani for his guidance and Professor Xing Li for his valuable suggestion

AUTHOR CONTRIBUTIONS

Muhammad Sikandar organized all the data and set with the help of other co authors Reeta Akram made illustrations for paper Umar Farooq and Zulfiqar ahmed make ammendments Shmasa Kanwal and Sumera Shabbir revise the paper

REFERENCES

Adesso, L., Calabretta, S., Barbagallo, F., Capurso, G., Pilozzi, E., Geremia, R., ... & Sette, C. (2013). Gemcitabine triggers a pro-survival response in pancreatic cancer cells through activation of the MNK2/eIF4E pathway. *Oncogene*, *32*(23), 2848-2857.

Adler, A. S., McCleland, M. L., Yee, S., Yaylaoglu, M., Hussain, S., Cosino, E., ... & Firestein, R. (2014). An integrative analysis of colon cancer identifies an essential function for PRPF6 in tumor growth. *Genes & development*, *28*(10), 1068-1084.

Alamancos, G. P., Pagès, A., Trincado, J. L., Bellora, N., & Eyras, E. (2015). Leveraging transcript quantification for fast computation of alternative splicing profiles. *Rna*, *21*(9), 1521-1531.

Arbeit, J. M., Howley, P. M., & Hanahan, D. (1996). Chronic estrogen-induced cervical and vaginal squamous carcinogenesis in human papillomavirus type 16 transgenic mice. *Proceedings of the National Academy of Sciences*, *93*(7), 2930-2935.

Arbyn, M., Redman, C. W., Verdoodt, F., Kyrgiou, M., Tzafetas, M., Ghaem-Maghami, S., ... & Moss, E. L. (2017). Incomplete excision of cervical precancer as a predictor of treatment failure: a systematic review and meta-analysis. *The Lancet Oncology*, *18*(12), 1665-1679.

Bacigalupo, I., Palladino, C., Leone, P., Toschi, E., Sgadari, C., Ensoli, B., & Barillari, G. (2017). Inhibition of MMP-9 expression by ritonavir or saquinavir is associated with inactivation of the AKT/Fra-1 pathway in cervical intraepithelial neoplasia cells. *Oncology letters*, *13*(5), 2903-2908.

Bai, B., Hales, C. M., Chen, P. C., Gozal, Y., Dammer, E. B., Fritz, J. J., ... & Peng, J. (2013). U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. *Proceedings of the National Academy of Sciences*, *110*(41), 16562-16567.

Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., ... & Frey, B. J. (2010). Deciphering the splicing code. *Nature*, *465*(7294), 53-59.

Barillari, G., Iovane, A., Bacigalupo, I., Labbaye, C., Chiozzini, C., Sernicola, L., ... & Ensoli, B. (2014). The HIV protease inhibitor indinavir down-regulates the expression of the pro-angiogenic MT1-MMP by human endothelial cells. *Angiogenesis*, *17*(4), 831-838.

Barillari, G., Iovane, A., Bacigalupo, I., Palladino, C., Bellino, S., Leone, P., ... & Ensoli, B. (2012). Ritonavir or saquinavir impairs the invasion of cervical intraepithelial neoplasia cells via a reduction of MMP expression and activity. *Aids*, *26*(8), 909-919.

Batman, G., Oliver, A. W., Zehbe, I., Richard, C.,

Hampson, L., & Hampson, I. N. (2011). Lopinavir up-regulates expression of the antiviral protein ribonuclease L in human papillomavirus-positive cervical carcinoma cells. *Antiviral therapy*, *16*(4), 515.

Bergers, G., Brekken, R., McMahon, G., Vu, T. H., Itoh, T., Tamaki, K., ... & Hanahan, D. (2000). Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nature cell biology*, *2*(10), 737-744.

Berman, E., Jhanwar, S., Hedvat, C., Arcila, M. E., Abdel-Wahab, O., Levine, R., ... & Albitar, M. (2016). Resistance to imatinib in patients with chronic myelogenous leukemia and the splice variant BCR-ABL135INS. *Leukemia research*, *49*, 108-112.

Blitz, S., Baxter, J., Raboud, J., Walmsley, S., Rachlis, A., Smaill, F., ... & Canadian Women's HIV Study Group. (2013). Evaluation of HIV and highly active antiretroviral therapy on the natural history of human papillomavirus infection and cervical cytopathologic findings in HIV-positive and high-risk HIV-negative women. *The Journal of infectious diseases*, *208*(3), 454-462.

Bodily, J., & Laimins, L. A. (2011). Persistence of human papillomavirus infection: keys to malignant progression. *Trends in microbiology*, *19*(1), 33-39.

Bourboulia, D., & Stetler-Stevenson, W. G. (2010, June). Matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs): Positive and negative regulators in tumor cell adhesion. In *Seminars in cancer biology* (Vol. 20, No. 3, pp. 161-168). Academic Press.

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology*, *34*(5), 525-527.

Bruni, L., Diaz, M., Castellsagué, M., Ferrer, E., Bosch, F. X., & de Sanjosé, S. (2010). Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. *Journal of Infectious Diseases*, *202*(12), 1789-1799.

Bruno, I. G., Jin, W., & Cote, G. J. (2004). Correction of aberrant FGFR1 alternative RNA splicing through targeting of intronic regulatory elements. *Human molecular genetics*, *13*(20), 2409-2420.

Carmeliet, P., & Jain, R. K. (2011). Molecular mechanisms and clinical applications of angiogenesis. *Nature*, *473*(7347), 298-307.

Charizanis, K., Lee, K. Y., Batra, R., Goodwin, M., Zhang, C., Yuan, Y., ... & Swanson, M. S. (2012). Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron*, *75*(3), 437-450.

Compeau, P. E., Pevzner, P. A., & Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nature biotechnology*, *29*(11), 987-991.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., ... & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome biology*, *17*(1), 1-19.

Dall, K. L., Scarpini, C. G., Roberts, I., Winder, D. M., Stanley, M. A., Muralidhar, B., ... & Coleman, N. (2008). Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. *Cancer research*, *68*(20), 8249-8259.

Das, S., & Krainer, A. R. (2014). Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Molecular Cancer Research*, *12*(9), 1195-1204.

De Bock, K., Cauwenberghs, S., & Carmeliet, P. (2011). Vessel abnormalization: another hallmark of cancer?: molecular mechanisms and therapeutic implications. *Current opinion in genetics & development*, *21*(1), 73-79.

Elit, L., Kennedy, E. B., Fyles, A., & Metser, U. (2016). Follow-up for cervical cancer: a program in evidence-based care systematic review and clinical practice guideline update. *Current Oncology*, *23*(2), 109-118.

Epstein, R. J. (2007). VEGF signaling inhibitors: more pro-apoptotic than anti-angiogenic. *Cancer and Metastasis Reviews*, *26*(3), 443-452.

Eriksson, N., Pachter, L., Mitsuya, Y., Rhee, S. Y., Wang, C., Gharizadeh, B., ... & Beerenwinkel, N. (2008). Viral population estimation using pyrosequencing. *PLoS computational biology*, *4*(5), e1000074.

Esposito, V., Palescandolo, E., Spugnini, E. P., Montesarchio, V., De Luca, A., Cardillo, I., ... & Chirianni, A. (2006). Evaluation of antitumoral properties of the protease inhibitor indinavir in a murine model of hepatocarcinoma. *Clinical cancer research*, *12*(8), 2634-2639.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S.,

Mathers, C., Rebelo, M., ... & Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*, *136*(5), E359-E386.

Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., ... & Piccirilli, J. A. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature*, *503*(7475), 229-234.

Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., ... & Piccirilli, J. A. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature*, *503*(7475), 229-234.

Forman, D., de Martel, C., Lacey, C. J., Soerjomataram, I., Lortet-Tieulent, J., Bruni, L., ... & Franceschi, S. (2012). Global burden of human papillomavirus and related diseases. *Vaccine*, *30*, F12-F23.

Fox, R. G., Lytle, N. K., Jaquish, D. V., Park, F. D., Ito, T., Bajaj, J., ... & Reya, T. (2016). Image-based detection and targeting of therapy resistance in pancreatic adenocarcinoma. *Nature*, *534*(7607), 407-411.

Fu, X. D., & Ares, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews Genetics*, *15*(10), 689-701.

Geib, T., & Hertel, K. J. (2009). Restoration of full-length SMN promoted by adenoviral vectors expressing RNA antisense oligonucleotides embedded in U7 snRNAs. *PloS one*, *4*(12), e8204.

Giraudo, E., Inoue, M., & Hanahan, D. (2004). An amino-bisphosphonate targets MMP-9–expressing macrophages and angiogenesis to impair cervical carcinogenesis. *The Journal of clinical investigation*, *114*(5), 623-633.

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith Jr, R. K., Hannick, L. I., ... & White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic acids research*, *31*(19), 5654-5666.

Hampson, L., Kitchener, H. C., & Hampson, I. N. (2006). Specific HIV protease inhibitors inhibit the ability of HPV16 E6 to degrade p53 and selectively kill E6-dependent cervical carcinoma cells in vitro. *Antiviral therapy*, *11*(6), 813.

Hampson, L., Maranga, I. O., Masinde, M. S., Oliver, A. W., Batman, G., He, X., ... & Hampson, I. N. (2016). A single-arm, proof-of-concept trial of lopimune (lopinavir/ritonavir) as a treatment for HPV-related pre-invasive cervical disease. *PLoS One*, *11*(1), e0147917.

Hua, Y., Sahashi, K., Hung, G., Rigo, F., Passini, M. A., Bennett, C. F., & Krainer, A. R. (2010). Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes & development*, *24*(15), 1634-1644.

Jackson, H. W., Defamie, V., Waterhouse, P., & Khokha, R. (2017). TIMPs: versatile extracellular regulators in cancer. *Nature Reviews Cancer*, *17*(1), 38-53.

Jain, R. K. (2014). Antiangiogenesis strategies revisited: from starving tumors to alleviating hypoxia. *Cancer cell*, *26*(5), 605-622.

Kahles, A., Ong, C. S., Zhong, Y., & Rätsch, G. (2016). SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. *Bioinformatics*, *32*(12), 1840-1847.

Kanitz, A., Gypas, F., Gruber, A. J., Gruber, A. R., Martin, G., & Zavolan, M. (2015). Comparative assessment of methods for the computational inference of transcript isoform abundance from RNA-seq data. *Genome biology*, *16*(1), 1-26.

Katz, Y., Wang, E. T., Airoldi, E. M., & Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature methods*, *7*(12), 1009-1015.

Kelly, H., Weiss, H. A., Benavente, Y., de Sanjose, S., Mayaud, P., Qiao, Y. L., ... & Hoffman, R. (2018). Association of antiretroviral therapy with high-risk human papillomavirus, cervical intraepithelial neoplasia, and invasive cervical cancer in women living with HIV: a systematic review and meta-analysis. *The lancet HIV*, *5*(1), e45-e58.

Kim, J., Hu, C., Moufawad El Achkar, C., Black, L. E., Douville, J., Larson, A., ... & Yu, T. W. (2019). Patient-customized oligonucleotide therapy for a rare genetic disease. *New England Journal of Medicine*, *381*(17), 1644-1652.

Kreimer, A. R., Schiffman, M., Herrero, R., Hildesheim, A., González, P., Burk, R. D., ... & Cecilia Rodriguez, A. (2012). Long-term risk of recurrent cervical human papillomavirus infection and precancer and cancer following excisional treatment. *International journal of cancer*, *131*(1), 211-218.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data

with or without a reference genome. *BMC bioinformatics*, *12*(1), 1-16.

Long, J. C., & Caceres, J. F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal*, *417*(1), 15-27.

Maione, F., Capano, S., Regano, D., Zentilin, L., Giacca, M., Casanovas, O., ... & Giraudo, E. (2012). Semaphorin 3A overcomes cancer hypoxia and metastatic dissemination induced by antiangiogenic treatment in mice. *The Journal of clinical investigation*, *122*(5), 1832-1848.

Maione, F., Molla, F., Meda, C., Latini, R., Zentilin, L., Giacca, M., ... & Giraudo, E. (2009). Semaphorin 3A is an endogenous angiogenesis inhibitor that blocks tumor growth and normalizes tumor vasculature in transgenic mouse models. *The Journal of clinical investigation*, *119*(11), 3356-3372.

Martin, J. D., Seano, G., & Jain, R. K. (2019). Normalizing function of tumor vessels: progress, opportunities, and challenges. *Annual review of physiology*, *81*, 505-534.

Martinez-Contreras, R., Fisette, J. F., Nasim, F. U. H., Madden, R., Cordeau, M., & Chabot, B. (2006). Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS biology*, *4*(2), e21.

Monini, P., Sgadari, C., Grosso, M. G., Bellino, S., Di Biagio, A., Toschi, E., ... & Ensoli, B. (2009). Clinical course of classic Kaposi's sarcoma in HIV-negative patients treated with the HIV protease inhibitor indinavir. *Aids*, *23*(4), 534-538.

Pan, D., Wei, K., Ling, Y., Su, S., Zhu, M., & Chen, G. (2015). The prognostic role of Ki-67/MIB-1 in cervical cancer: a systematic review with meta-analysis. *Medical science monitor: international medical journal of experimental and clinical research*, *21*, 882.

Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., ... & Fu, X. D. (2013). Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Molecular cell*, *50*(2), 223-235.

Passini, M. A., Bu, J., Richards, A. M., Kinnecom, C., Sardi, S. P., Stanek, L. M., ... & Cheng, S. H. (2011). Antisense oligonucleotides delivered to the mouse CNS ameliorate symptoms of severe spinal muscular atrophy. *Science translational medicine*, *3*(72), 72ra18-72ra18.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols*, *11*(9), 1650-1667.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols*, *11*(9), 1650-1667.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology*, *33*(3), 290-295.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology*, *33*(3), 290-295.

Pietras, K., Pahler, J., Bergers, G., & Hanahan, D. (2008). Functions of paracrine PDGF signaling in the proangiogenic tumor stroma revealed by pharmacological targeting. *PLoS medicine*, *5*(1), e19.

Qiu, Y., Maione, F., Capano, S., Meda, C., Picconi, O., Brundu, S., ... & Giraudo, E. (2020). HIV Protease Inhibitors Block HPV16-Induced Murine Cervical Carcinoma and Promote Vessel Normalization in Association with MMP-9 Inhibition and TIMP-3 Induction. *Molecular Cancer Therapeutics*, *19*(12), 2476-2489.

Quesada, V., Conde, L., Villamor, N., Ordóñez, G. R., Jares, P., Bassaganyas, L., ... & López-Otín, C. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature genetics*, *44*(1), 47-52.

Quesada, V., Conde, L., Villamor, N., Ordóñez, G. R., Jares, P., Bassaganyas, L., ... & López-Otín, C. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature genetics*, *44*(1), 47-52.

Ragnum, H. B., Vlatkovic, L., Lie, A. K., Axcrona, K., Julin, C. H., Frikstad, K. M., ... & Lyng, H. (2015). The tumour hypoxia marker pimonidazole reflects a transcriptional programme associated with aggressive prostate cancer. *British journal of cancer*, *112*(2), 382-390.

Riley, R. R., Duensing, S., Brake, T., Münger, K.,

Lambert, P. F., & Arbeit, J. M. (2003). Dissection of human papillomavirus E6 and E7 function in transgenic mouse models of cervical carcinogenesis. *Cancer research*, *63*(16), 4862-4871.

Roberts, A. (2013). *Ambiguous fragment assignment for high-throughput sequencing experiments*. University of California, Berkeley.

Rodriguez-Martin, T., Anthony, K., Garcia-Blanco, M. A., Mansfield, S. G., Anderton, B. H., & Gallo, J. M. (2009). Correction of tau mis-splicing caused by FTDP-17 MAPT mutations by spliceosome-mediated RNA trans-splicing. *Human molecular genetics*, *18*(17), 3266-3273.

Rositch, A. F., Koshiol, J., Hudgens, M. G., Razzaghi, H., Backes, D. M., Pimenta, J. M., ... & Smith, J. S. (2013). Patterns of persistent genital human papillomavirus infection among women worldwide: a literature review and meta-analysis. *International journal of cancer*, *133*(6), 1271-1285.

Seiler, M., Peng, S., Agrawal, A. A., Palacino, J., Teng, T., Zhu, P., ... & Van Den Berg, D. J. (2018). Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. *Cell reports*, *23*(1), 282-296.

Sgadari, C., Barillari, G., Toschi, E., Carlei, D., Bacigalupo, I., Baccarini, S., ... & Ensoli, B. (2002). HIV protease inhibitors are potent anti-angiogenic molecules and promote regression of Kaposi sarcoma. *Nature medicine*, *8*(3), 225-232.

Sgadari, C., Monini, P., Barillari, G., & Ensoli, B. (2003). Use of HIV protease inhibitors to block Kaposi's sarcoma and tumour growth. *The lancet oncology*, *4*(9), 537-547.

SP, Y. D., & Hewett, P. W. (1997). Johnson ir, Carmichael J, murray JC. angiogenesis is associated with vascular endothelial growth factor expression in cervical intraepithelial neoplasia. *Br J Cancer*, *76*, 14140-5.

Steijger, T., Abril, J. F., Engström, P. G., Kokocinski, F., Hubbard, T. J., Guigó, R., ... & Bertone, P. (2013). Assessment of transcript reconstruction methods for RNA-seq. *Nature methods*, *10*(12), 1177-1184.

Toschi, E., Sgadari, C., Malavasi, L., Bacigalupo, I., Chiozzini, C., Carlei, D., ... & Ensoli, B. (2011). Human immunodeficiency virus protease inhibitors reduce the growth of human tumors via a proteasome-independent block of angiogenesis and matrix metalloproteinases. *International journal of cancer*, *128*(1), 82-93.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., ... & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, *28*(5), 511-515.

Witkiewicz, A. K., McMillan, E. A., Balaji, U., Baek, G., Lin, W. C., Mansour, J., ... & Knudsen, E. S. (2015). Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nature communications*, *6*(1), 1-11.

Yeo, G. W., Van Nostrand, E., Holste, D., Poggio, T., & Burge, C. B. (2005). Identification and analysis of alternative splicing events conserved in human and mouse. *Proceedings of the National Academy of Sciences*, *102*(8), 2850-2855.

Zhang, C., Zhang, B., Vincent, M. S., & Zhao, S. (2016). Bioinformatics tools for RNA-seq gene and isoform quantification. *Next Generat. Sequenc. & Applic*, *3*, 140.