ALTERNATIVE SPLICING OF RNA

Abhinav Pandey
B.Tech Computer Science, AI
Amrita Vishwa Vidyapeetham
Kollam, Kerala
AM.EN.U4AIE21088

Abhishek.A
B.Tech Computer Science, AI
Amrita Vishwa Vidyapeetham
Kollam, Kerala
AM.EN.U4AIE21003

Aravind.MJ
B.Tech Computer Science, AI
Amrita Vishwa Vidyapeetham
Kollam, Kerala
AM.EN.U4AIE21017

Anfas Hassan
B.Tech Computer Science, AI
Amrita Vishwa Vidyapeetham
Kollam, Kerala
AM.EN.U4AIE21014

Rithesh.R
B.Tech Computer Science, AI
Amrita Vishwa Vidyapeetham
Kollam, Kerala
AM.EN.U4AIE21054

DSS Sandeep Chandra
B.Tech Computer Science, AI
Amrita Vishwa Vidyapeetham
Kollam, Kerala
AM.EN.U4AIE21079

ABSTRACT

The overall function of alternative splicing is to increase the diversity of mRNAs expressed from the genome. Alternative splicing changes proteins encoded by mRNAs, which has profound functional effects. Experimental analysis of these protein isoforms showed that alternative splicing regulates binding between proteins, between proteins and nucleic acids as well as between proteins and membranes. Alternative splicing regulates the localization of proteins, their enzymatic properties and their interaction with ligands. In most cases, changes caused by individual splicing isoforms are small. However, cells typically coordinate numerous changes in 'splicing programs', which can have strong effects on cell proliferation, cell survival and properties of the nervous system. Due to its widespread usage and molecular versatility, alternative splicing emerges as a central element in gene regulation that interferes with almost every biological function analyzed.

Keywords: transcription, RNA, DNA, translation, splicing, entrons, introns

INTRODUCTION

Alternative splicing is a molecular mechanism that modifies pre-mRNA constructs prior to translation. This process can produce a diversity of mRNAs from a single gene by arranging coding sequences (exons) from recently spliced RNA transcripts into different combinations. The mRNA transcripts created from alternative splicing can translate into varying amino acid sequences that produce protein isoforms with different functions. The mechanisms that regulate alternative splicing play a fundamental role in gene expression due to their spatial and temporal functions throughout biology.

Frequently transcripts contain several alternative exons and their usage can be combined, largely increasing the diversity of the mRNA expressed from the genome and giving alternative splicing a central role in forming complex organisms. Alternative splicing patterns constantly change under physiological conditions, allowing an organism to respond to changes in the environment by determining which part of the genome it expresses. Most of the changes in alternative splicing are studied in artificial experimental systems, but alternative exon usage changes in real life scenarios.

The stress of exams on medical students causes a change in alternative pre-mRNA splicing of the phosphatidylinositol 3-kinase-related protein kinase (SMG-1). This change may have later effects on nonsense-mediated RNA decay and the p53 pathway).

Alternative splicing can play a role even before life and after death. The importance of alternative splicing before fertilization is illustrated by Nitric Oxide Synthase 1 where splicing isoforms are involved in controlling the erectile function. The role after death is shown by the poor meat quality of turkeys that underwent transport and heat stress prior to slaughtering. This stress changes the splicing patterns of ryanodine receptors, ultimately

leading to an increase of water content in the meat and lowering the quality.

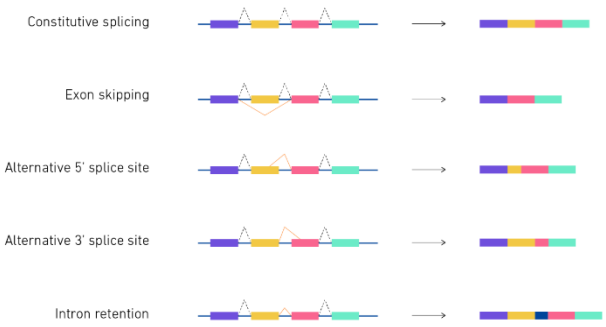
When introns are removed from the primary RNA transcript by RNA splicing, the remaining exons are spliced together to generate the final, mature mRNA. However, for most genes, the primary transcript can follow multiple alternative splicing pathways, leading to the synthesis of multiple related but different mRNAs, each of which can be subsequently translated to generate different protein products. Some of these alternative events are highly tissue- or cell type–specific, and, to the extent that such events are determined by primary sequence, they are subject to allelic variation between different individuals. Nearly all human genes undergo alternative splicing to some degree, and it has been estimated that there are an average of two or three alternative transcripts per gene in the human genome, thus greatly expanding the information content of the human genome beyond the approximately 20,000 protein-coding genes. The regulation of alternative splicing appears to play a particularly impressive role during neuronal development, where it may contribute to generating the high levels of functional diversity needed in the nervous system. Consistent with this, susceptibility to a number of neuropsychiatric conditions has been associated with shifts or disruption of alternative splicing patterns.

METHODOLOGY:

Prior to RNA splicing, RNA polymerase II produces pre-mRNA transcripts by transcribing gene sequences into a collection of non-coding introns and protein-coding exons. When these pre-mRNA sequences undergo constitutive splicing, the removal of introns is followed by the joining of exons in their DNA-corresponding order. Alternative splicing deviates from this process through mechanisms that rearrange the pattern of exons into alternative coding sequences that translate to different proteins.

The production of different proteins from a single gene results from the interplay between RNA-binding proteins and splice sites located throughout the pre-mRNA transcript.The primary RBP directly involved in alternative splicing is the spliceosome, a multi-unit complex consisting of small nuclear RNAs and various proteins. Once the spliceosome snRNAs recognize splicer sites along the pre-mRNA construct, the splicing proteins within this complex can alternate exons and remove introns accordingly.

Alternative splicing can be done in mostly 3 ways. They are as follows:



- Exon Skipping:

This process involves the removal of certain exons and their adjacent introns from mRNA constructs prior to translation.

Algorithm:

```
firstList = []
for x in range(0,len(exons)):
  combined = ""
  for y in range(0,len(exons)):
    if y ==x :
      continue
    combined = combined + exons[y]
  firstList.append(combined)
```

- Alternate 5’ or 3’ splicing:

Alternative splicing can also be mediated by the joining of exons at alternative 5’ or 3’ splice sites.

Algorithm:

```
secondList = []
combined1 = ""
combined2 = ""
for x in range(0,len(exons)):
  if x%2==0:
    combined1 = combined1 + exons[x]
  if x%2!=0:
    combined2 = combined2 + exons[x]
secondList.append(combined1)
secondList.append(combined2)
```

- Intron Retention:

This type happens when non-coding portions of a gene are retained in the final mRNA transcript.

Algorithm:

```
thirdList = []
for i in range(0, len(introns)):
    mixup = ""
    for j in range(0, len(exons)):
      mixup = mixup + exons[j];
      if(i == j):
        mixup = mixup + introns[i]
    thirdList.append(mixup)
```

We get three list of rna which are spliced and ready to be coverted into proteins. Now, if do the conversion on the protein according to the 3-split method, we will get different types of protein formation.

RESULTS:

When RNA polymerase I are created by transcription, we had to remove the non-coding regions called introns. After applying different methods of alternative splicing, we found that different types of RNA polymerase can be created from the same DNA. The RNA will go under translation and produce different proteins. So, different kinds of proteins are created in our body using the same gene by alternative splicing of RNA.

CONCLUSION:

Even though RNA splicing was originally discovered in the 1970s, the significance of alternative splicing for humans could not be thoroughly appreciated until the Human Genome Project determined that there are approximately 22,000 protein-coding genes that translate into over 90,000 different proteins. The completion of this global project provided a foundation of genomic data for the advancement of subsequent research projects like the Encyclopedia of DNA Elements (ENCODE) and the Human Proteome Project.

REFRENCES:

1. Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A. 1977;74(8):3171-3175. doi: 10.1073/pnas.74.8.3171
2. Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell. 1977;12(1):1-8. doi: 10.1016/0092-8674(77)90180-5
3. Darnell JE Jr. Implications of RNA-RNA splicing in evolution of eukaryotic cells. Science. 1978;202(4374):1257-1260. doi: 10.1126/science.364651
4. Ramanouskaya TV, Grinev VV. The determinants of alternative RNA splicing in human cells. Mol Genet Genomics. 2017;292(6):1175-1195. doi: 10.1007/s00438-017-1350-0
5. Wang Y, Liu J, Huang B, Xu Y, Li J, Huang L, et al. Mechanism of alternative splicing and its regulation (Review). Biomed Rep. 2015;2:152-158. doi: 10.3892/br.2014.407
6. Tharp CA, Haywood ME, Sbaizero O, Taylor MRG, Mestroni L. The giant Protein Titin's role in cardiomyopathy: genetic, transcriptional, and post-translational modifications of TTN and their contribution to cardiac disease. Front Physiol. 2019;10:1436. 2019. doi: 10.3389/fphys.2019.01436

APPENDIX: (CODE)

```
# Alternative-Splicing

# input is the DNA string and the introns present in it.

# STEP-1: Formatting a fasta file

f = open("dataset.fasta", "r")

lines = f.readlines()

DNA_string = ""

introns = []

for x in range(0,len(lines)):

  if x==1:

    DNA_string = lines[x]

    continue

  if x%2 != 0:

    lines[x] = lines[x].split("\n")[0]

    introns.append(lines[x])


# STEP2: Splitting of exons from introns:

exons = []

for x in introns:

  temp = DNA_string.split(x)

  exons.append(temp[0])

  DNA_string = temp[1]

exons.append(temp[1])


# STEP-3: Converting DNA to pre-RNA:

def dnaToRna(exon):
```

```
  output = ""
  for x in exon:
    if x == "G":
      output = output + "C"
    elif x == "C":
      output = output + "G"
    elif x == "T":
      output = output + "A"
    elif x == "A":
      output = output + "U"
  return output

for x in range(0,len(exons)):
  exons[x] = dnaToRna(exons[x])
for x in range(0,len(introns)):
  introns[x] = dnaToRna(introns[x])

# STEP-4: ALTERNATIVE SPLICING: 3 METHODS

# 1st METHOD: Exon skipping
#skipping only one exon
firstList = []
for x in range(0,len(exons)):
  combined = ""
  for y in range(0,len(exons)):
    if y ==x :
```

```
      continue
    combined = combined + exons[y]
  firstList.append(combined)



# 2nd METHOD: 5' or 3' splicing
secondList = []
combined1 = ""
combined2 = ""
for x in range(0,len(exons)):
  if x%2==0:
    combined1 = combined1 + exons[x]
  if x%2!=0:
    combined2 = combined2 + exons[x]
secondList.append(combined1)
secondList.append(combined2)

# 3rd METHOD: Intron retention
thirdList = []
for i in range(0, len(introns)):
    mixup = ""
    for j in range(0, len(exons)):
      mixup = mixup + exons[j];
      if(i == j):
        mixup = mixup + introns[i]
    thirdList.append(mixup)
```

```
# STEP-5: Converting spliced rna into proteins

table = dict()

table["phe"] = ["UUU","UUC"]

table["leu"] = ["UUA","UUG","CUU","CUC","CUA","CUG"]

table["ile"] = ["AUU","AUC","AUA"]

table["met"] = ["AUG"]

table["val"] = ["GUU","GUC","GUA","GUG"]

table["ser"] = ["UCU","UCC","UCA","UCG","AGU","AGC"]

table["pro"] = ["CUU","CCC","CCA","CCG"]

table["thr"] = ["ACU","ACC","ACA","ACG"]

table["ala"] = ["GCU","GCC","GCA","GCG"]

table["tyr"] = ["UAU","UAC"]

table["his"] = ["CAU","CAC"]

table["gin"] = ["CAA","CAG"]

table["asn"] = ["AAU","AAC"]

table["lys"] = ["AAA","AAG"]

table["asp"] = ["GAU","GAC"]

table["glu"] = ["GAA","GAG"]

table["cys"] = ["UGU","UGC"]

table["trp"] = ["UGG"]

table["arg"] = ["AGA","AGG","CGU","CGC","CGA","CGG"]

table["gly"] = ["GGU","GGC","GGA","GGG"]


def split_three(sequence):

  string = ""
```

```
  count = 0
  for i in sequence:
    string += i
    count += 1
    if count ==  3:
      string += " "
      count = 0
  return string


def returnKeyFromDict(splitted):
  protein = ""
  for x in table:
    if splitted in table[x]:
      protein = x.upper()
  if protein == "":
    return ""
  return protein


proteinForFirstMethod = []
for x in firstList:
  first = split_three(x).split(" ")
  outputFirst = ""
  for y in first:
    outputFirst = outputFirst + returnKeyFromDict(y) + " "
  proteinForFirstMethod.append(outputFirst)
```

```
proteinForSecondMethod = []
for x in secondList:
  second = split_three(x).split(" ")
  outputSecond = ""
  for y in second:
    outputSecond = outputSecond + returnKeyFromDict(y) + " "
  proteinForSecondMethod.append(outputSecond)

proteinForThirdMethod = []
for x in thirdList:
  third = split_three(x).split(" ")
  outputThird = ""
  for y in third:
    outputThird = outputThird + returnKeyFromDict(y) + " "
  proteinForThirdMethod.append(outputThird)

# OUTPUT
print("The proteins formed from Exon skipping are: ")
print(proteinForFirstMethod)
print("The proteins formed from 5'or 3' splicing are: ")
print(proteinForSecondMethod)
print("The proteins formed from Intron retention are: ")
print(proteinForThirdMethod)
```

SAMPLE DATASET:

>Rosalind_10

ATGGTCTACATAGCTGACAAACAGCACGTAGCAATCGGTCGAATCTCGAGAGGCATATGGT
CACATGATCGGTCGAGCGTGTTTCAAAGTTTGCGCCTAG

>Rosalind_11

TACATAG

>Rosalind_12

AGCACGT

>Rosalind_13

CGAGAGG

>Rosalind_15

AGCGTGT