Studying Alternative Splicing

Meelis Kull

PhD student in the University of Tartu

supervisor: Jaak Vilo

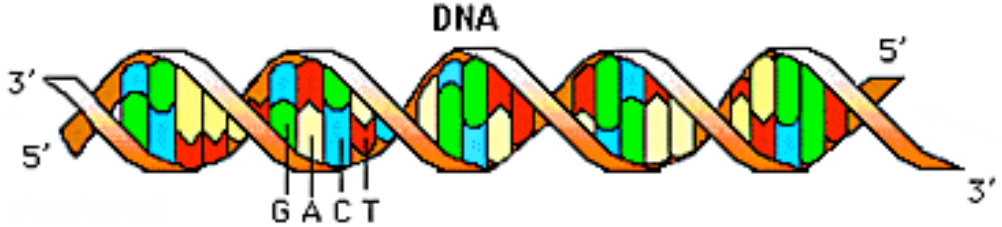
CS Theory Days

Rõuge 2007

Overview

- Alternative splicing
- Its biological function
- Studying splicing
  - Technology
  - Computational tasks
- Tissue-specific alternative splicing

The Basis of Life on Earth



RNA Splicing

- Folded DNA
  - Unfolding
- DNA
  - Transcription
- pre-mRNA
  - **Splicing**
- mRNA
  - Translation
- Poly-peptide
  - Protein folding
- Protein

Splicing

..GUAUCGUGGAAGGACUCAUGGUAUGAGAGCUGGGGAAUG..

..GGACUGAGGCUCCCACCUUUCUCAUCCAAGACUGGCUCC..

..UCCCUGCCGGGGCUGCGUGCAACCCUGGGGUUGGGGGUU..

..CUGGGGACUGGCUUUCCCAUAAUUUCCUUUCAAGGUGGG..

..GAGGGAGGUAGAGGGGUGAUGUGGGGAGUACGCUGCAGG..

..GCCUCACUCCUUUUGCAGACCACAGUCCAUGCCAUCACU..

Splicing

..GUAUCGUGGAAGGACUCAUGGUAUGAGAGCUGGGGAAUG..

..GGACUGAGGCUCCCACCUUUCUCAUCCAAGACUGGCUCC..

..UCCCUGCCGGGGCUGCGUGCAACCCUGGGGUUGGGGGUU..

..CUGGGGACUGGCUUUCCCAUAAUUUCCUUUCAAGGUGGG..

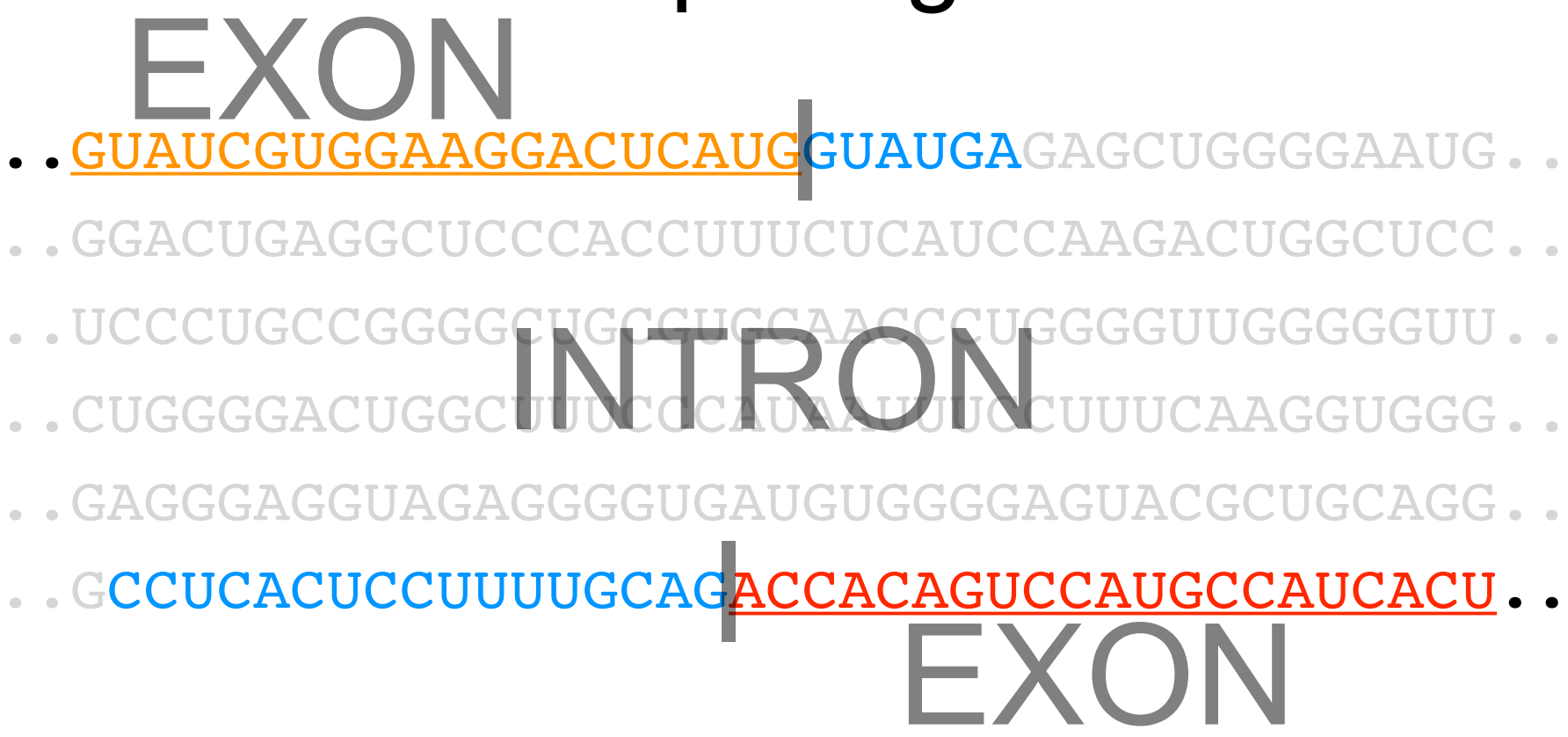
..GAGGGAGGUAGAGGGGUGAUGUGGGGAGUACGCUGCAGG..

..GCCUCACUCCUUUUGCAGACCACAGUCCAUGCCAUCACU..

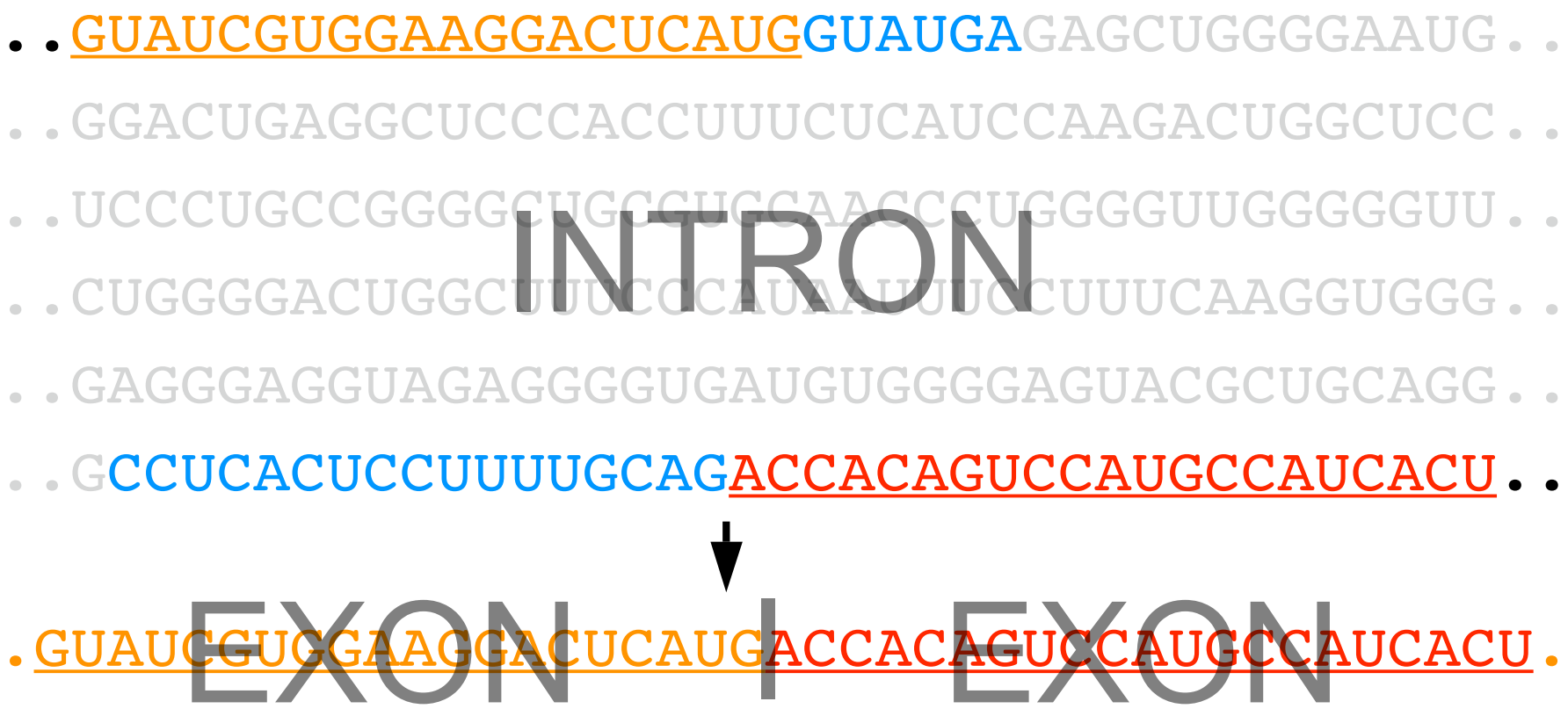
(A|C)(A|U)GGU(A|G)(A|U)G(A|U)

(C|U)(A|C|G|U)(C|U)U(A|C|G)A(C|U){10-20}(A|G)(C|U)AG

Splicing


EXON
..GUAUCGUGGAAGGACUCAUGGUAUGAGAGCUGGGGAAUG..
..GGACUGAGGCUCCCACCUUUCUCAUCCAAGACUGGCUCC..
..UCCCUGCCGGGGCUGCCUGCAACCCUGGGGUUGGGGGUU..
INTRON
..CUGGGGACUGGCUUUCCCAUAAUUUCCUUUCAAGGUGGG..
..GAGGGAGGUAGAGGGGUGAUGUGGGGAGUACGCUGCAGG..
..GCCUCACUCCUUUUGCAGACCACAGUCCAUGCCAUCACU..
EXON

Splicing


..GUAUCGUGGAAGGACUCAUGGUAUGAGAGCUGGGGAAUG..
..GGACUGAGGCUCCCACCUUUCUCAUCCAAGACUGGCUCC..
..UCCCUGCCGGGGCUGCCUGCAACCCUGGGGUUGGGGGUU..
..CUGGGGACUGGCUUUCCCAUAAUUUCCUUUCAAGGUGGG..
..GAGGGAGGUAGAGGGGUGAUGUGGGGAGUACGCUGCAGG..
..GCCUCACUCCUUUUGCAGACCACAGUCCAUGCCAUCACU..
INTRON
.GUAUCGUGGAAGGACUCAUGACCACAGUCCAUGCCAUCACU.
EXON | EXON

Splicing

http://www.shenghua.sdu.edu.cn/gif/mRNA%20splicing.gif

Alternative Splicing

..GUAUCGUGGAAGGACUCAUGGUAUGAGAGCUGGGGAAUG..

..GGACUGAGGCUCCCACCUUUCUCAUCCAAGACUGGCUCC..

..UCCCUGCCGGGGCUGCGUGCAACCCUGGGGUUGGGGGUU..

..CUGGGGACUGGCUUUCCCAUAAUUUCCUUUCAAGGUGGG..

..GAGGGAGGUAGAGGGGUGAUGUGGGGAGUACGCUGCAGG..

..GCCUCACUCCUUUUGCAGACCACAGUCCAUGCCAUCACU..

Alternative Splicing

..GUAUCGUGGAAGGACUCAUGGUAUGAGAGCUGGGGAAUG..

..GGACUGAGGCUCCCACCUUUCUCAUCCAAGACUGGCUCC..

..UCCCUGCCGGGGCUGCGUGCAACCCUGGGGUUGGGGGUU..

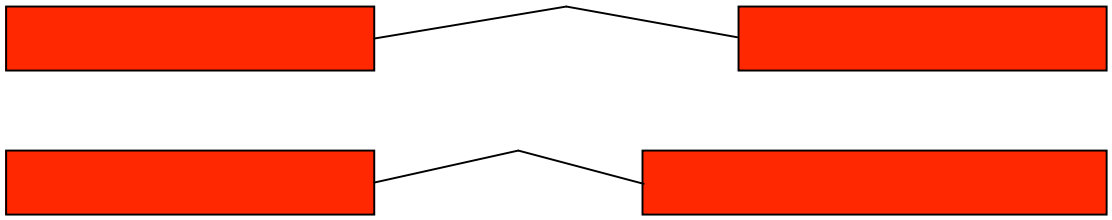
..CUGGGGACUGGCUUUCCCAUAAUUUCCUUUCAAGGUGGG..

..GAGGGAGGUAGAGGGGUGAUGUGGGGAGUACGCUGCAGG..

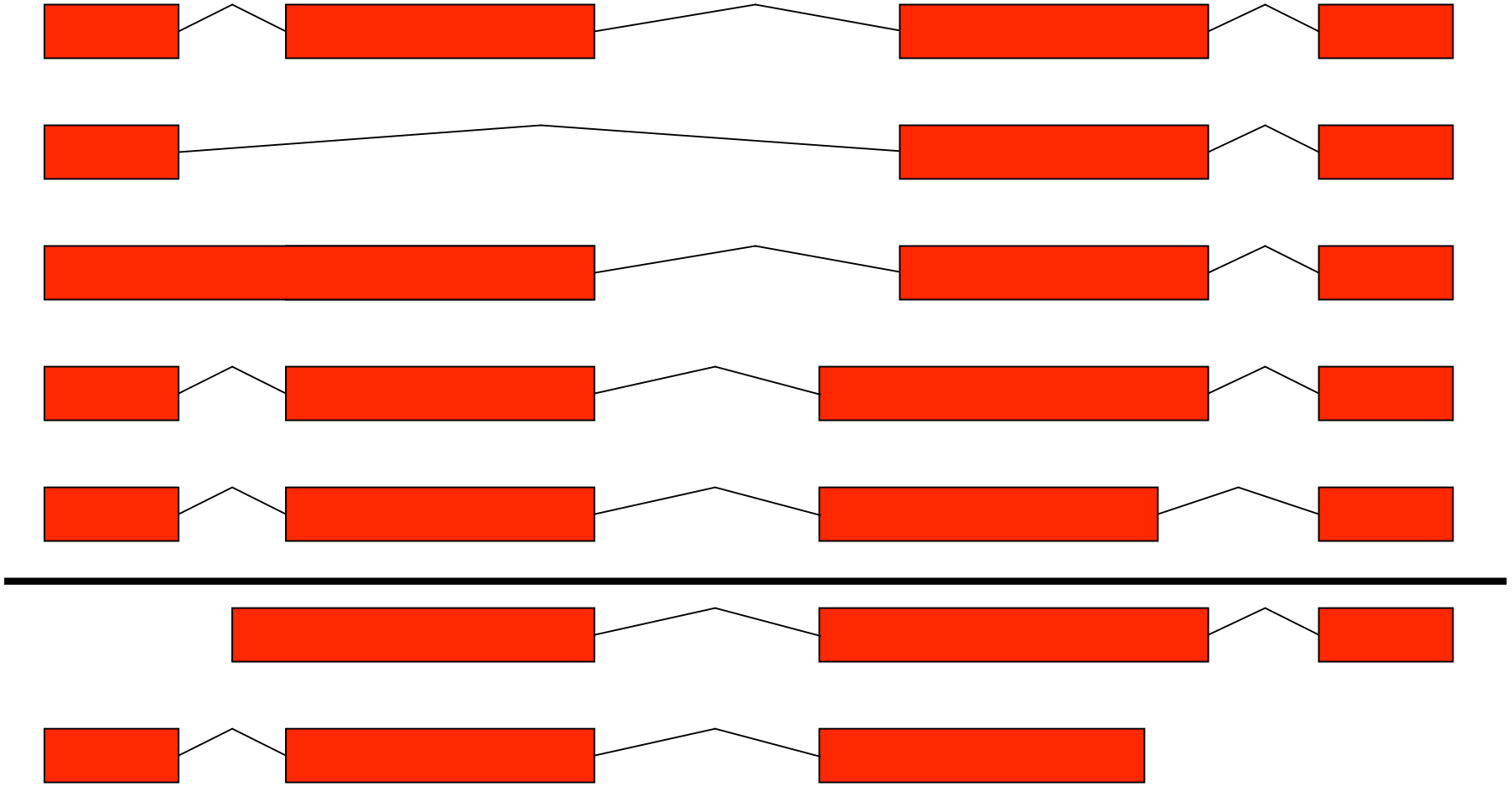
..GCCUCACUCCUUUUGCAGACCACAGUCCAUGCCAUCACU..

Alternative Splicing

Alternative Splicing



Alternative Splicing



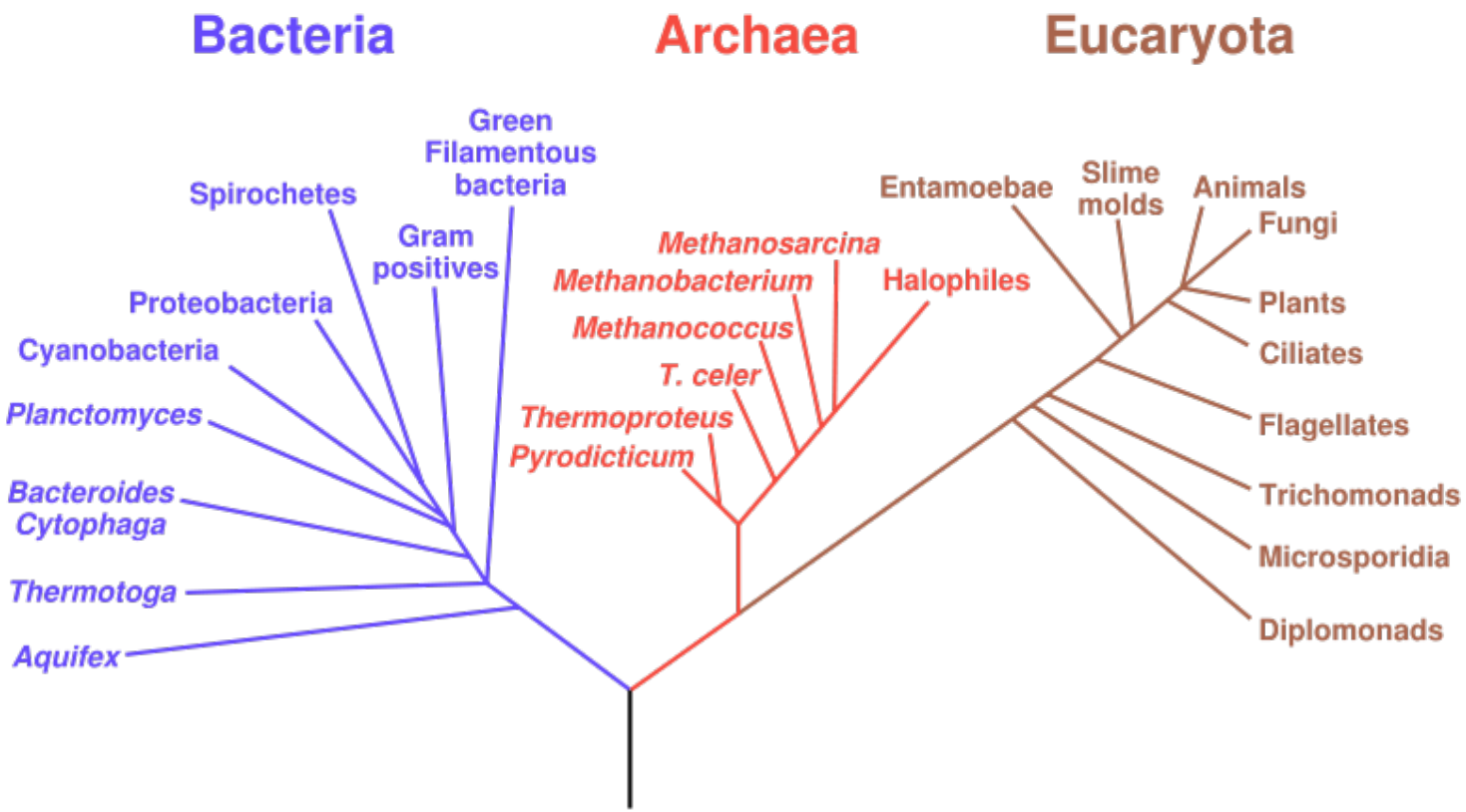
Overview

- ✔ Alternative splicing
- Its biological function
- Studying splicing
  - Technology
  - Computational tasks
- Tissue-specific alternative splicing

Origins of Splicing

- **Prokaryotes** - splicing occurs but there is no general machinery for performing it
- **Eukaryotes** - have got spliceosome - a protein complex specialised in carrying out splicing

Phylogenetic Tree of Life



Biological Function of Alternative Splicing

- **Protein Diversity** - AS enables the cell to produce more than one protein from the same gene (no need for larger genome)
- **Fast Protein Evolution** - new splicing variants emerge by replacing only a few nucleotides (DNA letters).

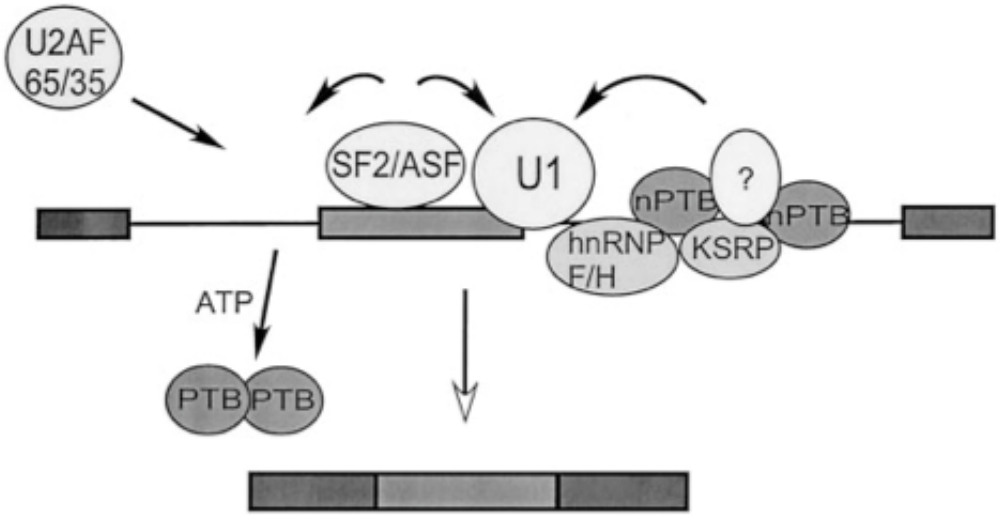
Alternative Splicing in Action

There can be different splice variants in

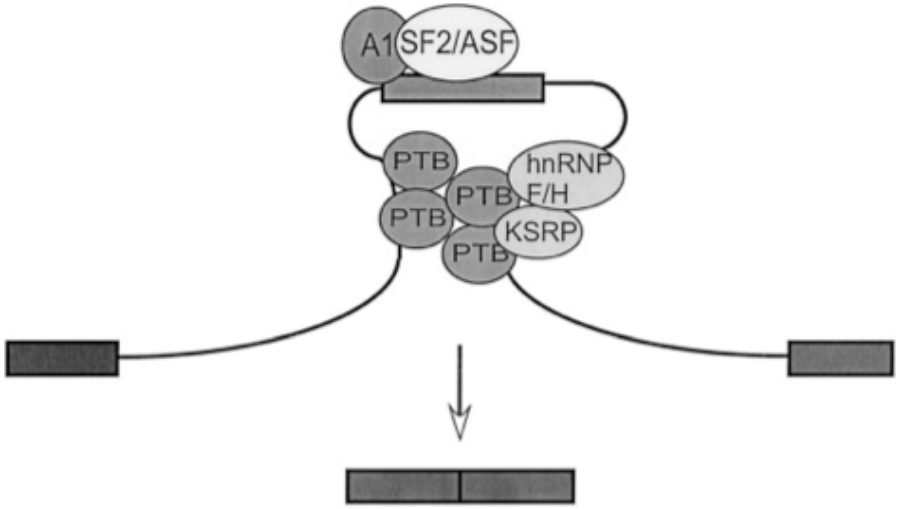
- Embryos and Adults (development)
- Brain and Muscle (anatomy)
- Normal and Cancer (pathology)

Studying alternative splicing helps to decrypt anatomy, development, and pathology

Regulation of Alternative Splicing



Neurons



Non-neuronal cells

Our Task

- Computationally predict cases of anatomy-, pathology-, or development-specific splicing

Overview

- ✔ Alternative splicing
- ✔ Its biological function
- Studying splicing
  - Technology
  - Computational tasks
- Tissue-specific alternative splicing

Technology

- How was splicing performed in a cell?
- Let us find out the final result of splicing
  - Take out molecules of spliced mRNA
  - Find out the sequences of nucleotides
- Compare it with the original mRNA (=DNA)
  - Scan the genome to find out the matching parts
- Matching parts - Exons,
  Regions between matches - Introns

Example
(CS Theory Days, right?)

```
..GUAUCGUGGAAGGACUCAUGGUAUGAGAGCUGGGGAAUG..
..GGACUGAGGCUCCCACCUUUCUCAUCCAAGACUGGCUCC..
..UCCCUGCCGGGGCUGCGUGCAACCCUGGGGUUGGGGGUU..
..CUGGGGACUGGCUUUCCCAUAAUUUCCUUUCAAGGUGGG..
..GAGGGAGGUAGAGGGGUGAUGUGGGGAGUACGCUGCAGG..
..GCCUCACUCCUUUUGCAGACCACAGUCCAUGCCAUCACU..

GGCAGGAUUCAACAGCCCAUUCGAUCCC
```

Example

```
                GGCAGGAUUCA
..GUAUCGUGGAAGGACUCAUGGUAUGAGAGCUGGGGAAUG..
..GGACUGAGGCUCCCACCUUUCUCAUCCAAGACUGGCUCC..
..UCCCUGCCGGGGCUGCGUGCAACCCUGGGGUUGGGGGUU..
..CUGGGGACUGGCUUUCCCAUAAUUUCCUUUCAAGGUGGG..
..GAGGGAGGUAGAGGGGUGAUGUGGGGAGUACGCUGCAGG..
..GCCUCACUCCUUUUGCAGACCACAGUCCAUGCCAUCACU..
                        ACAGCCCAUUCGAUCCC

GGCAGGAUUCA ACAGCCCAUUCGAUCCC - 6 mismatch
```

Example

```
            GGCAGGAUUCA
..GUAUCGUGGAAGGACUCAUGGUAUGAGAGCUGGGGAAUG..
..GGACUGAGGCUCCCACCUUUCUCAUCCAAGACUGGCUCC..
..UCCCUGCCGGGGCUGCGUGCAACCCUGGGGUUGGGGGUU..
..CUGGGGACUGGCUUUCCCAUAAUUUCCUUUCAAGGUGGG..
..GAGGGAGGUAGAGGGGUGAUGUGGGGAGUACGCUGCAGG..
..GCCUCACUCCUUUUGCAGACCACAGUCCAUGCCAUCACU..
                        ACAGCCCAUUCGAUCCC
```

GGCAGGAUUCA ACAGCCCAUUCGAUCCC - 6 mismatch

Example

```
                GGCAGGAUUCAAC
..GUAUCGUGGAAGGACUCAUGGUAUGAGAGCUGGGGAAUG..
..GGACUGAGGCUCCCACCUUUCUCAUCCAAGACUGGCUCC..
..UCCCUGCCGGGGCUGCGUGCAACCCUGGGGUUGGGGGUU..
..CUGGGGACUGGCUUUCCCAUAAUUUCCUUUCAAGGUGGG..
..GAGGGAGGUAGAGGGGUGAUGUGGGGAGUACGCUGCAGG..
..GCCUCACUCCUUUUGCAGACCACAGUCCAUGCCAUCACU..
                          AGCCCAUUCGAUCCC
```

GGCAGGAUUCA ACAGCCCAUUCGAUCCC - 6 mismatch

GGCAGGAUUCAAC AGCCCAUUCGAUCCC - 9 mismatch

Result of Alignment

Chromosome: 12

Gene: ENSG00000179088 (FLJ25323)

Location: 102195188 - 102395213

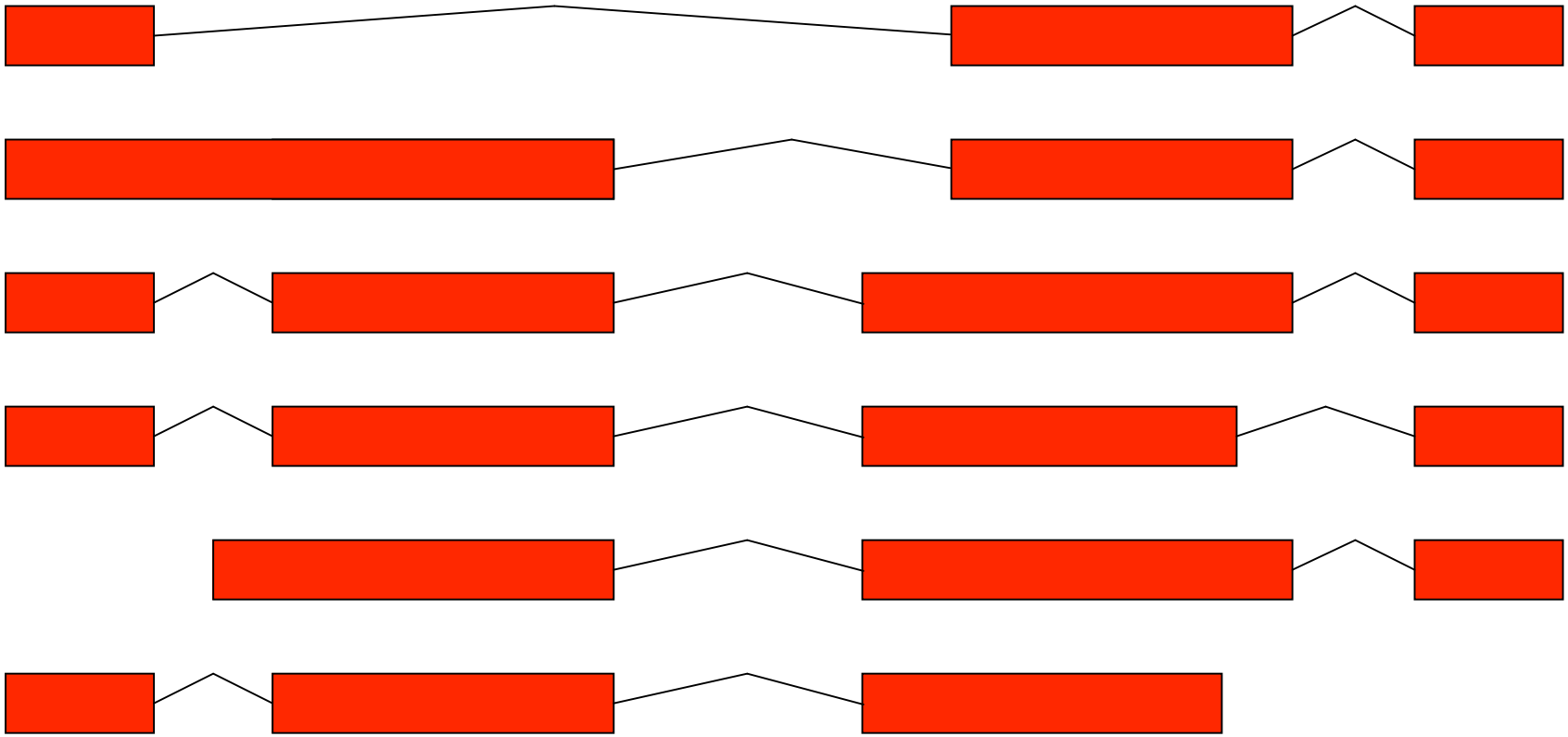
Mapping:

g(3016..3202)	e(1..186)
g(20522..20620)	e(187..285)
g(97294..97362)	e(286..354)
g(129971..130082)	e(355..466)
g(192624..192983)	e(467..830)

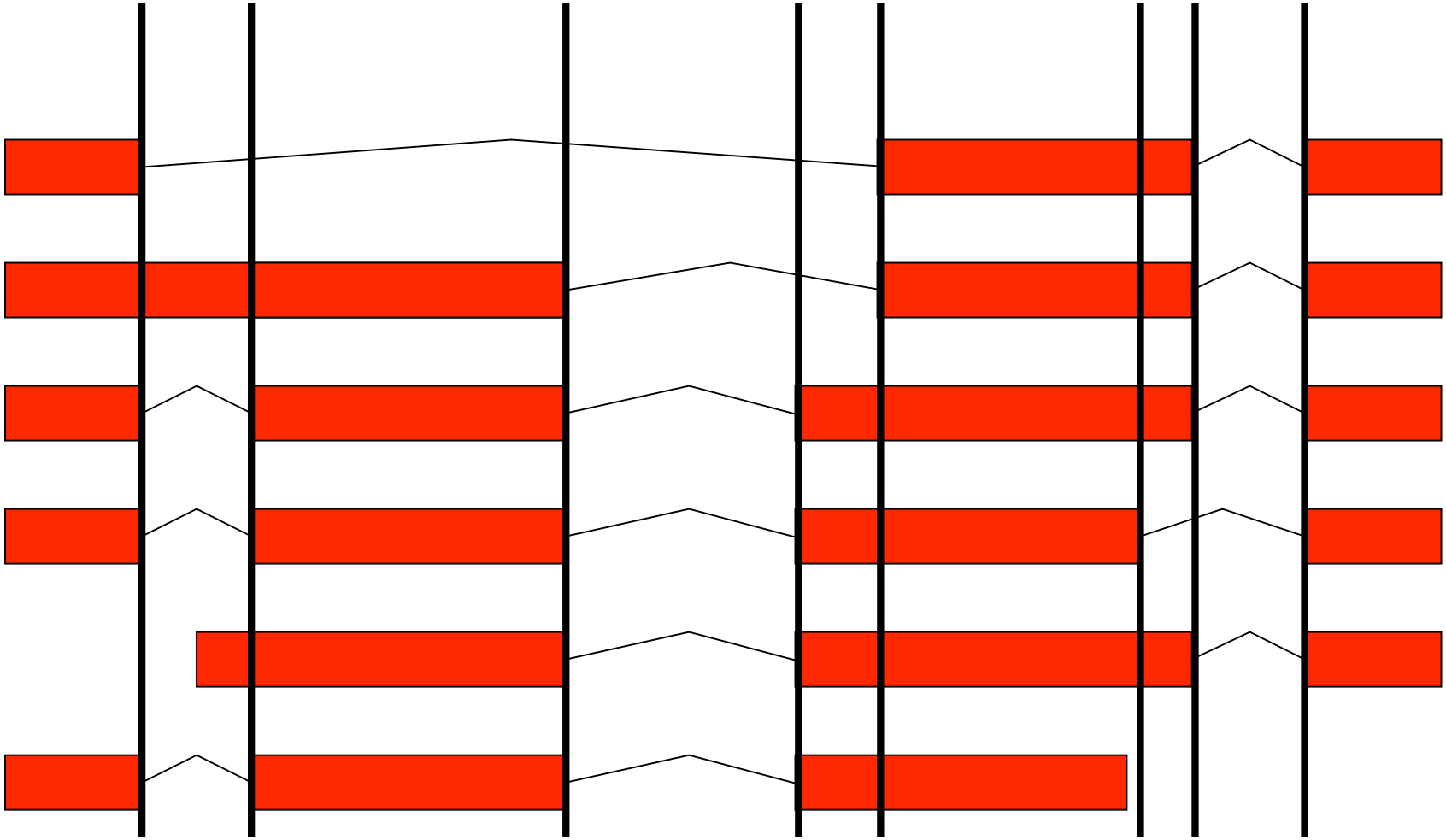
Overview

- ✔ Alternative splicing
- ✔ Its biological function
- ✔ Studying splicing
  - Technology
  - Computational tasks
- Tissue-specific alternative splicing

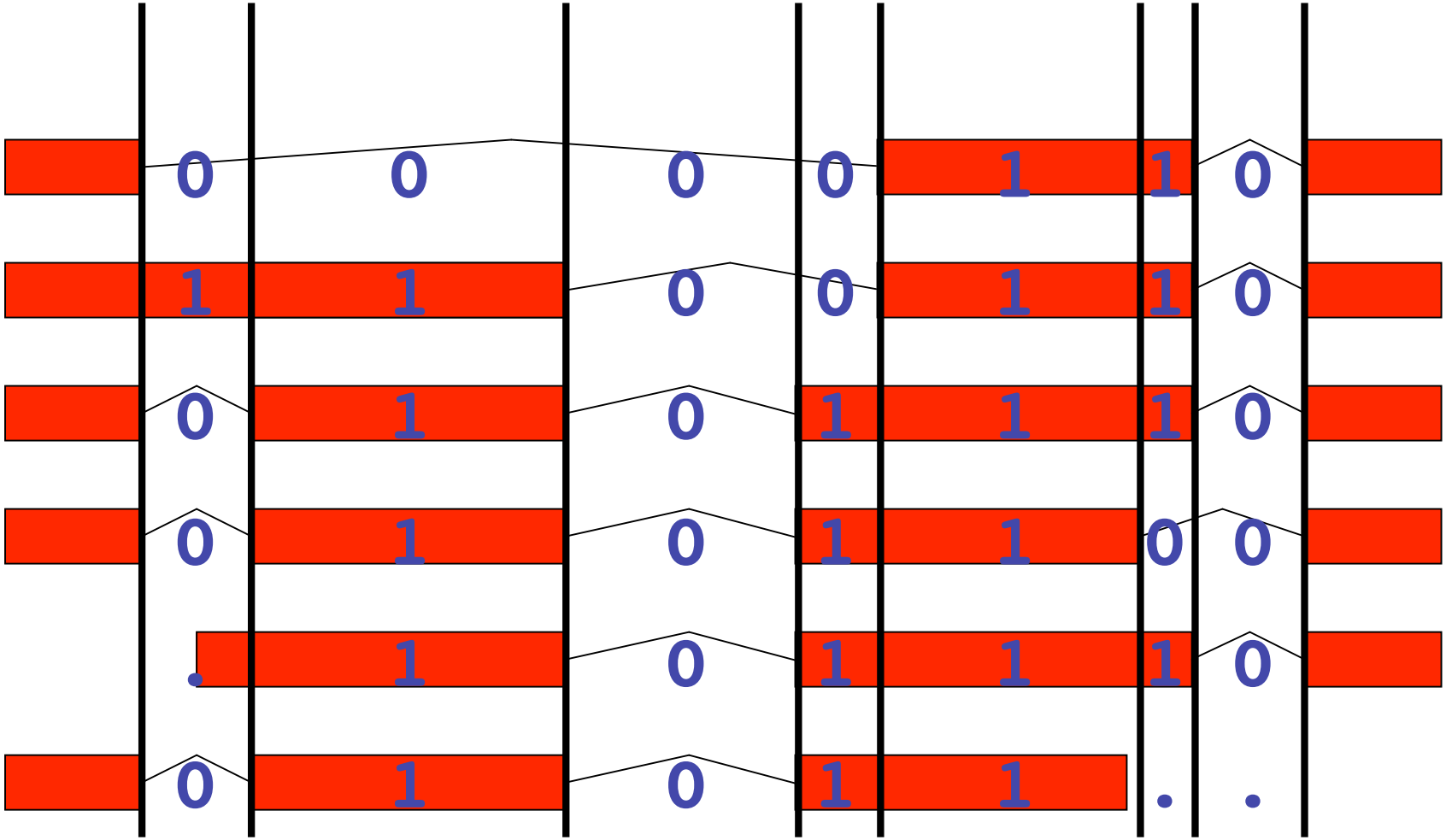
Combine all mRNA produced from the same gene



Find all splice sites



Encode the mRNA alignments



Include mRNA annotation data

Anatomical site: nervous system -> central nervous system -> brain -> brain stem -> medulla oblongata

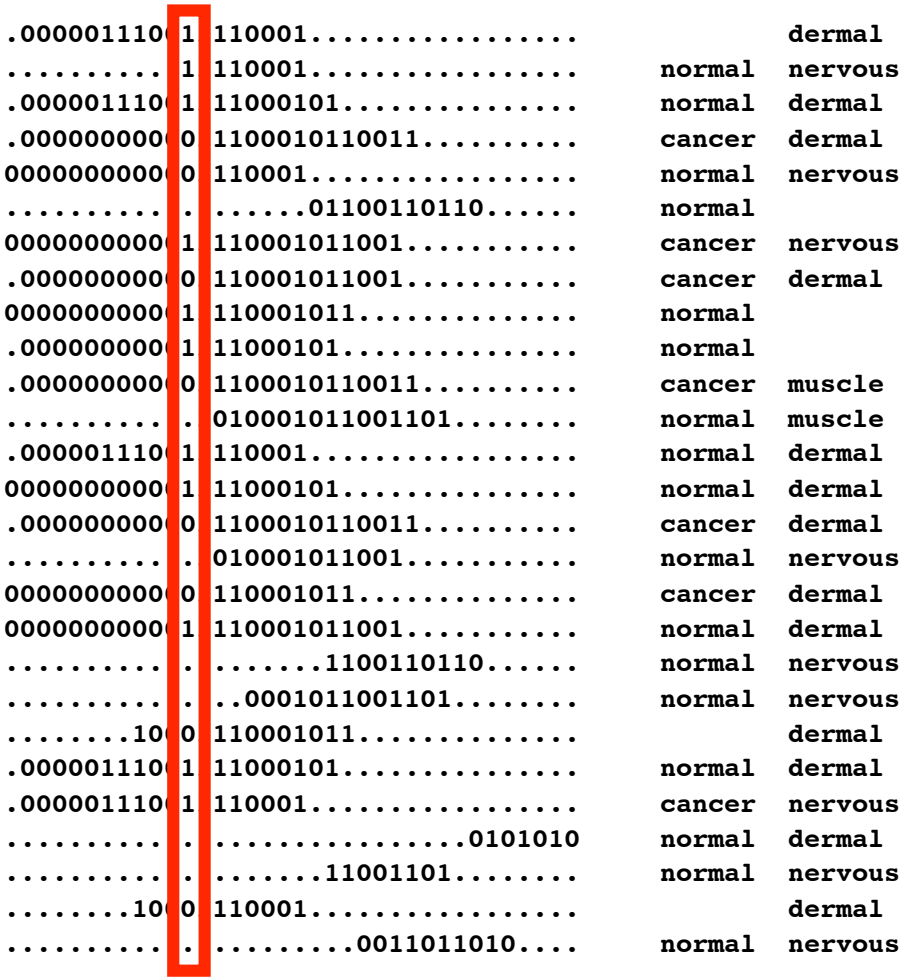
Developmental stage: adult -> 27 years

Pathological state: normal

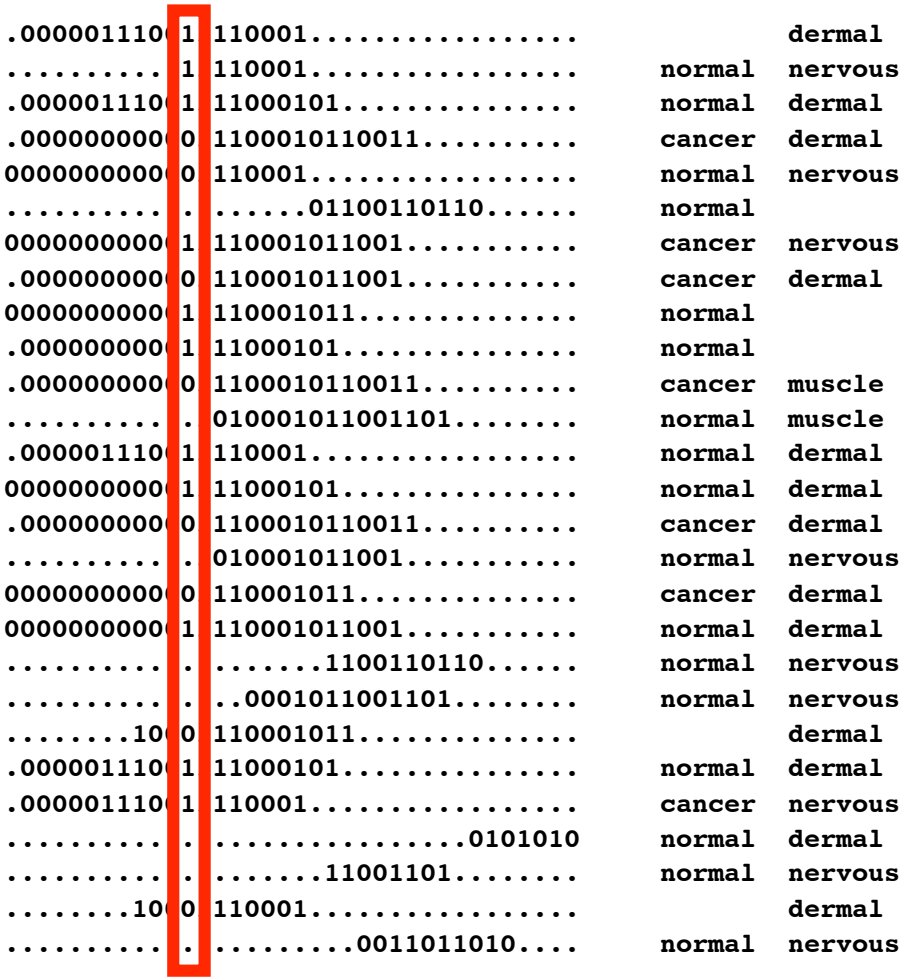
Find alternative splicing

```
.000001110011110001.................          dermal
..........11110001.................    normal  nervous
.0000011100111100010...............    normal  dermal
.00000000000111000010110011.........    cancer  dermal
000000000000111000010..............    normal  nervous
...................01100110110......    normal
0000000000011110001011001...........    cancer  nervous
.000000000001110001011001...........    cancer  dermal
00000000001111000101..............     normal
.0000000001111000101...............    normal
.00000000000111000010110011.........    cancer  muscle
............1010001011001101........    normal  muscle
.000001110011110001.................    normal  dermal
00000000001111000101...............     normal  dermal
.00000000000111000010110011.........    cancer  dermal
............1010001011001...........    normal  nervous
00000000000111000101..............     cancer  dermal
0000000000011110001011001...........    normal  dermal
...................1100110110......    normal  nervous
..............0001011001101........    normal  nervous
........10001110001011..............          dermal
.0000011100111100010...............    normal  dermal
.000001110011110001.................    cancer  nervous
.............................0101010    normal  dermal
...................11001101........    normal  nervous
........10001110001.................          dermal
.....................0011011010....    normal  nervous
```

Find alternative splicing



Find alternative splicing



		dermal
	normal	nervous
	normal	dermal
	cancer	dermal
	normal	nervous
	normal	
	cancer	nervous
	cancer	dermal
	normal	
	normal	
	cancer	muscle
	normal	muscle
	normal	dermal
	normal	dermal
	cancer	dermal
	normal	nervous
	cancer	dermal
	normal	dermal
	normal	nervous
	normal	nervous
		dermal
	normal	dermal
	cancer	nervous
	normal	dermal
	normal	nervous
		dermal
	normal	nervous

	dermal	other
Intron	6	2
Exon	6	3

	cancer	normal
Intron	5	1
Exon	2	8

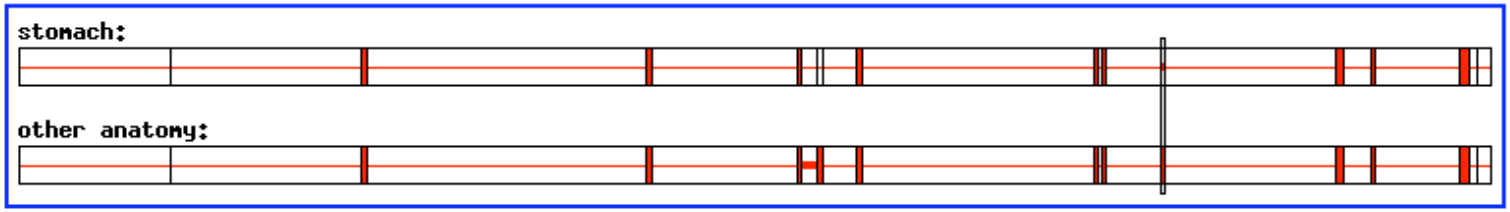
	dermal cancer	dermal normal
Intron	4	0
Exon	0	5

Rank 33

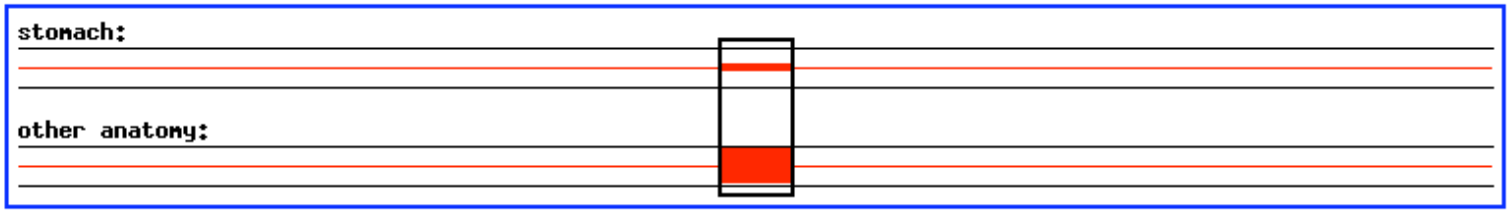
Showing all results
SHOW ONE RESULT FROM EACH GENE
SHOW ALL RESULTS WITH GENE ENSG00000100983

Gene name: **GSS** (GeneCard)
Ensembl ID: **ENSG00000100983** (EnsEMBL, AltSplice)
Region: **23157..23223**

Map:



Zoomed map:



	Code	Description
Developmental:	d	-
Pathological:	p	-
Anatomical:	a_1_8	**Anatomical System--->alimentary system (digestive system)--->stomach**

Number of ESTs:

Number of ESTs	stomach	other anatomy
23157..23223 intronic:	**6** (3-7)	**1** (0-2)
23157..23223 exonic:	**2** (1-3)	**50** (43-51)
All ESTs:	**24133**	**595100**
23157..23223 intronic per million ESTs:	**248** (124-290)	**1** (0-3)
23157..23223 exonic per million ESTs:	**82** (41-124)	**84** (72-85)

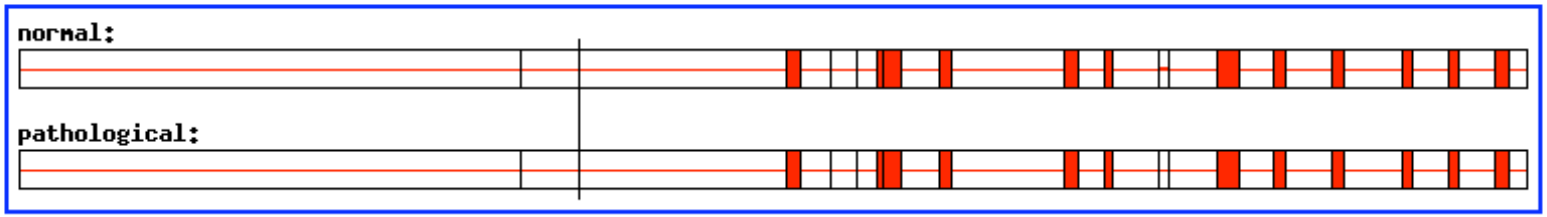
P-value: **4.2093e-06**
Change: **21.5**-fold

Rank 19

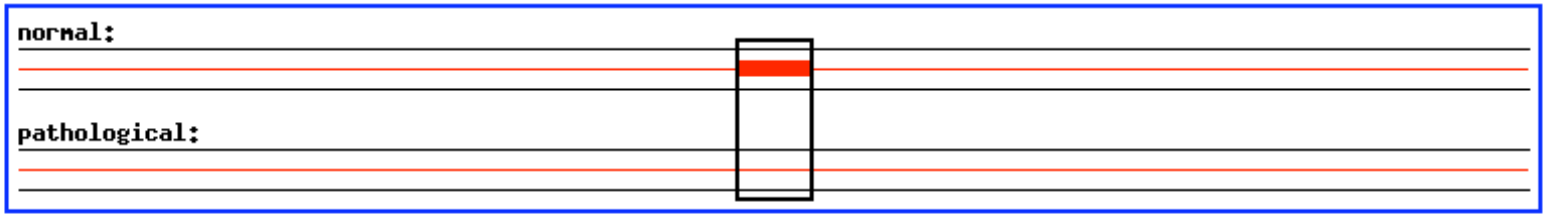
Showing all results
SHOW ONE RESULT FROM EACH GENE
SHOW ALL RESULTS WITH GENE ENSG00000185813

Gene name: **PCYT2** (GeneCard)
Ensembl ID: **ENSG00000185813** (EnsEMBL, AltSplice)
Region: **3514..3518**

Map:



Zoomed map:



	Code	Description
Anatomical:	a	-
Developmental:	d	-
Pathological:	p_6	**Pathology--->normal**

Number of ESTs:

Number of ESTs	normal	pathological
3514..3518 intronic:	**16** (13-17)	**34** (25-35)
3514..3518 exonic:	**14** (6-15)	**0** (0-1)
All ESTs:	**235504**	**326554**
3514..3518 intronic per million ESTs:	**67** (55-72)	**104** (76-107)
3514..3518 exonic per million ESTs:	**59** (25-63)	**0** (0-3)

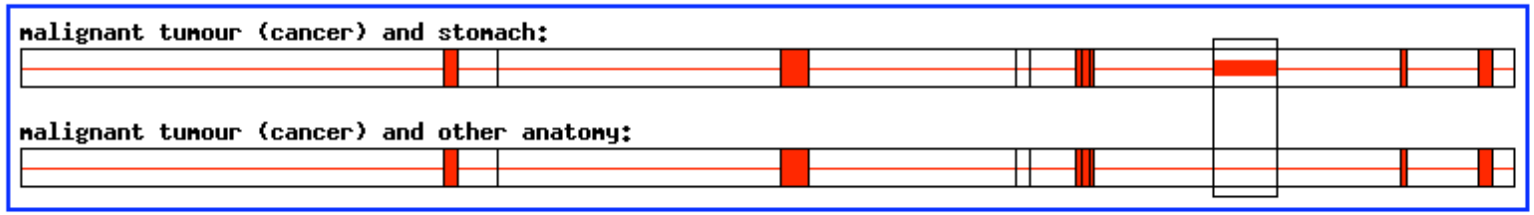
P-value: **3.03877e-06**
Change: **8.31970582240641**-fold

Rank 24

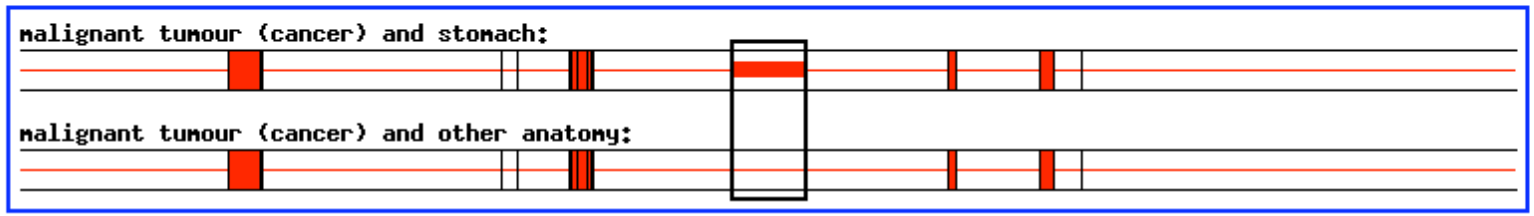
Showing all results
SHOW ONE RESULT FROM EACH GENE
SHOW ALL RESULTS WITH GENE ENSG00000112081

Gene name: **SFRS3** (GeneCard)
Ensembl ID: **ENSG00000112081** (EnsEMBL, AltSplice)
Region: **8439..8894**

Map:



Zoomed map:



	Code	Description
Developmental:	d	-
Pathological:	p_5_1_2	**Pathology--->neoplasia (leukoplakia)--->tumour (tumor,neoplasm)--->malignant tumour (cancer)**
Anatomical:	a_1_8	**Anatomical System--->alimentary system (digestive system)--->stomach**

Number of ESTs:

Number of ESTs	malignant tumour (cancer) and stomach	malignant tumour (cancer) and other anatomy
8439..8894 intronic:	**4** (2-5)	**330** (286-331)
8439..8894 exonic:	**3** (1-4)	**0** (0-1)
All ESTs:	**11085**	**270947**
8439..8894 intronic per million ESTs:	**360** (180-451)	**1217** (1055-1221)
8439..8894 exonic per million ESTs:	**270** (90-360)	**0** (0-3)

P-value: **5.53612e-06**
Change: **24.4426702751466**-fold

Summary

- Alternative splicing is important for protein diversity and evolution
- Studying splicing involves text algorithms
- Our contribution is a method for predicting tissue-specific alternative splicing
- Experiments are yet to be carried out (by others)