

Identifikácia variantov v dátach nanopórového sekvenovania

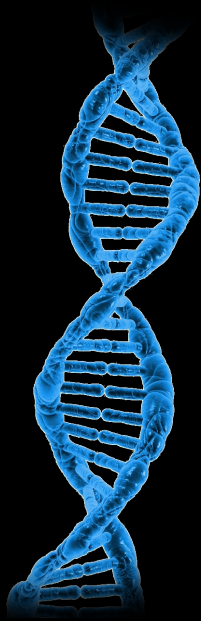
Eduard Batmendijn
Vedúci: doc. Mgr. Tomáš Vinař, PhD.

26. júna 2018

Identifikácia variantov v dátach nanopórového sekvenovania

Identifikácia variantov v dátach nanopórového sekvenovania

Sekvenovanie



ACGTCTAAC

AAGTTACCT

TCAAGTCAT

AGTCTACAT

GCTAGTCAA

Sekvenovanie



ACGTCTAAC

AAGTTACCT

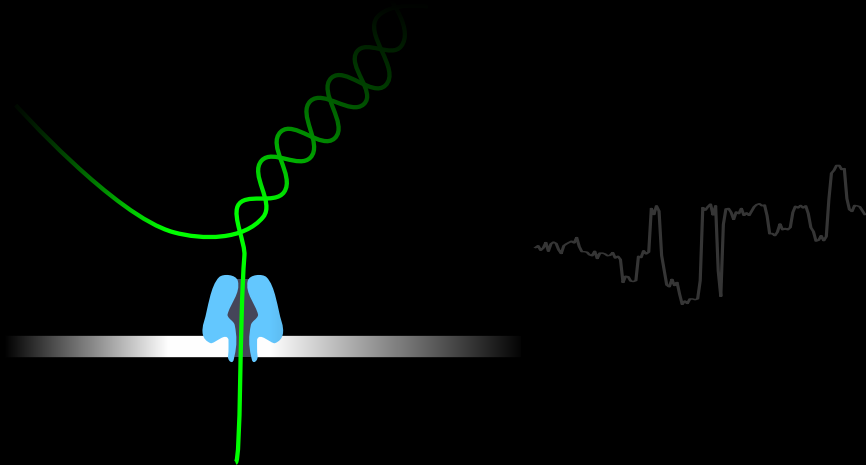
TCAAGTCAT

AGTCTACAT

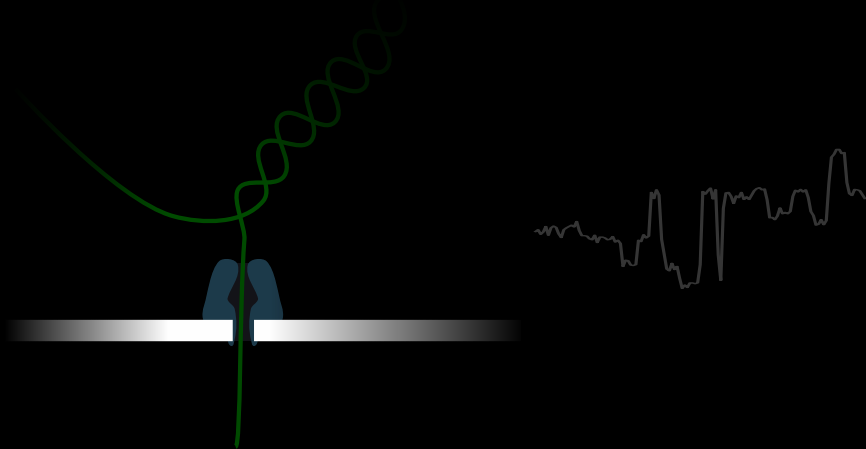
GCTAGTCAA

Identifikácia variantov v dátach nanopórového sekvenovania

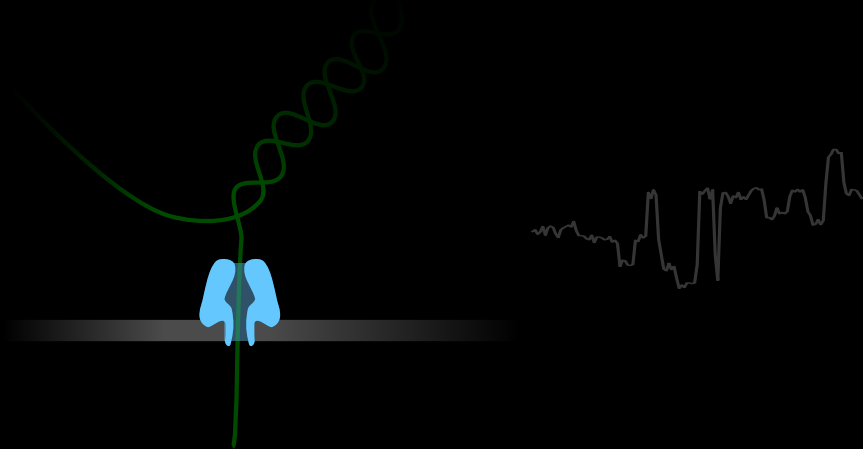
Nanopórové sekvenovanie



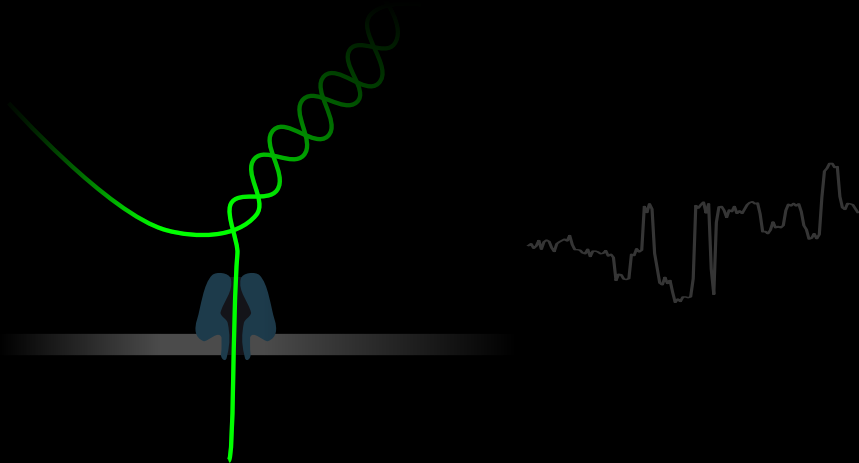
Nanopórové sekvenovanie



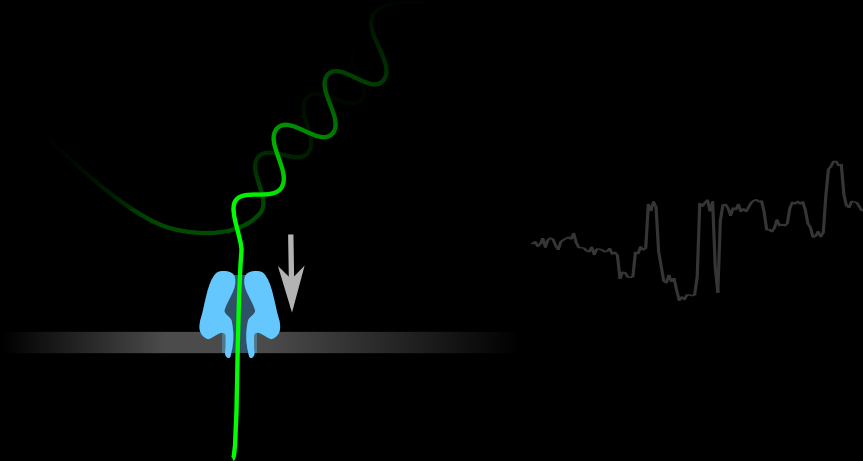
Nanopórové sekvenovanie



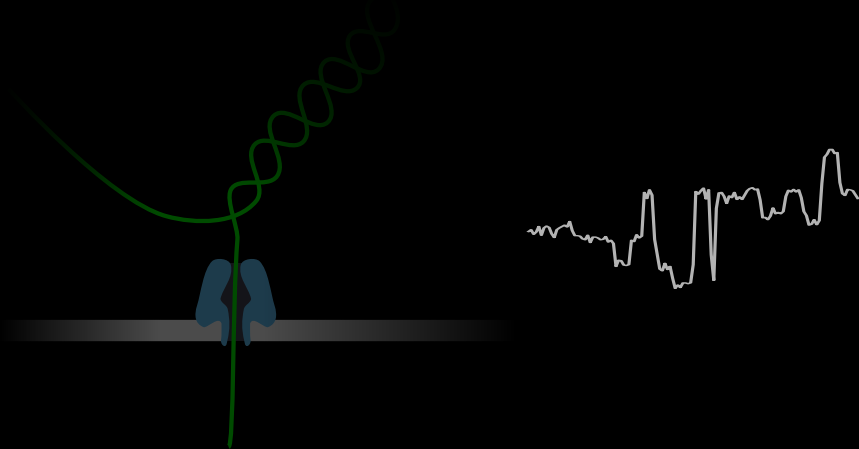
Nanopórové sekvenovanie



Nanopórové sekvenovanie



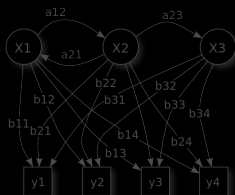
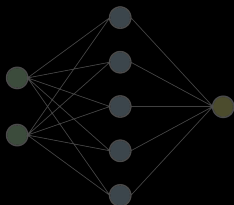
Nanopórové sekvenovanie



Identifikácia variantov v dátach nanopórového sekvenovania

Spracovanie signálu

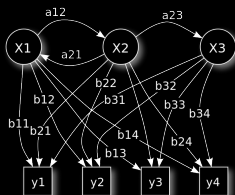
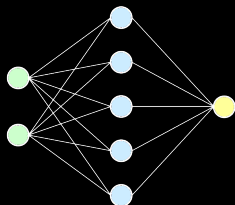
0	521
1	522
2	535
3	515
4	529
5	523
6	529
7	515
8	538
9	510
10	507
11	493
12	484
13	482
14	490
-	-



ACGTCTA
AAGTTAC
TCAAGTC
AGTCTAC
GCTAGTC

Spracovanie signálu

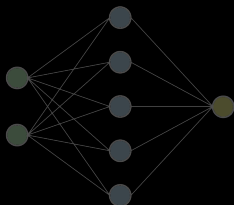
0	521
1	522
2	535
3	515
4	529
5	523
6	529
7	515
8	538
9	510
10	507
11	493
12	484
13	482
14	490
...	...



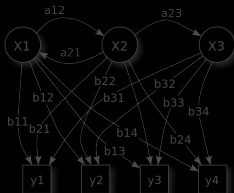
ACGTCTA
AAGTTAC
TCAAGTC
AGTCTAC
GCTAGTC

Spracovanie signálu

0	521
1	522
2	535
3	515
4	529
5	523
6	529
7	515
8	538
9	510
10	507
11	493
12	484
13	482
14	490
...	...



ACGTCTA
AAGTTAC
TCAAGTC
AGTCTAC
GCTAGTC



85% – 90%

Identifikácia variantov v dátach nanopórového sekvenovania

Varianty

ACGGTACCGCAGGACACGACT
ACCTTACCAGGTCACACGGCT

Varianty

ACGGTACCGCAGGACACGACT

ACCTTACCAGGTCACACGGCT

Varianty

ACGGTACCGCAGGACACGACT
ACCTTACCAGGTCACACGGCT

Varianty

ACGGTACCGCAGGACACGACT
ACCTTACCAGGTCACACGGCT

Varianty

ACGGTACCGCAGGACACGACT

ACCTTACCAGGTCACACGGCT

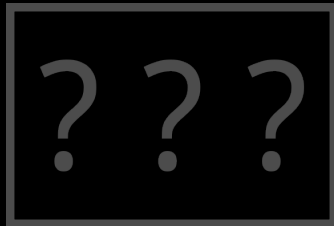
SNP

Identifikácia variantov v dátach nanopórového sekvenovania

Identifikácia variantov

ACAGGTAGTCGCAGATCGACACG

0	521
1	522
2	535
3	515
4	529
5	523
6	529
7	515
8	538
9	510
10	507
11	493
12	484
13	482
14	490
--	--



Identifikácia variantov

ACAGGTAGTCGCAGATCGACACG

0	521
1	522
2	535
3	515
4	529
5	523
6	529
7	515
8	538
9	510
10	507
11	493
12	484
13	482
14	490
--	--



Identifikácia variantov

ACAGGTAGTCGCAGATCGACACG



0	521
1	522
2	535
3	515
4	529
5	523
6	529
7	515
8	538
9	510
10	507
11	493
12	484
13	482
14	490
--	--

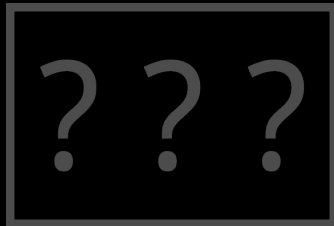


Identifikácia variantov

ACAGG**G**TAGTCGCAGATCGACACG



0	521
1	522
2	535
3	515
4	529
5	523
6	529
7	515
8	538
9	510
10	507
11	493
12	484
13	482
14	490
--	--



Priamočiary prístup

ACGGTACGCAGGACACGACT

ACCTTACCAGGTCACACGGCT

Priamočiary prístup

ACGGTACGCAGGACACGACT

ACCTTACCAGGTCACACGGCT

Priamočiary prístup

ACGGTACGCAGG ACACGACT
ACCTTAC CAGGTCACACGGCT

Priamočiary prístup

ACGGTACGCAGG ACACGACT
ACCTTAC CAGGTCACACGGCT

Priamočiary prístup

ACGGTACGCAGG ACACGACT
ACCTTAC CAGGTCACACGGCT

Priamočiary prístup

ACGGTACGCAGG ACACGACT
ACCTTAC CAGGTCACACGGCT

Priamočiary prístup

ACGGTACGCAGG ACACGACT
ACCTTAC CAGGTCACACGGCT

Priamočiary prístup

ACGGTACGCAGG ACACGACT
ACCTTAC CAGGTCACACGGCT

Pravdepodobnostný model

Pravdepodobnostný model

ACGGTACGCAGGACACGACTAACA



Pravdepodobnostný model

ACGGTACGCAGGACACGACTAACA



Pravdepodobnostný model

ACGGTACGCAGG**A**CACGACTAACA



Pravdepodobnostný model

ACGGTACGCAGGACACGACTAACA



Pravdepodobnostný model

ACGGTACGCAGGACACGACTAACA



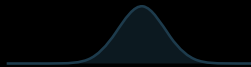
Pravdepodobnostný model

ACGGTACGCAGGACACGACTAACA

The image features a DNA sequence 'ACGGTACGCAGGACACGACTAACA' centered on a black background. The sequence is rendered in a light gray, monospace font. Two bright blue arrows, one pointing left and one pointing right, are positioned above and below the sequence, respectively, with their tips meeting at the center over the 'GACACG' segment. The arrows have a slight shadow effect, giving them a three-dimensional appearance.

Pravdepodobnostný model

AAAAAA



AAAAAC



AAAAAG



AAAAAT



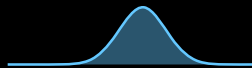
AAAACA



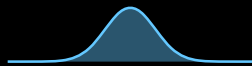
...

Pravdepodobnostný model

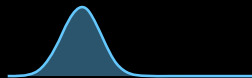
AAAAAA



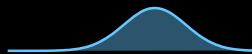
AAAAAC



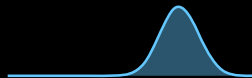
AAAAAG



AAAAAT



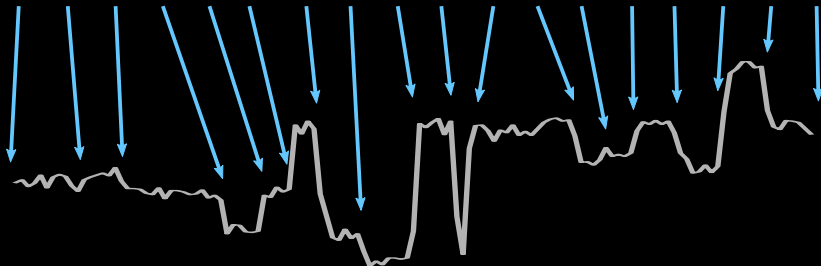
AAAACA



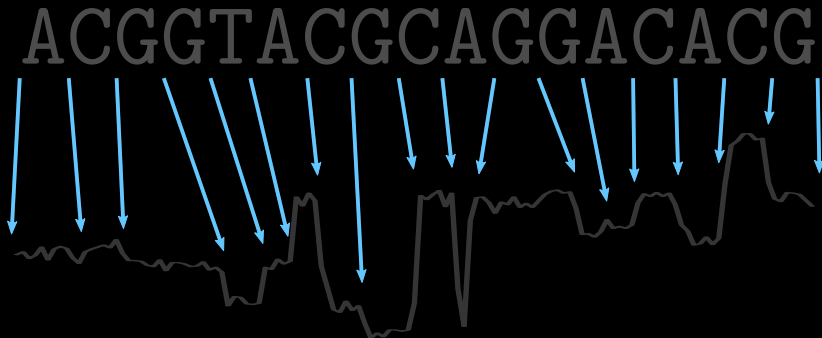
...

Pravdepodobnostný model

ACGGTACGCAGGACACG

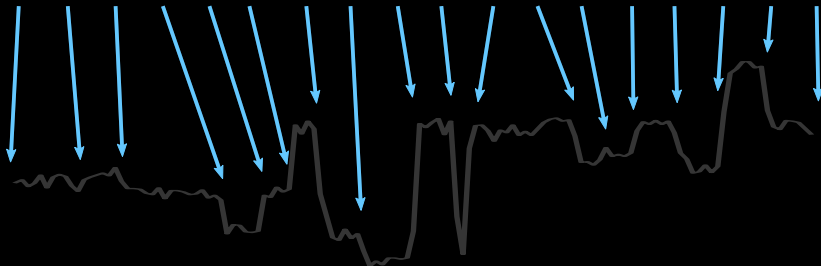


Pravdepodobnostný model



Pravdepodobnostný model

ACGGTACGCAGGACACG



Pravdepodobnostný model

$$P(\text{[waveform]} \mid \text{ACGGTACGCAGGACACG}, \text{[arrows]})$$

Pravdepodobnostný model


$$P(\text{signal} \mid \text{ACGGTACGCAGGACACG})$$

Hypotézy

ACGGTACGCAGGACACG


Hypotézy

ACGGTACGC**A**GGACACG



Hypotézy

ACGGTACGC**C**GGACACG




Hypotézy

ACGGTACGC**G**GGACACG



Hypotézy

ACGGTACGCTGGACACG



Hypotézy


$$P(\text{ACGGTACGCAGGACACG} \mid \text{[Spectral Plot]})$$

Hypotézy

ACGGTACGCAGGACACG


Hypotézy

ACGGTACGC**A**GGACACG




Hypotézy

ATGGGTACGC**A**GGACACG



Hypotézy

ACGGTACTGC**A**GGACACG

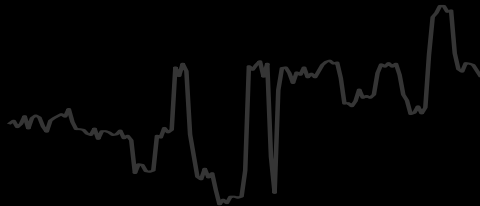


Testovanie

Dáta

ACGGTACGCAGGACACGACT

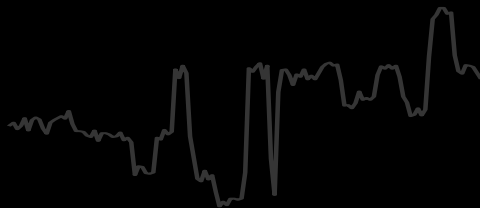
ACCGCACGCAGGACTCGACT



Dáta

ACGGTACGCAGGACACGACT

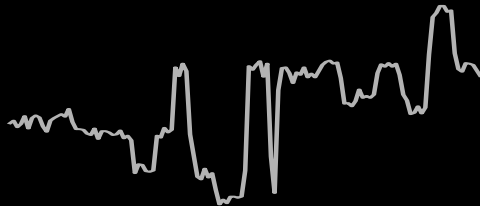
ACCGCACGCAGGACTCGACT



Dáta

ACGGTACGCAGGACACGACT

ACCGCACGCAGGACTCGACT

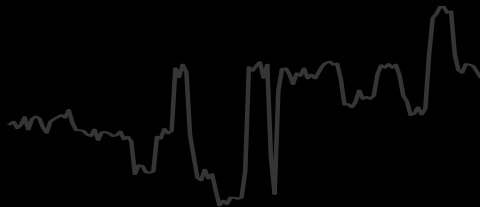


Dáta

ACGGTACGCAGGACACGACT



ACCGCACGCAGGACTCGACT



Úspešnosť identifikácie

1741: 97.87 %

39: 95.17 %

1222: 93.15 %

900: 89.91 %

529: 82.69 %

684: 69.56 %

722: 63.47 %

709: 59.70 %

20: 48.30 %

674: 42.83 %

1343: 31.58 %

1344: 27.68 %

886: 25.81 %

856: 22.24 %

1561: 22.22 %

676: 19.43 %

Úspešnosť identifikácie

1741: 97.87 %

39: 95.17 %

1222: 93.15 %

900: 89.91 %

529: 82.69 %

684: 69.56 %

722: 63.47 %

709: 59.70 %

20: 48.30 %

674: 42.83 %

1343: 31.58 %

1344: 27.68 %

886: 25.81 %

856: 22.24 %

1561: 22.22 %

676: 19.43 %

Úspešnosť identifikácie

1741: 97.87 %

39: 95.17 %

1222: 93.15 %

900: 89.91 %

529: 82.69 %

684: 69.56 %

722: 63.47 %

709: 59.70 %

20: 48.30 %

674: 42.83 %

1343: 31.58 %

1344: 27.68 %

886: 25.81 %

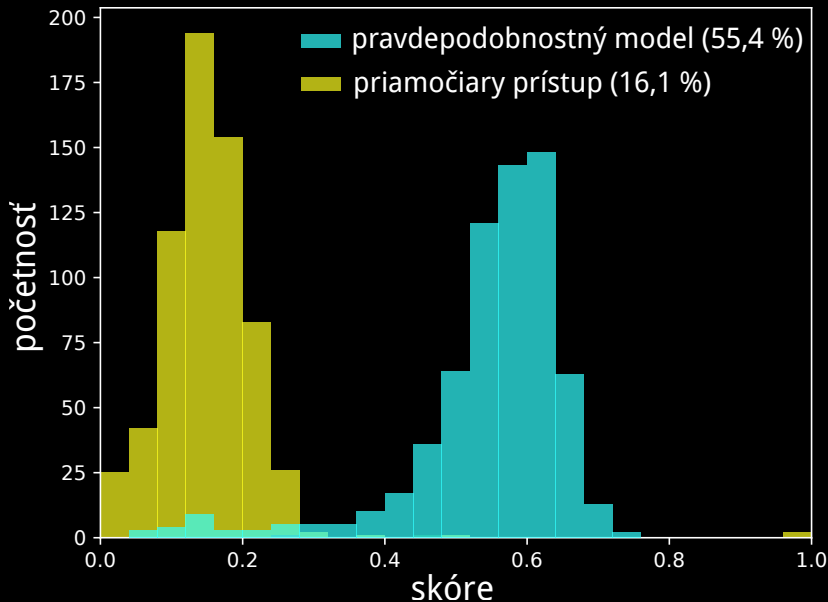
856: 22.24 %

1561: 22.22 %

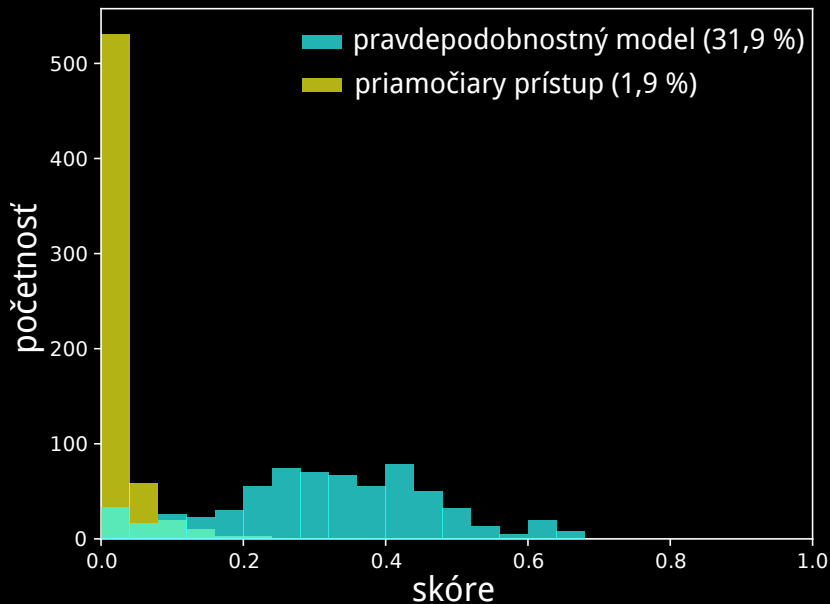
676: 19.43 %

Výsledky

1% SNP



0,1% SNP



Budúcnosť

- ▶ Inzercie, delécie
- ▶ Viac čítaní
- ▶ Iný prístup

Budúcnosť

- ▶ Inzercie, delécie
- ▶ **Viac čítaní**
- ▶ Iný prístup

Budúcnosť

- ▶ Inzercie, delécie
- ▶ Viac čítaní
- ▶ Iný prístup

Použité obrázky

- ▶ https://commons.wikimedia.org/wiki/File:Neural_network.svg
- ▶ <https://commons.wikimedia.org/wiki/File:HiddenMarkovModel.svg>