

# Detekcia tandemových opakovaní v nanopórových dátach

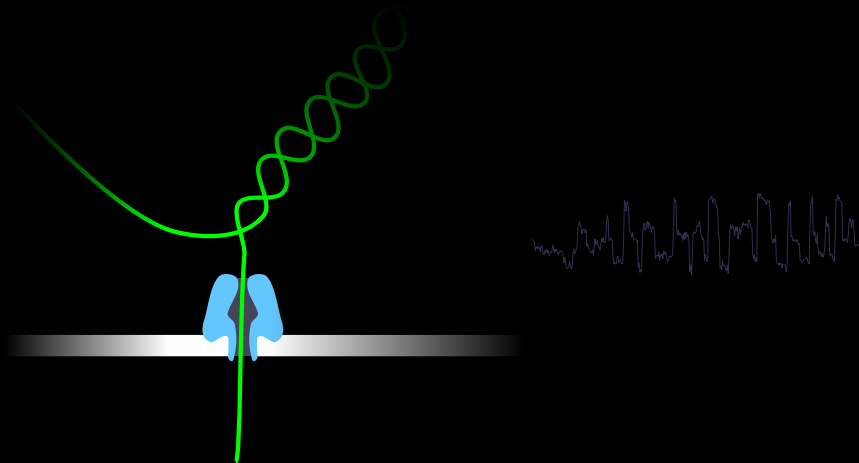
Eduard Batmendijn  
Školiteľ: doc. Mgr. Tomáš Vinař, PhD.

21. februára 2020

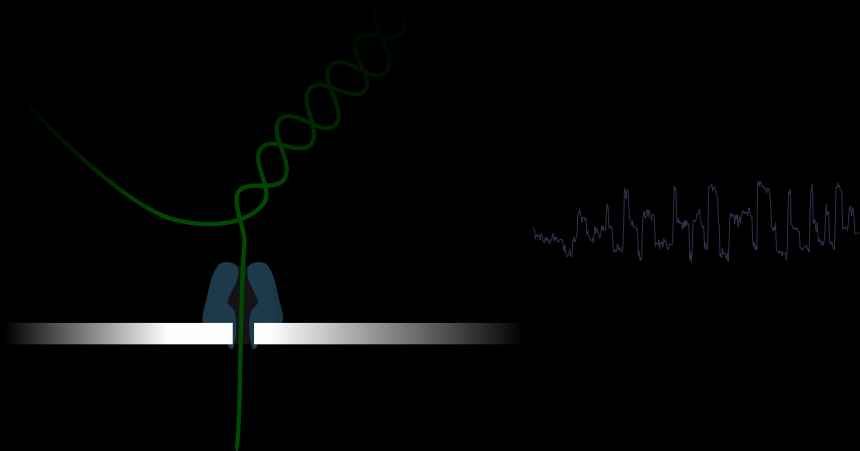
# Detekcia tandemových opakovaní v nanopórových dátach

Detekcia tandemových  
opakovaní v nanopórových  
dátach

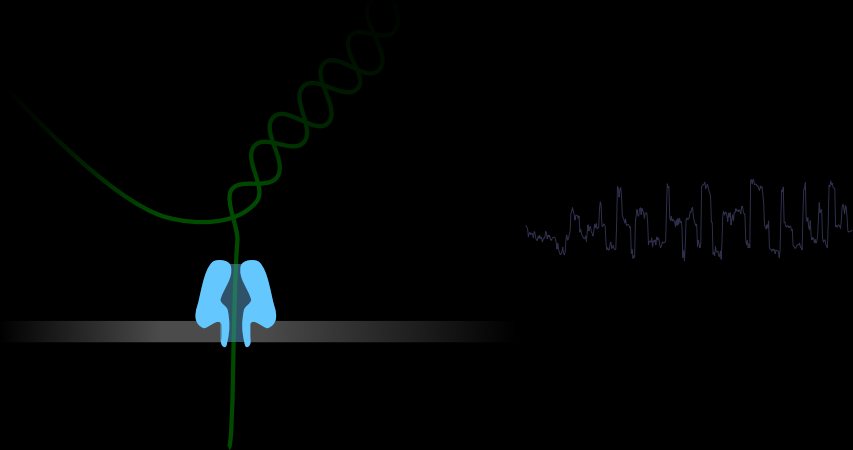
# Nanopórové sekvenovanie



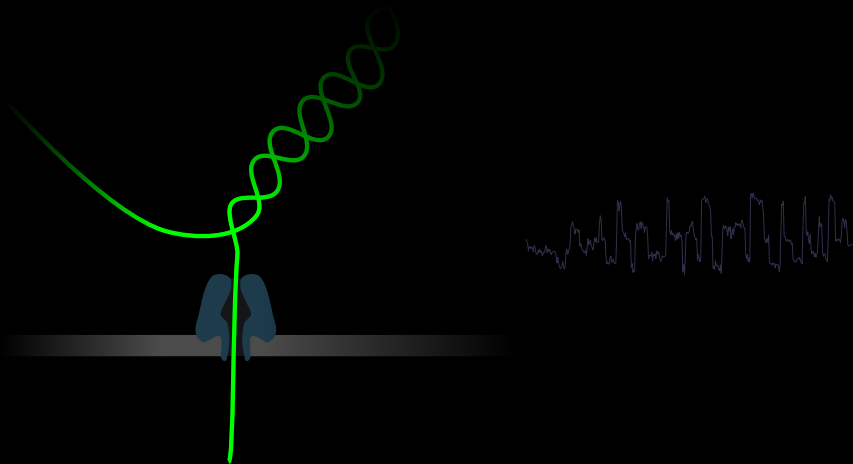
# Nanopórové sekvenovanie



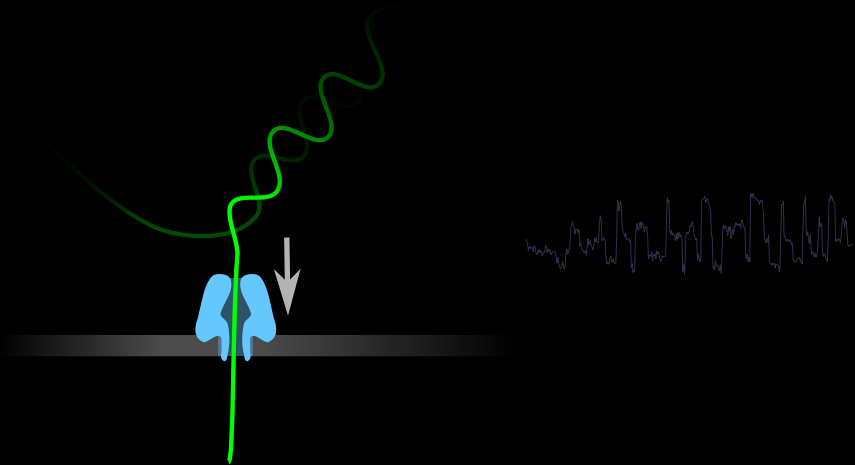
# Nanopórové sekvenovanie



# Nanopórové sekvenovanie

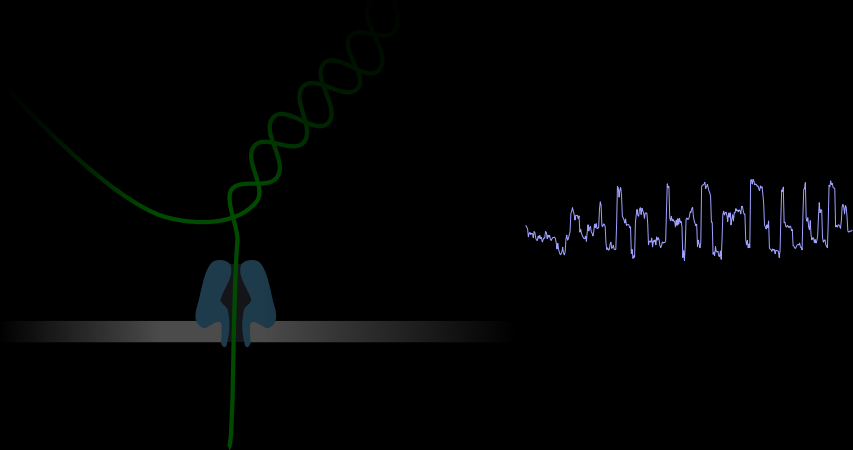


# Nanopórové sekvenovanie

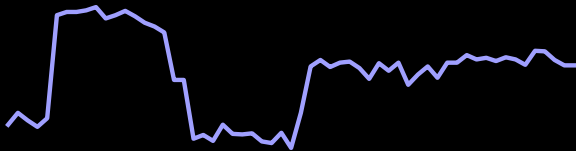
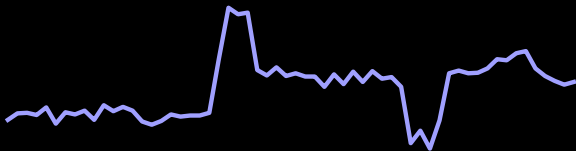




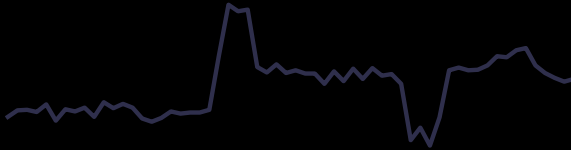
# Nanopórové sekvenovanie



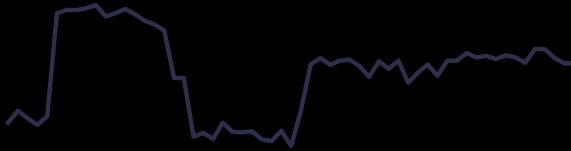
# Nerovnomernosť signálu



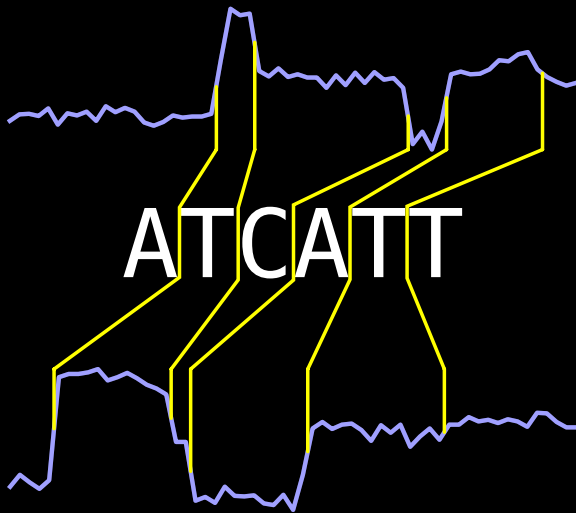
Nerovnomernosť signálu



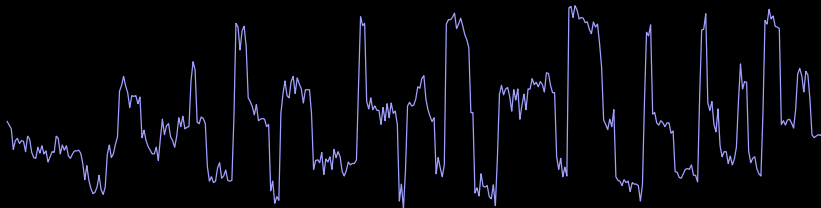
**ATCATT**



Nerovnomernosť signálu

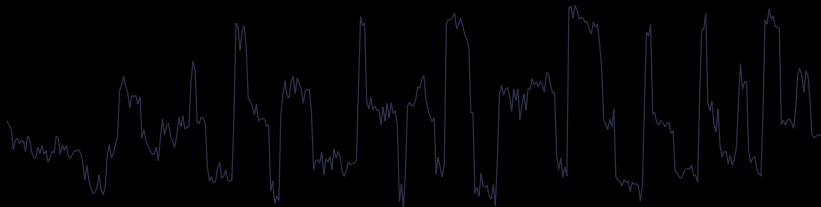


# Určovanie báz (basecalling)



AAAAAAAAACTTTCATCATTATCA  
TTATCATTATCATCATATCTA

# Určovanie báz (basecalling)



AAAAAAAAACTTTCATCATTATCA  
TTATCATTATCATCATATCTA

85% – 90%

# Detekcia tandemových opakovaní v nanopórových dátach



# Detekcia tandemových opakovaní v nanopórových dátach

Tandemové opakovanie

TTCATCATTATCATTATCATTATCATCATC

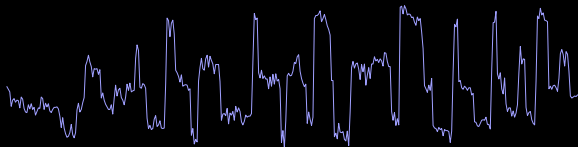
# Tandemové opakovanie

TTCATCATTATCATTATCATTATCATCATC

# Tandemové opakovanie

TTCATCATTATCATTATCATTATCATCATC

# Naivná detekcia

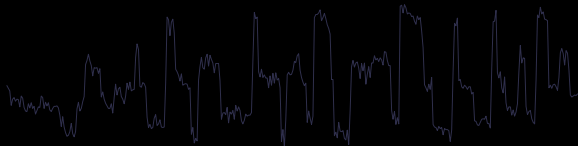


AAAAAAACACTTTCATCATTATCA  
TTATCATTATCATCATATCTA



AAAAAAACACTTTCATCATTATCA  
TTATCATTATCATCATATCTA

# Naivná detekcia

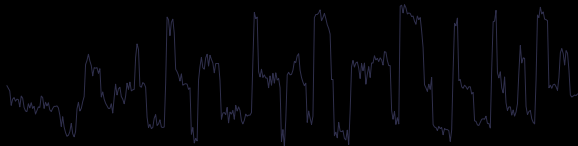


AAAAAAACACTTTCATCATTATCA  
TTATCATTATCATCATATCTA



AAAAAAACACTTTCATCATTATCA  
TTATCATTATCATCATATCTA

# Naivná detekcia



AAAAAAACACTTTCATTATCA  
TTATCATTATCATCATATCTA



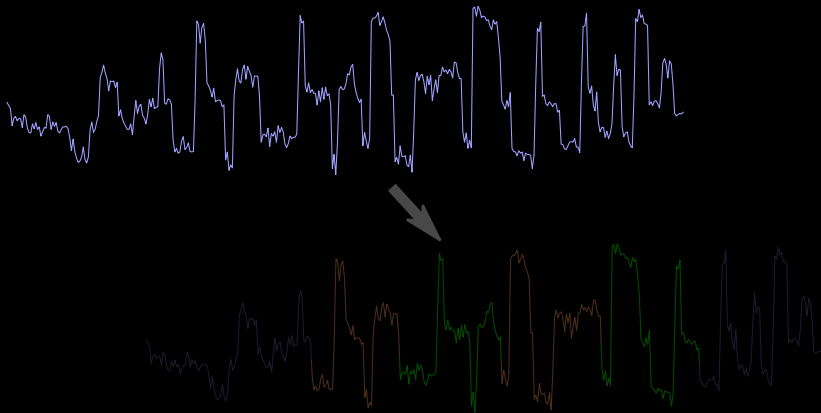
AAAAAAACACTTTCATCATTATCA  
TTATCATTATCATCATATCTA

6%

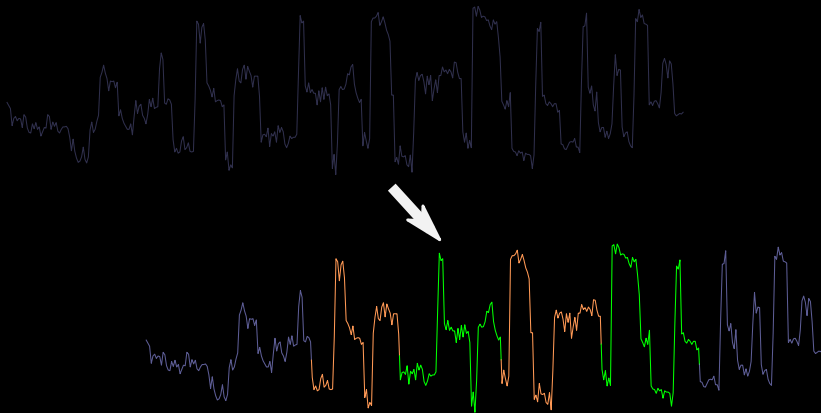


Riešenie

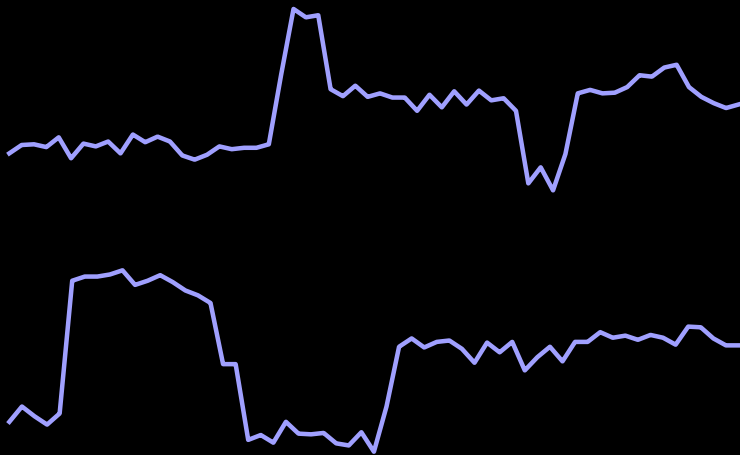
# Detekcia opakovaní v signáli



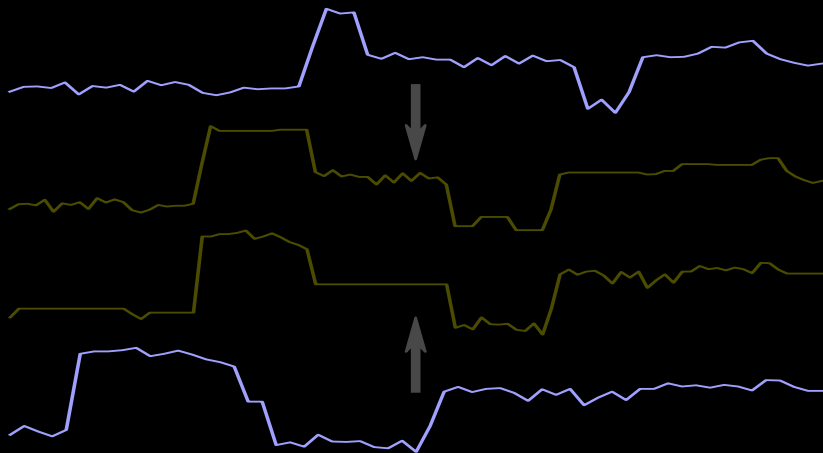
# Detekcia opakovaní v signáli



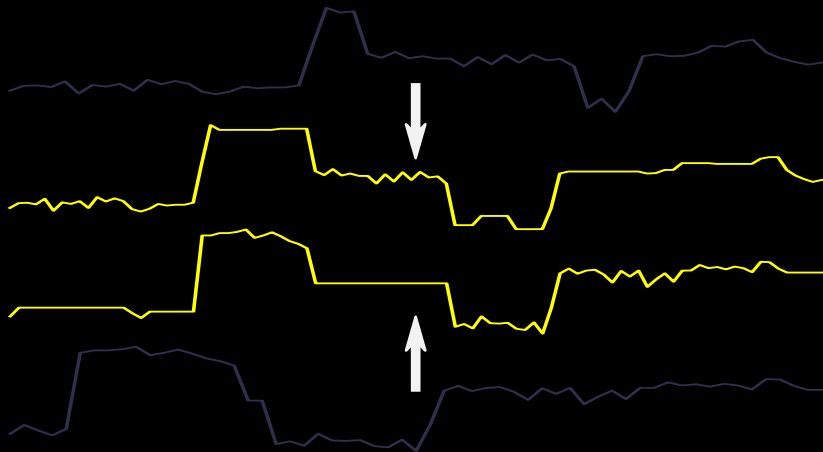
# Nerovnomernosť signálu



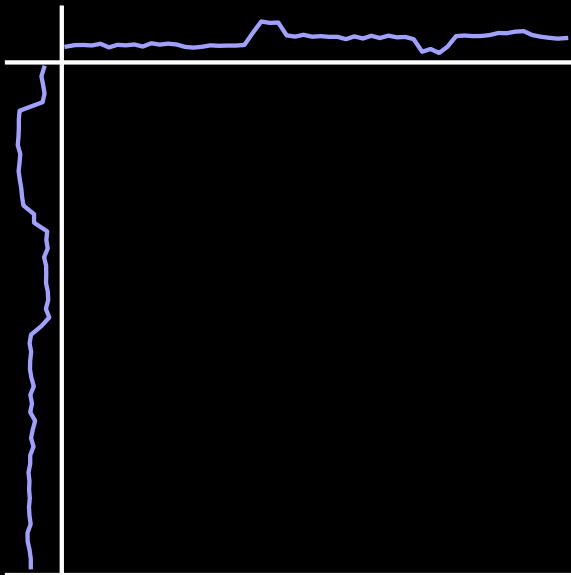
# Dynamic time warping



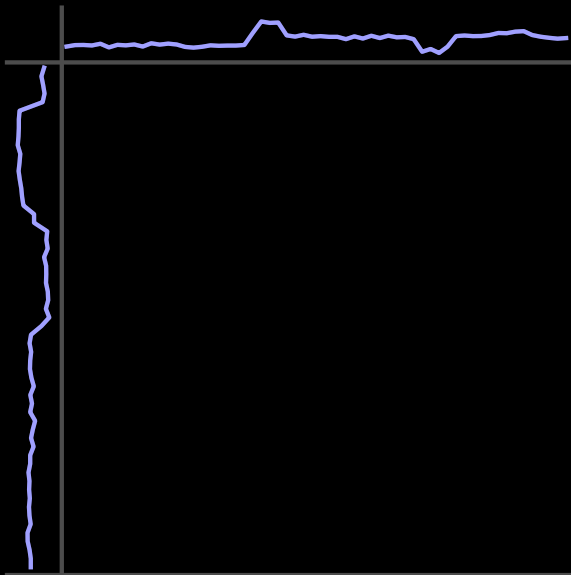
# Dynamic time warping



# Dynamic time warping

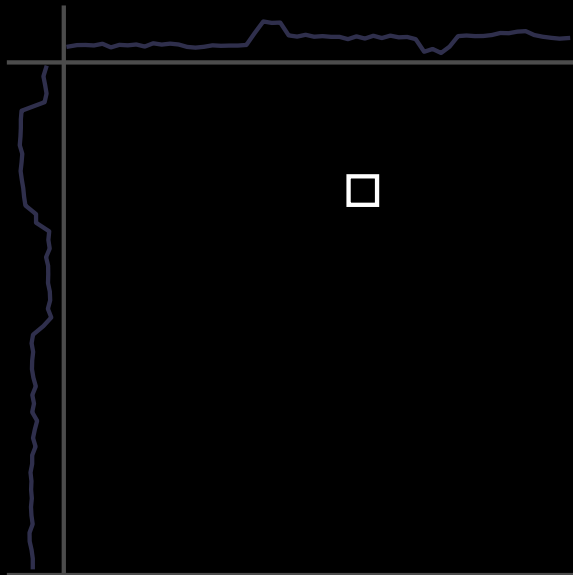


# Dynamic time warping

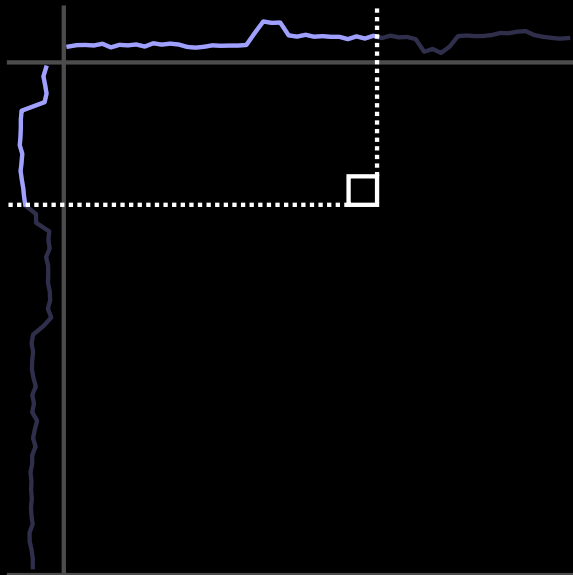




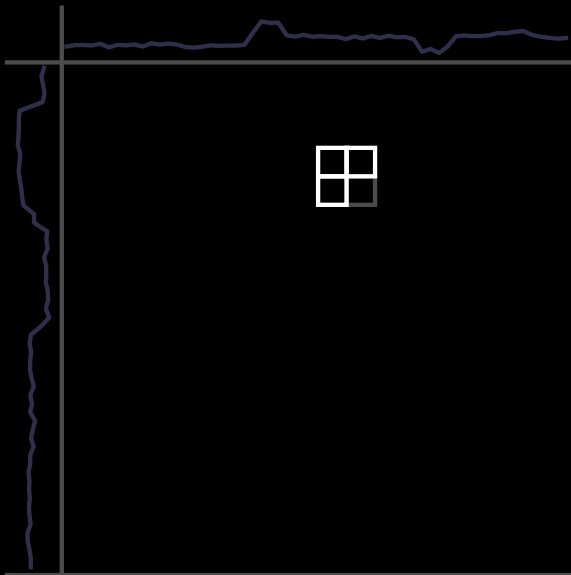
# Dynamic time warping



# Dynamic time warping



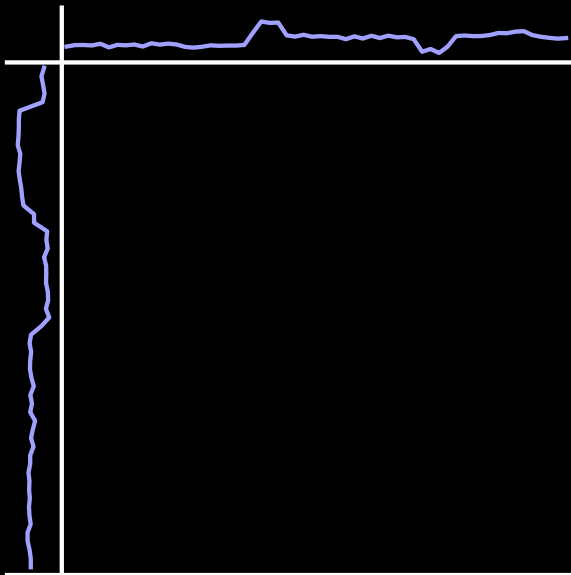
# Dynamic time warping



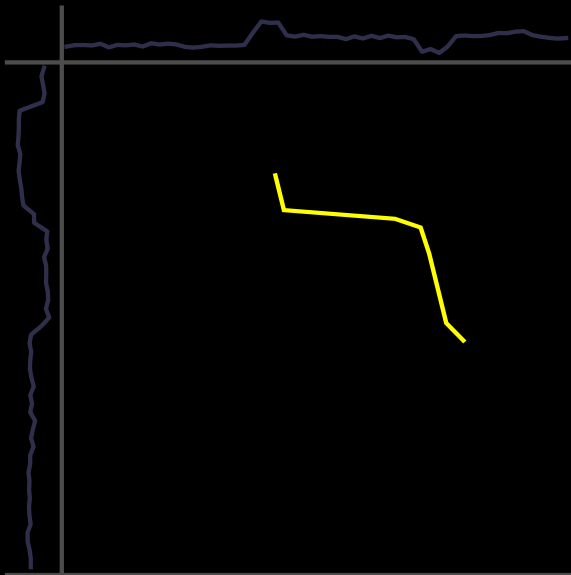
# Dynamic time warping



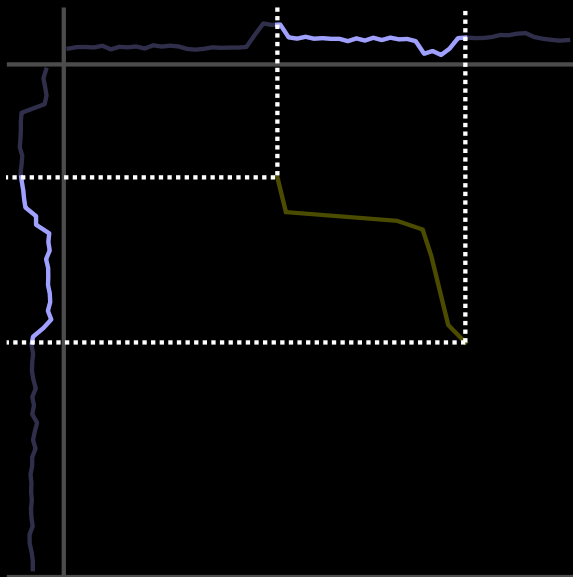
# Lokálne zarovnanie



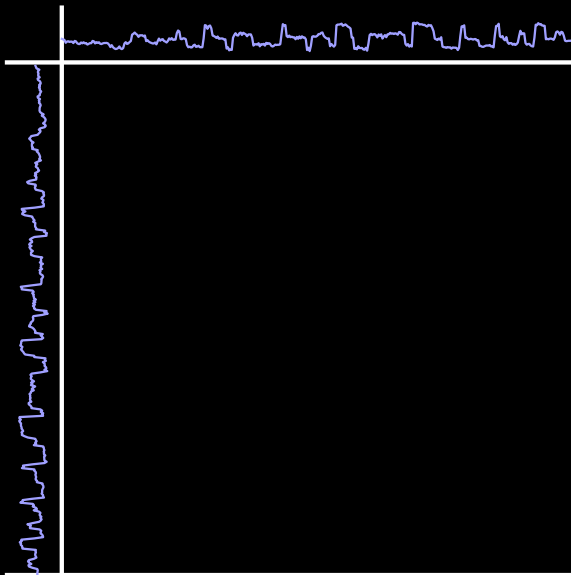
# Lokálne zarovnanie



# Lokálne zarovnanie

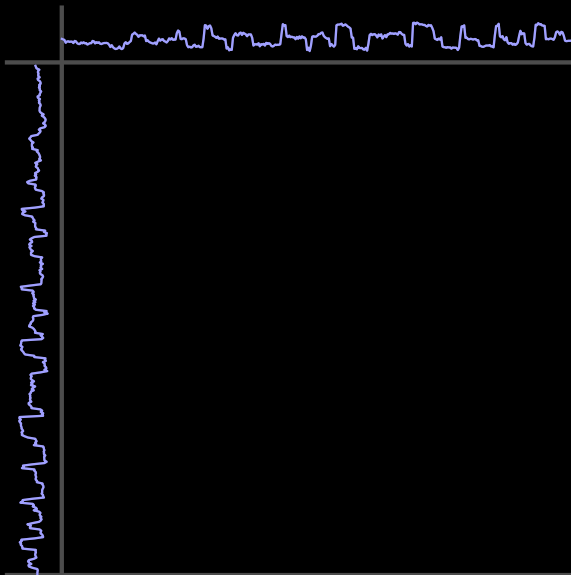


# Samozarovnanie

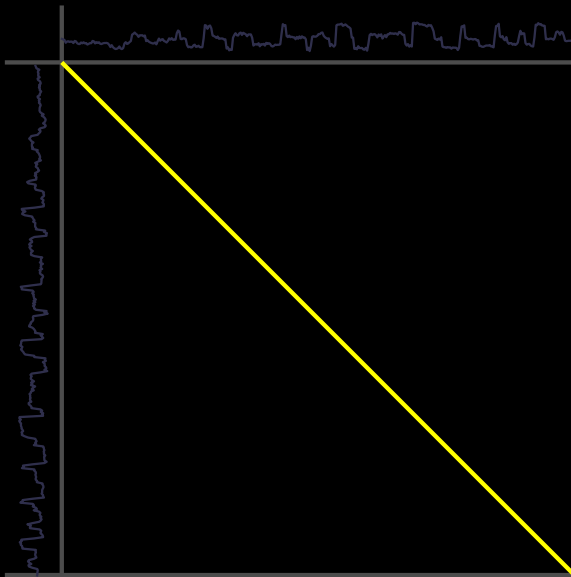




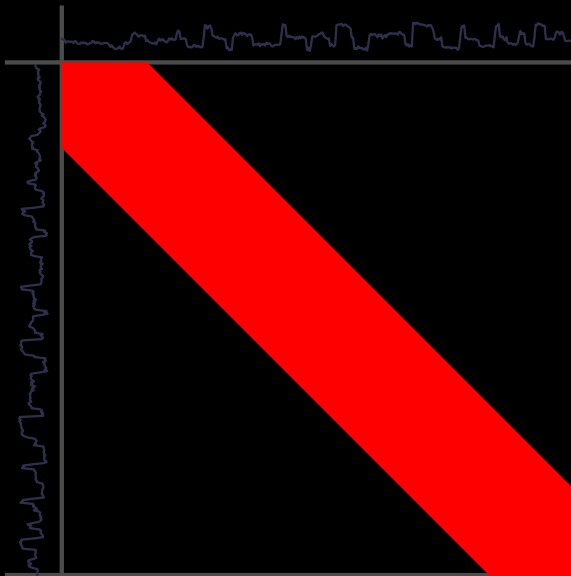
# Samozarovnanie



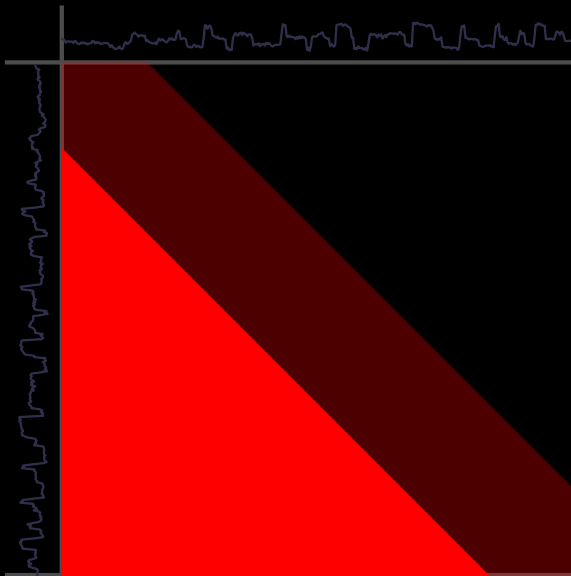
# Samozarovnanie



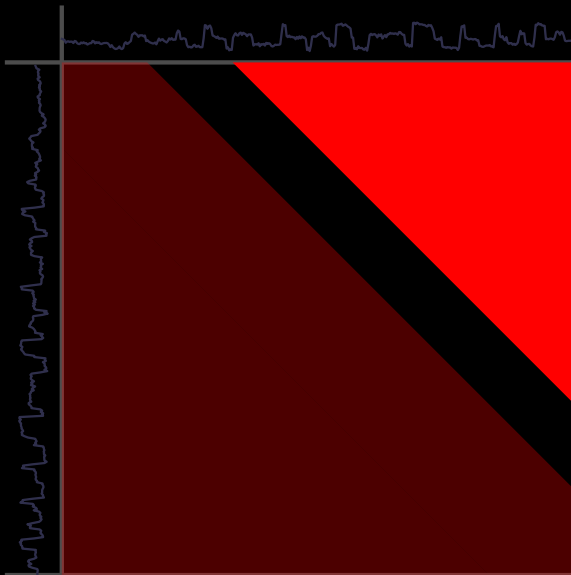
# Samozarovnanie



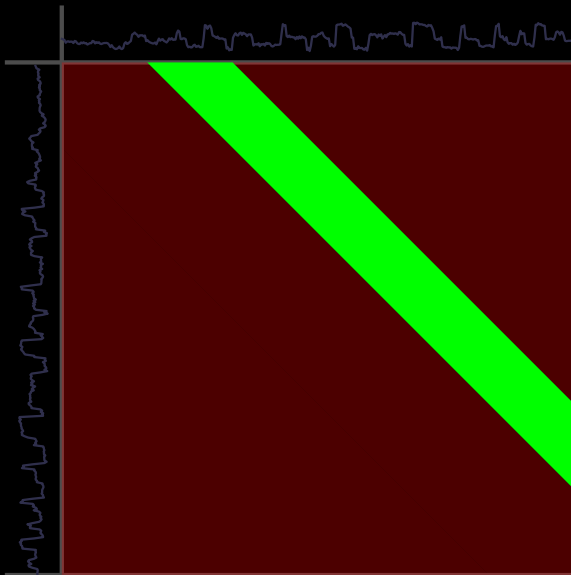
# Samozarovnanie



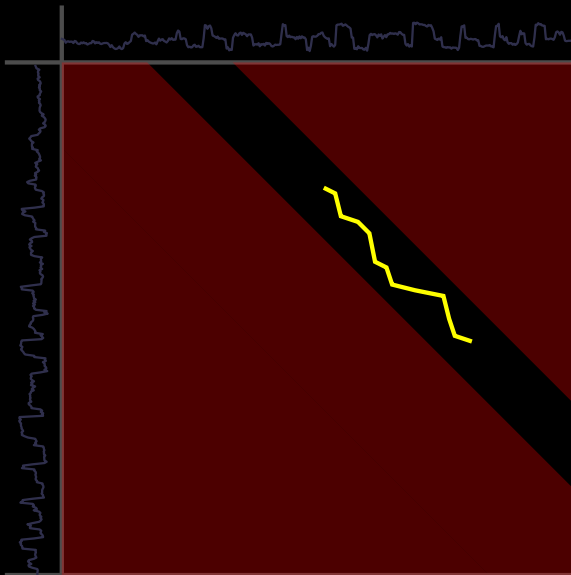
# Samozarovnanie



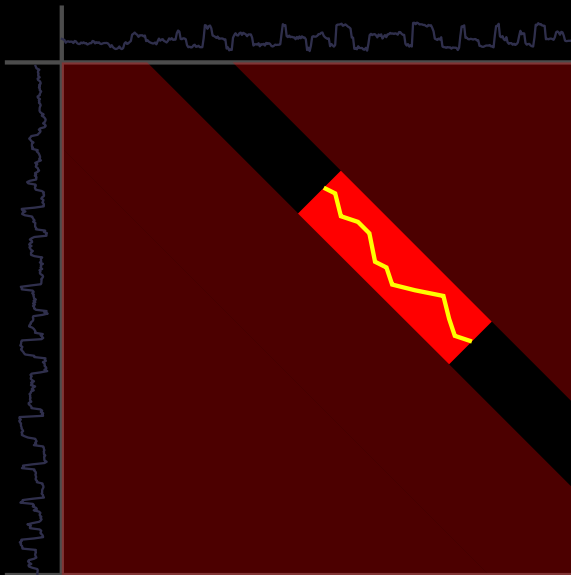
# Samozarovnanie



# Samozarovnanie

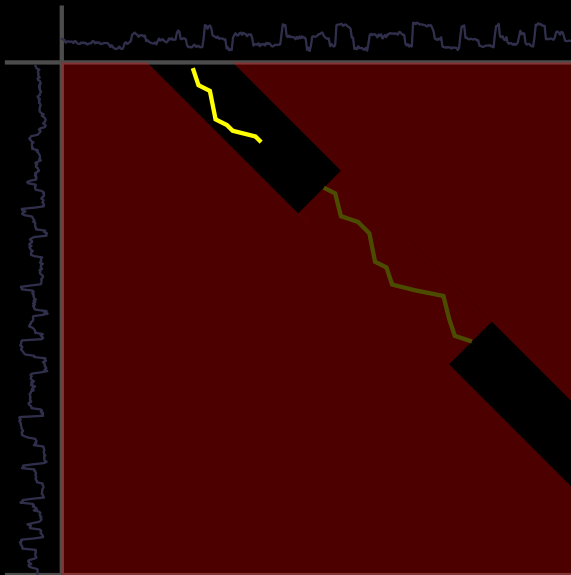


# Samozarovnanie

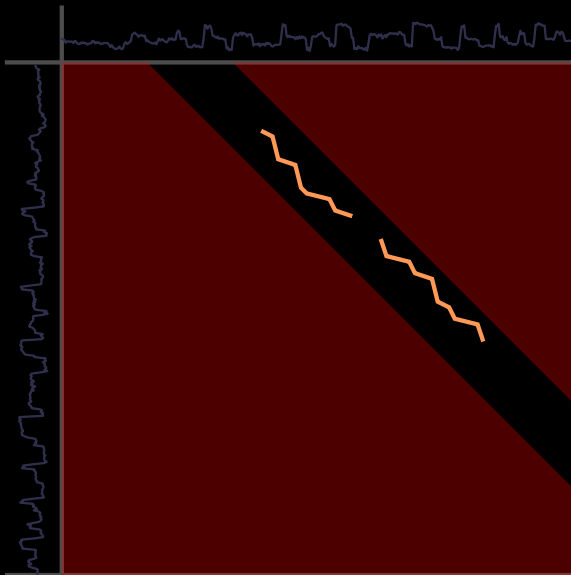




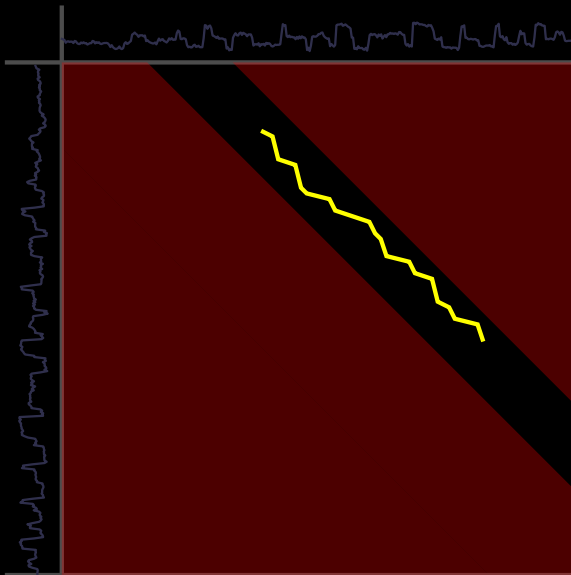
# Samozarovnanie



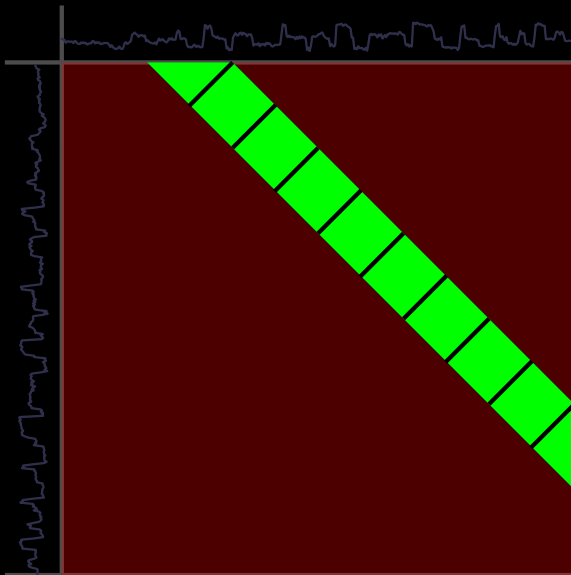
# Zlievanie ciest



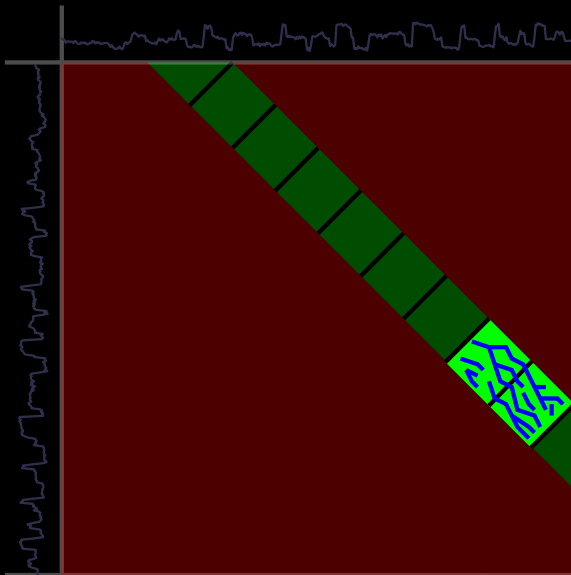
# Zlievanie ciest



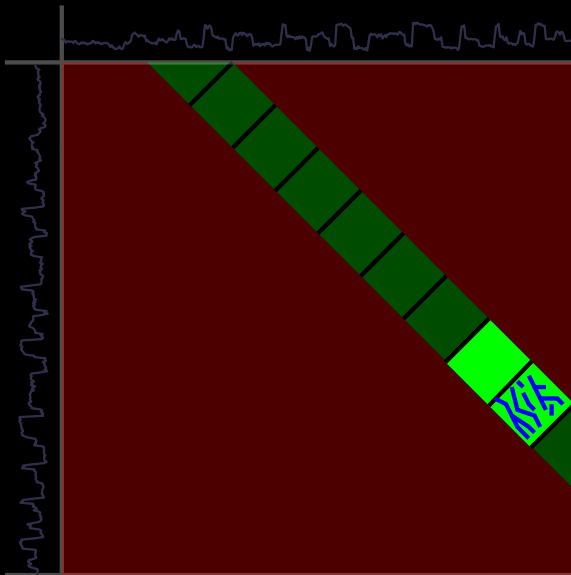
# Okienková heuristika



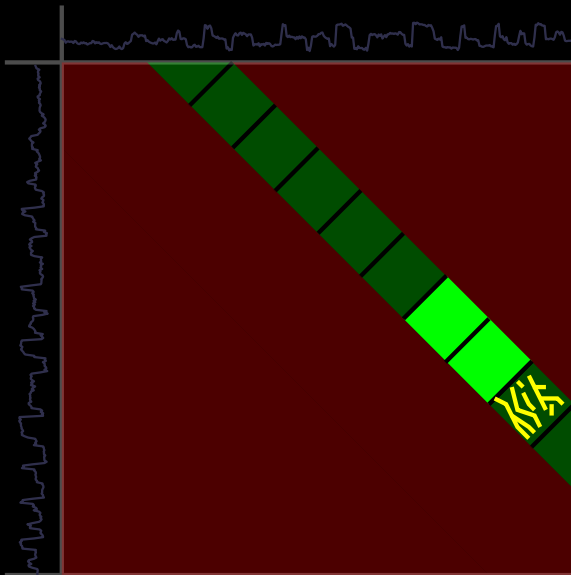
# Okienková heuristika



# Okienková heuristika



# Okienková heuristika



Efekt variantov

ATCATC

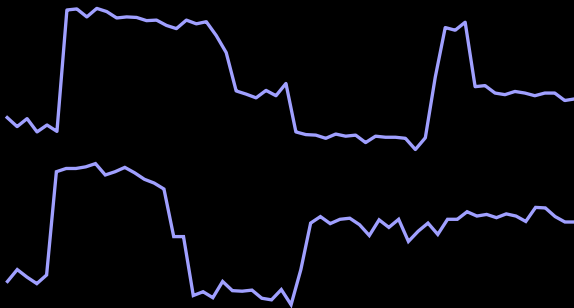


ATCATT



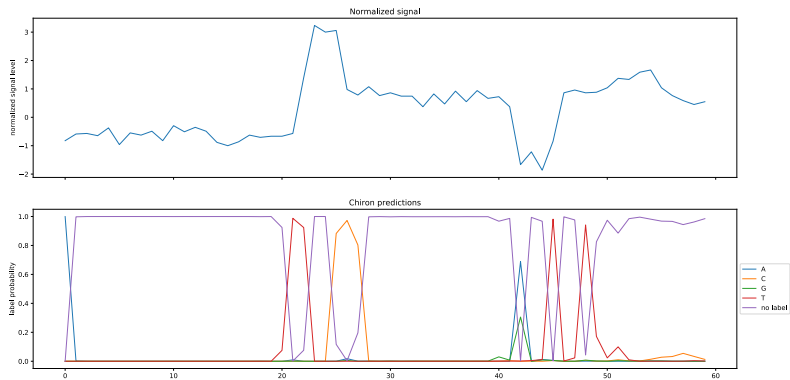
Efekt variantov

ATCATC



ATCATT

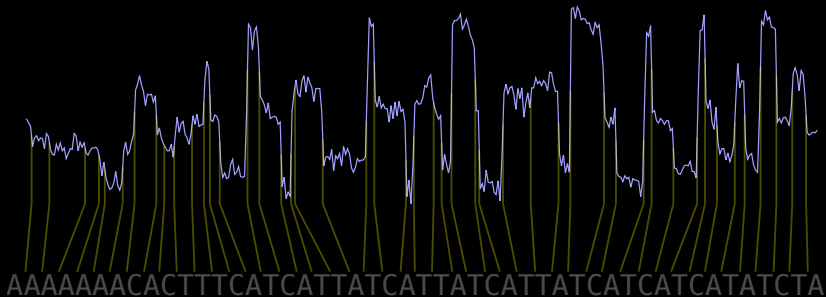
# Medzivýstup z basecallera (CNN)



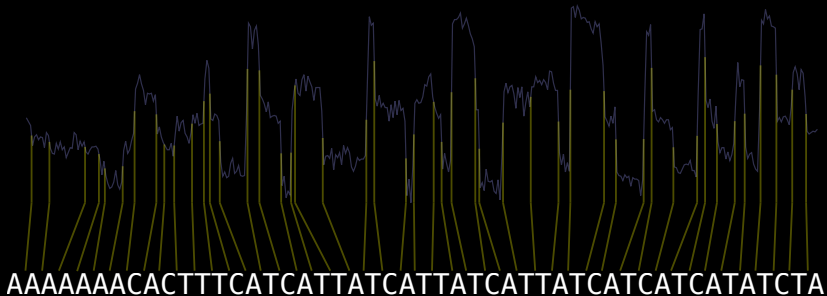
Ako definovať skóre?

Vyhodnotenie

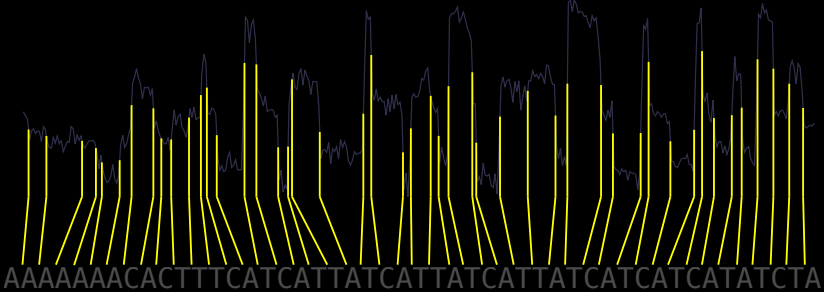
# Dáta



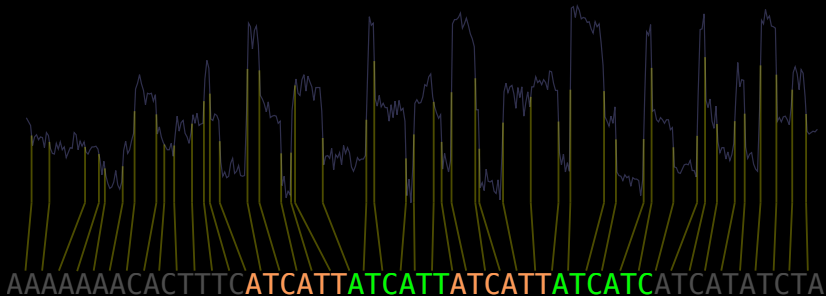
Dáta



# Dáta

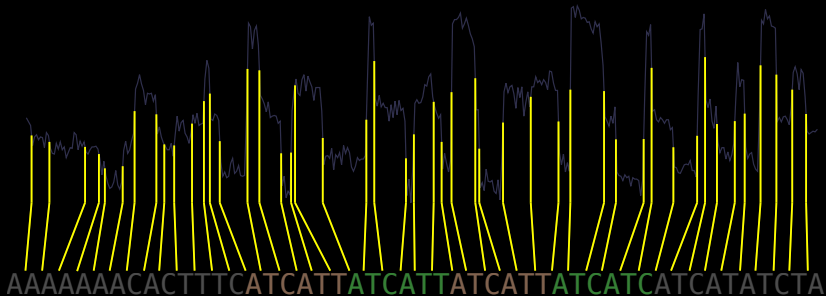


# Dáta

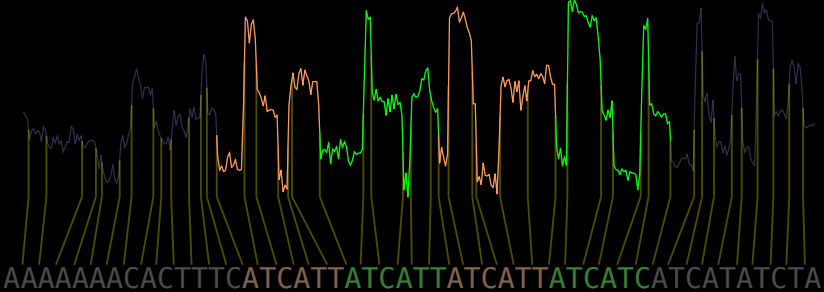




# Dáta



# Dáta



## Klasifikácia signálu

- ▶ Senzitivita – podiel pozitívnych príkladov klasifikovaných správne
- ▶ Specificita – podiel negatívnych príkladov klasifikovaných správne

## Klasifikácia signálu

- ▶ Senzitivita – podiel pozitívnych príkladov klasifikovaných správne
- ▶ Specificita – podiel negatívnych príkladov klasifikovaných správne

# Výsledky

<b>metóda</b>	<b><i>b</i></b>	<b>senzitivita</b>	<b>specificita</b>
naivná	-	6.15%	99.56%
štvorec rozdielu	0.12	61.77%	73.75%
štvorec rozdielu	0.125	63.37%	71.76%
štvorec rozdielu	0.135	66.78%	67.86%
štvorec rozdielu	0.15	71.27%	61.21%
medzivýstup basecallera	0.11	68.59%	65.35%

## Zostávajúca práca

- ▶ Skórovacia schéma pre medzivýstup z basecallera
- ▶ Metrika presnosti zohľadňujúca štruktúru
- ▶ Poriadne otestovať na viac dátach

## Zostávajúca práca

- ▶ Skórovacia schéma pre medzivýstup z basecallera
- ▶ **Metrika presnosti zohľadňujúca štruktúru**
- ▶ Poriadne otestovať na viac dátach

## Zostávajúca práca

- ▶ Skórovacia schéma pre medzivýstup z basecallera
- ▶ Metrika presnosti zohľadňujúca štruktúru
- ▶ Poriadne otestovať na viac dátach