

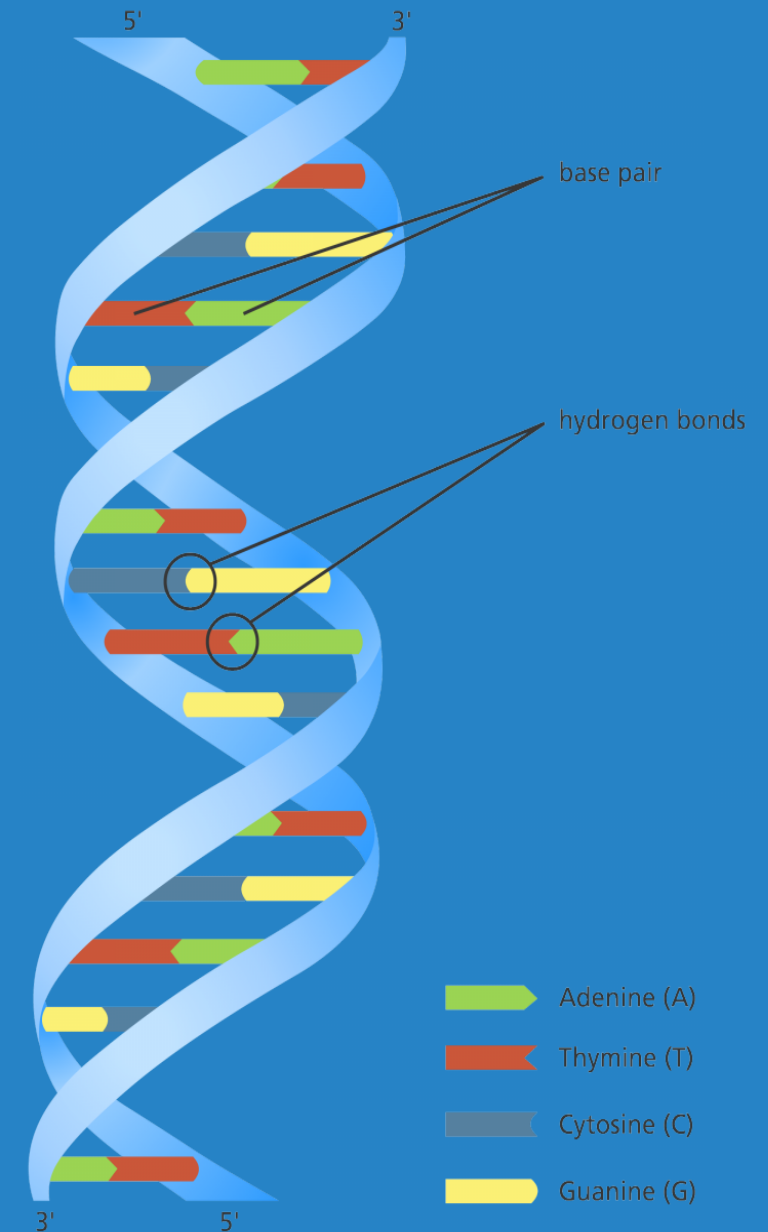
Kombinovaná predikcia štrukturálnych variantov genómu pomocou sekvenovania s nízkym pokrytím

ZUZANA KLINOVSKÁ

DR. TECHN MARCEL KUCHARÍK

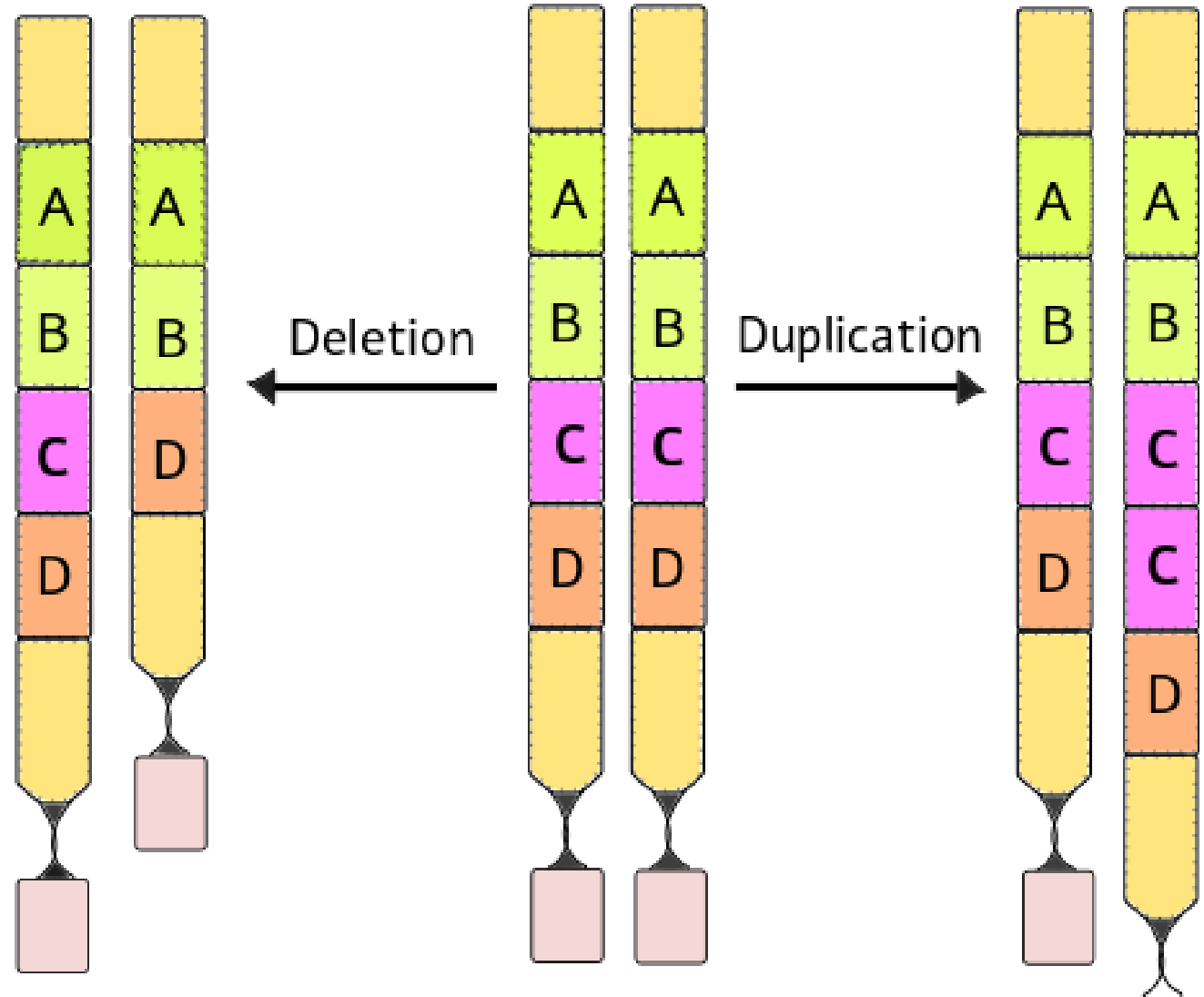
Dôležité pojmy

- DNA
- Bázy: A,C,T,G
- Mutácia-zmena



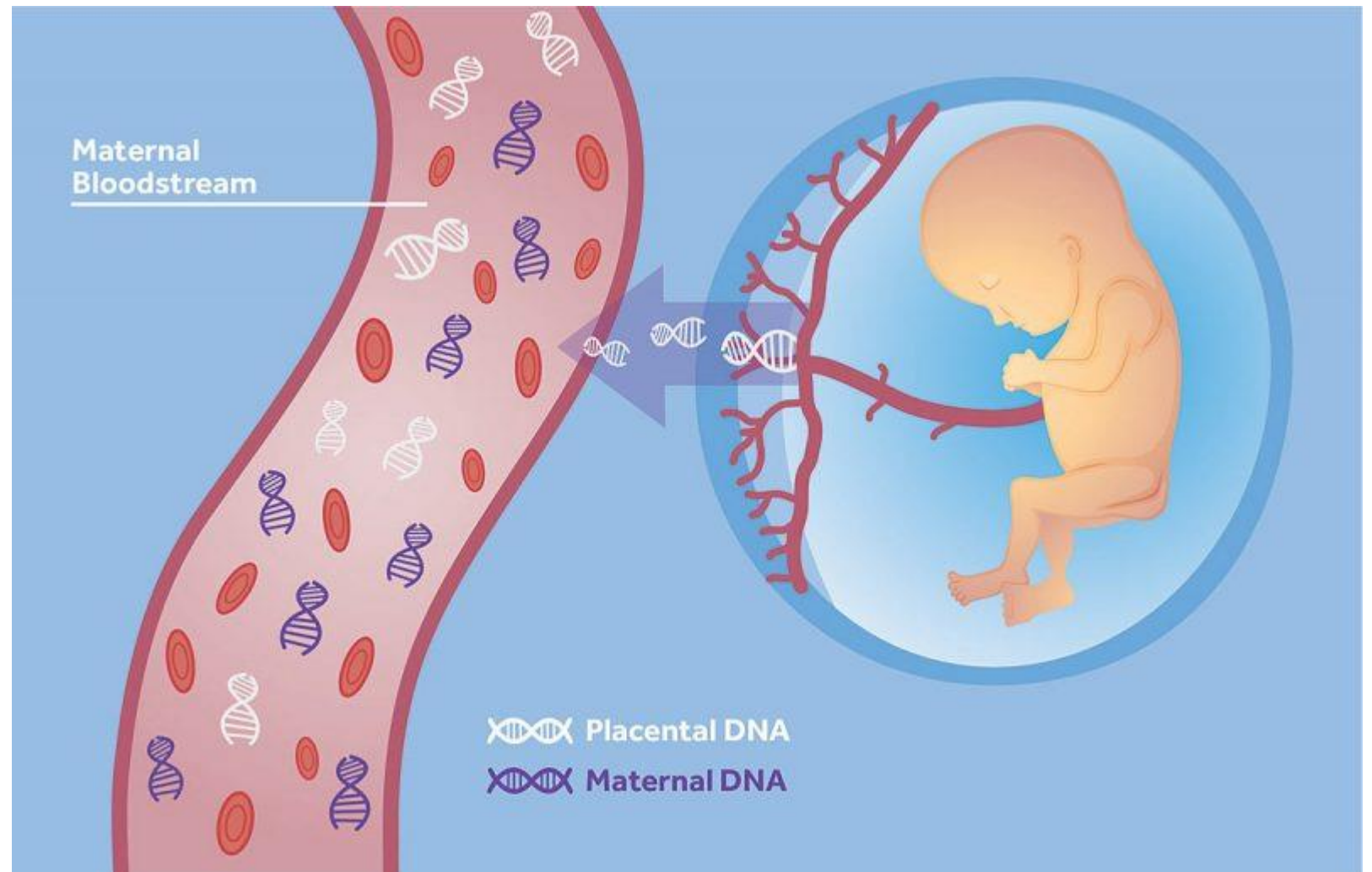
Copy Number Variations

- Štrukturálny variant
- Duplikácie, Delécie
- Pozitívne
- Negatívne –
Rázštepý, srdcové
problémy, duševné
vady



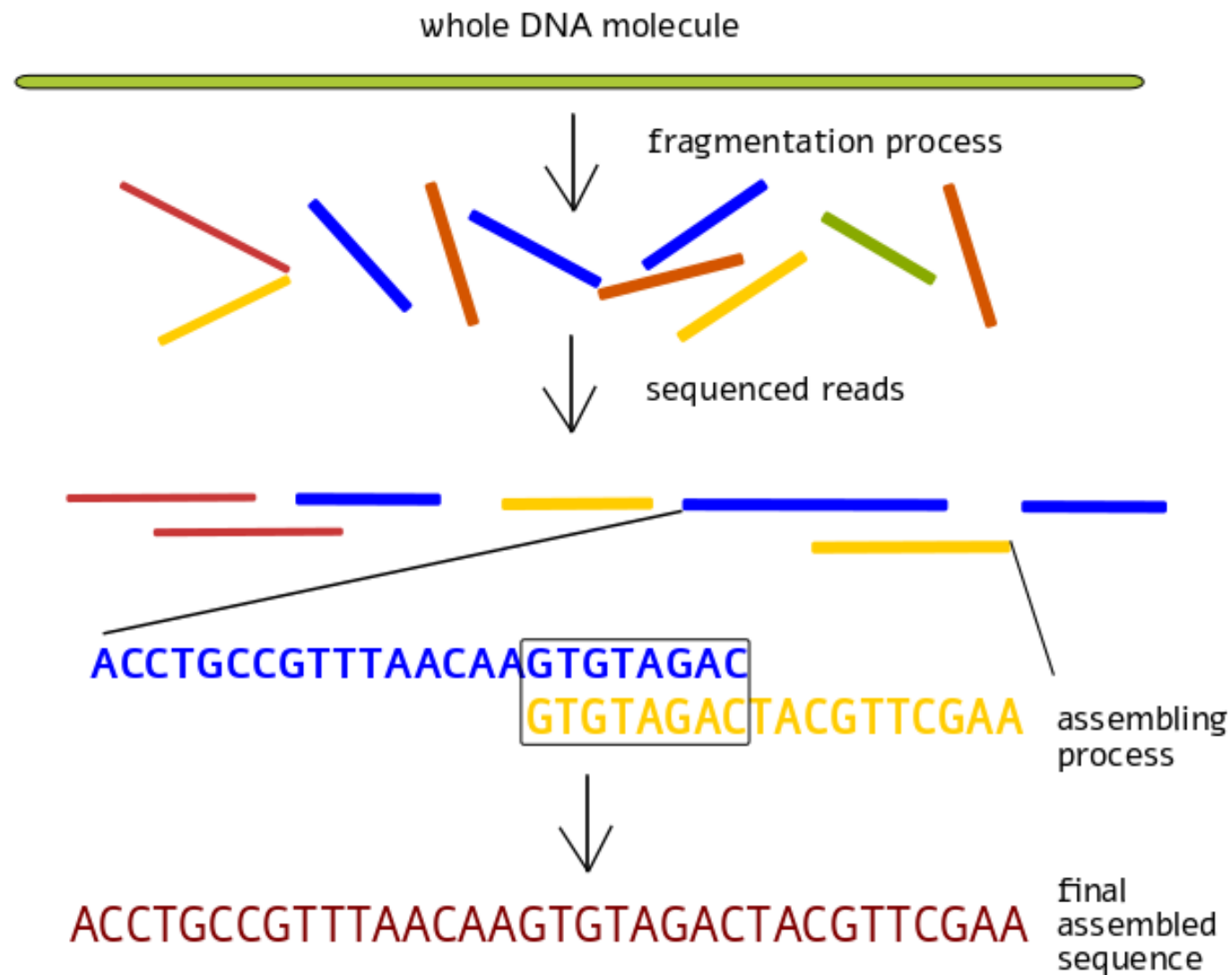
NIPT

- Non invasive prenatal testing
- Vyšetrenie plodu z krvi matky
- Bezpečné
- Min 5% DNA dieťaťa
- Fetálne, maternálne CNV



Získavanie DNA sekvencie

- Čítania
(Reads)
- Pokrytie
(Coverage)





Deep sequencing vs Low sequencing

V našej práci pracujeme s pokrytímenej ako 0.5x

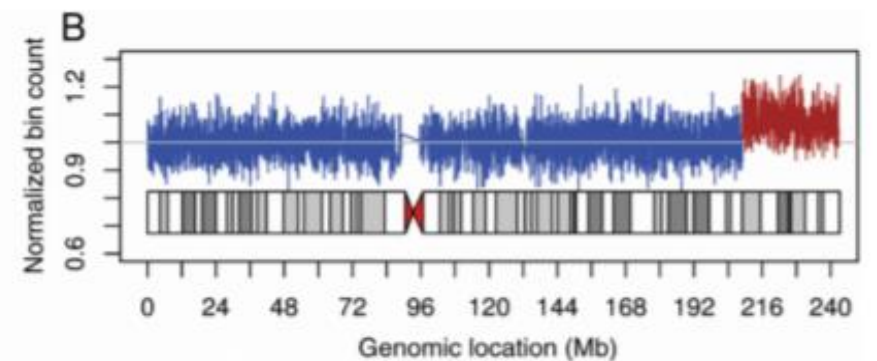
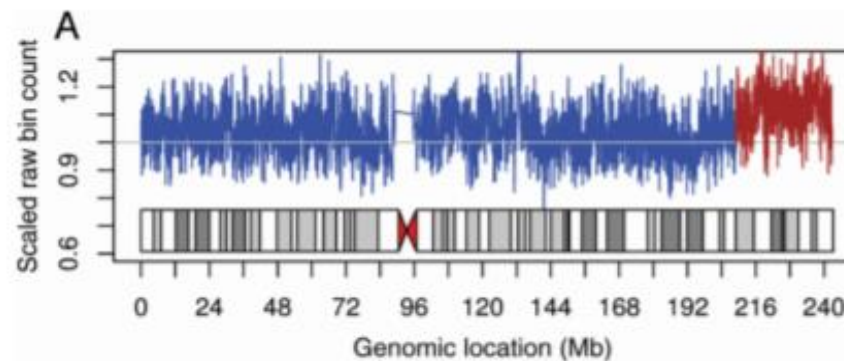
Ciele

- ✓ Porovnať nástroje na detekciu CNV
- ✓ Charakterizovať jednotlivé nástroje
- ✓ Navrhnuť predikčný model na detekciu



Predikcia CNV

- Binning
 - Rozdelenie na rovnaké segment - biny
 - Veľkosť 20 000 báz
- Normalizácia
 - 'Urovnanie'' signálu
- Segmentácia
 - CNV regióny



Porovnanie


- **Nástroje:** CNVcaller, CNVkit, iCopyDAV, WisecondorX
- Trénovanie (okrem iCopyDAV) – 134 vzoriek
- 60 CNV (19 umelých)
- Faktory: fetálna frakcia, veľkosť CNV
- Zložitosť použitia nástroja
- Dokumentácia

Porovnávanie - software

- Snakemake
 - Python prostredie
 - Spúšťanie na viacerých vzorkách

Rule: Vytvor Bam
Input: {vzorky}.sam
Output: {vzorky}.bam
Shell: Príkaz

Rule: Vytvor index
Input: {vzorky}.bam
Output: {vzorky}.bam.bai
Shell: Príkaz



Porovnanie – umelé vzorky

Size	R. count	f	CNV-caller	WisecondorX	CNVkit	iCopyDAV
0.9Mb	19.24M	5.90%	N	N	N	N
	20.36M	11.50%	D	D	N	N
	21.52M	17.30%	D	D	N	N
2.6Mb	19.6M	8.70%	N	N	N	N
	14.54M	16.69%	D	D	N	N
	19.45M	17.30%	D	D	N	N
3Mb	25.27M	4.50%	N	N	N	N
	20.59M	11.20%	D	N	D	N
	20.39M	20.10%	D	D	N	N
5.3Mb	15.3M	7.30%	D	N	N	N
	24.3M	13.40%	D	D	D	N

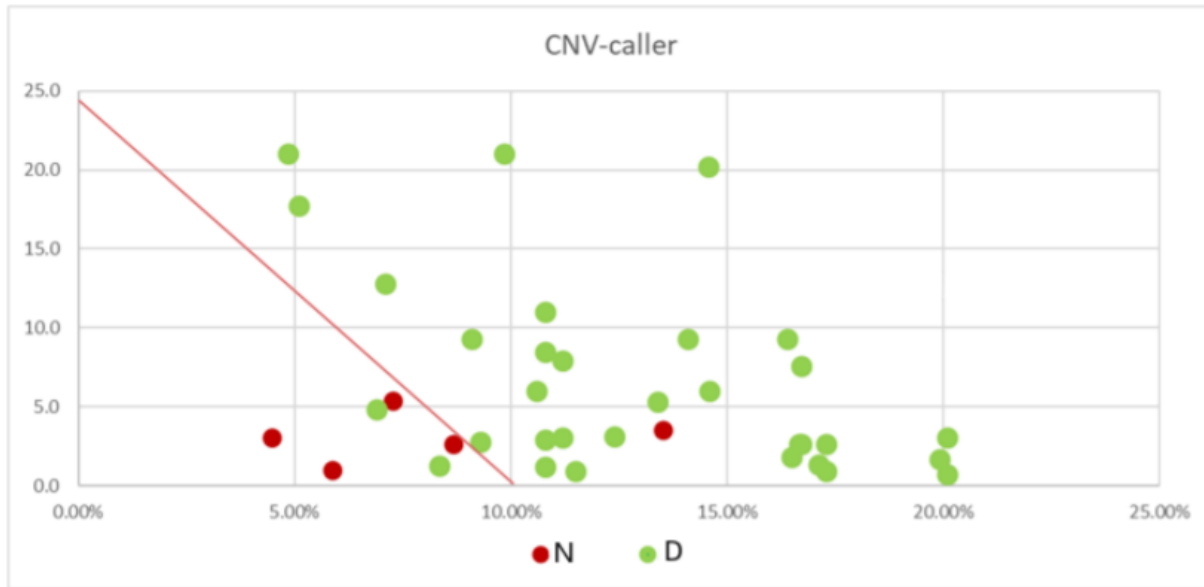
Porovnanie

1. CNV-caller
2. WisecondorX
3. CNVkit
4. iCopyDAV

Tool	Success rate (SR)	Fetal CNVs SR	Maternal CNVs SR
CNV-caller	92%	80%	100%
WisecondorX	60%	56%	73%
CNVkit	46%	16%	100%
iCopyDAV	25%	16%	70%

Porovnanie

CNV-CALLER



WISECONDORX



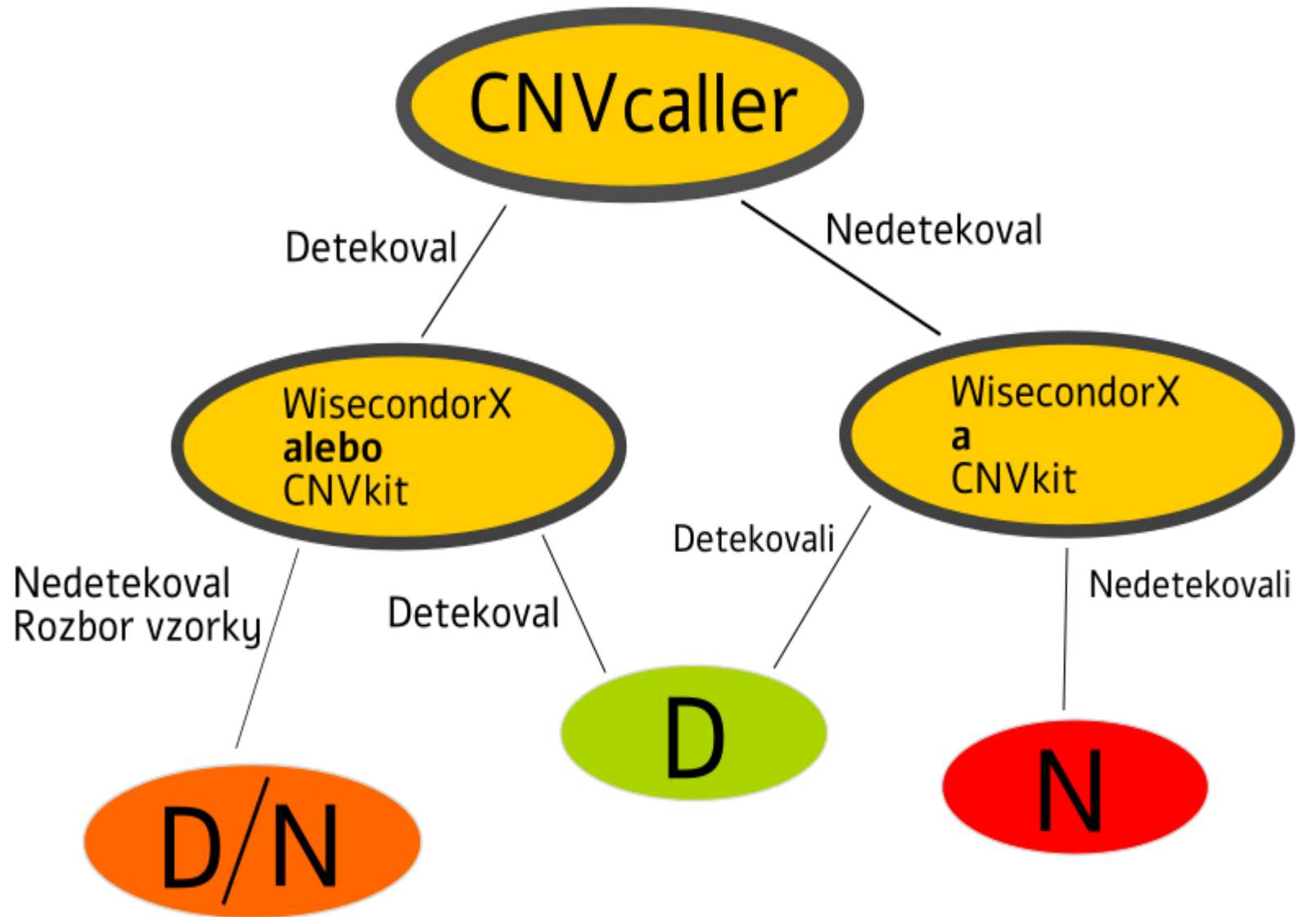
Os X - Fetálna frakcia

Os Y – Veľkosť CNV v Mb (Megabázach)

Predikčný model

Obmedziť detekciu

- Menej false positives
- Menej odhalených CNV
- Zlepšenie pri hlbšom výskume



Problémy

- Pochopenie dokumentácie, postupov, manuálov
- Server
- Porovnanie - zdĺhavé
- Nástroje – niekoľko neúspešných

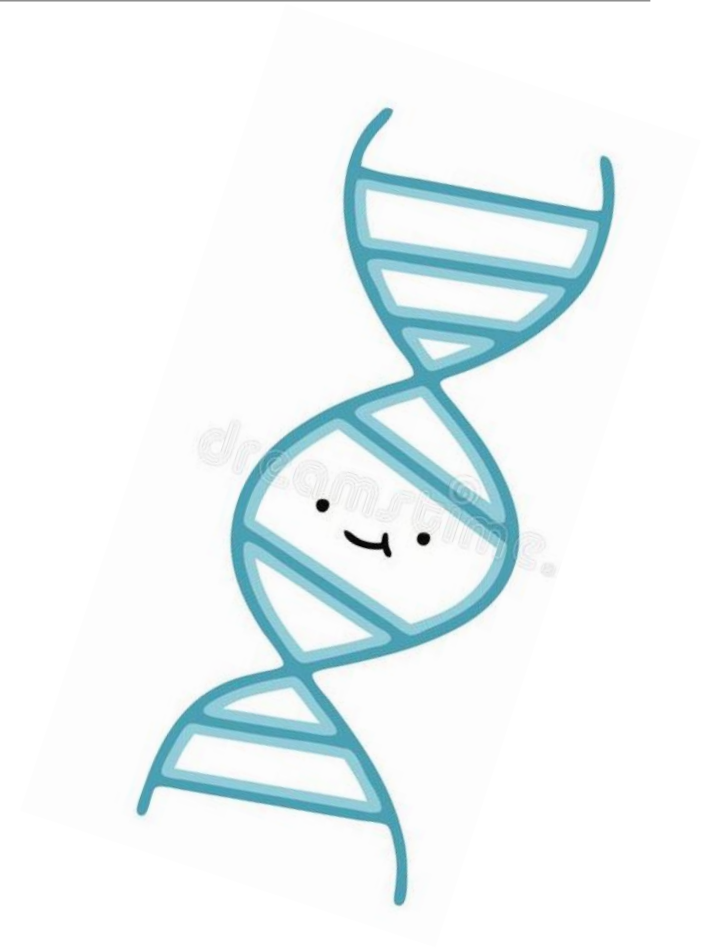


Dosiahli sme...

- Porovnanie efektivity
- Praktické využitie
- Umelé aj reálne vzorky
- Štatistiky pre umelé, reálne
- Návrh predikčného modelu



Ďakujem za pozornosť



Otázky

- Ukážte príklad technicky náročnejšej časti vašej práce, s akými problémami ste sa stretli a ako ste ich riešili.
 - Inštalovanie potrebných SW, výber vhodných nástrojov, spúšťanie nástrojov, server
- V popise algoritmu Circular binary segmentation uvádzate, že ide o rekurzívny algoritmus. Ako a na čo sa v ňom používa rekurzia