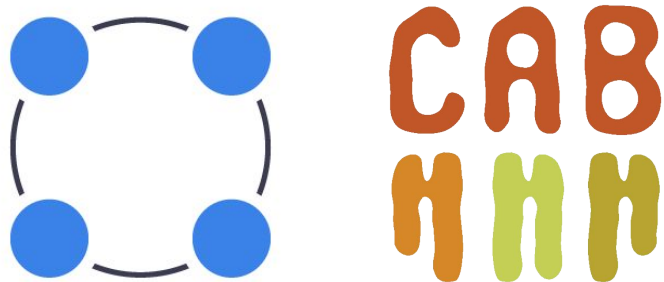


De novo CDR3 annotation in VDJ rna sequences



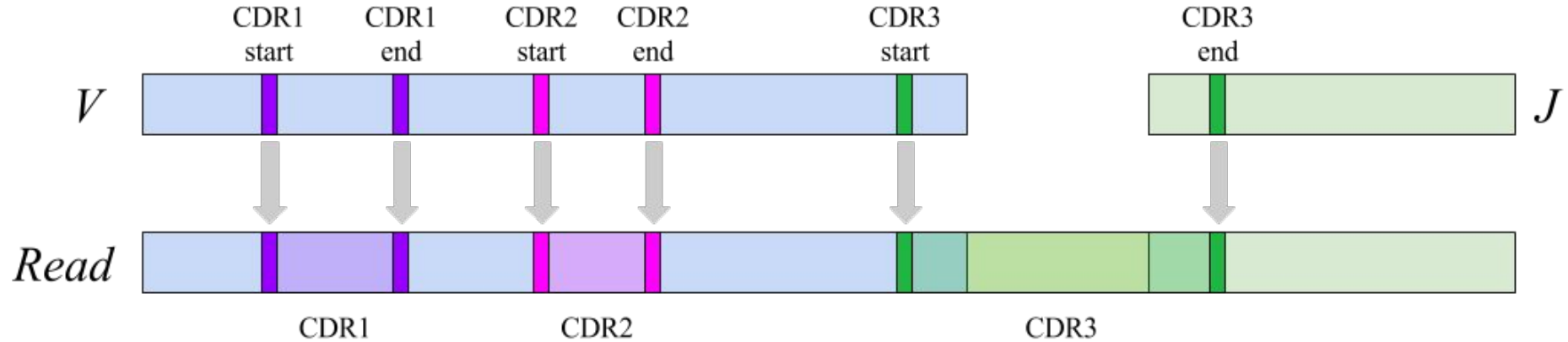
Кривоносова Кристина

руководители:

Андрей Слабодкин

Мария Черниговская

Разметчики используют *выравнивание* на germline



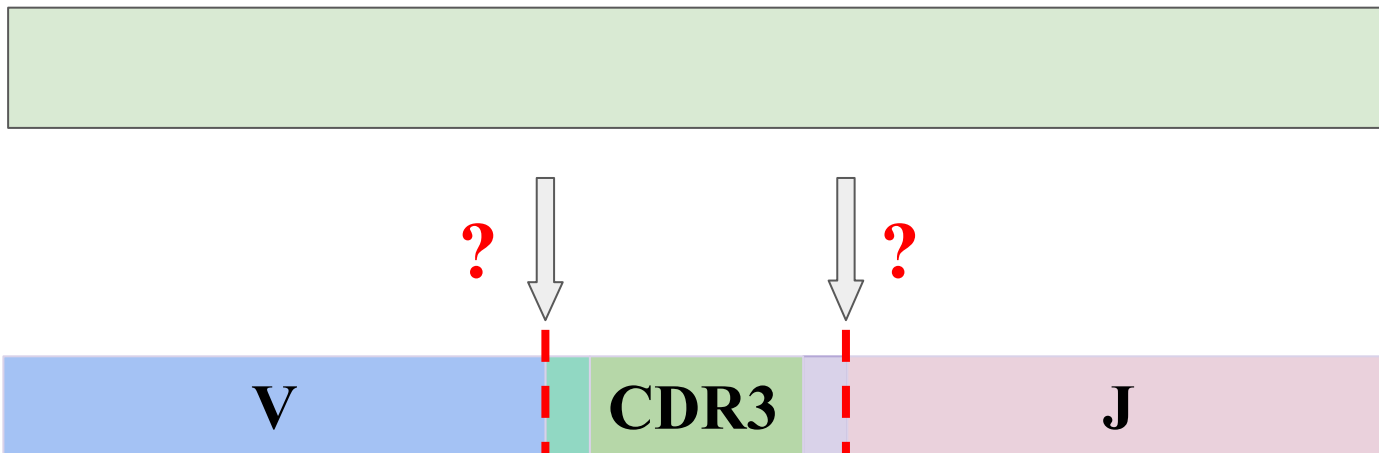
Проблемные случаи:

1. Выравнивание не существует (плохие данные)
2. Сложное строение сиквенса (например, VH replacement)

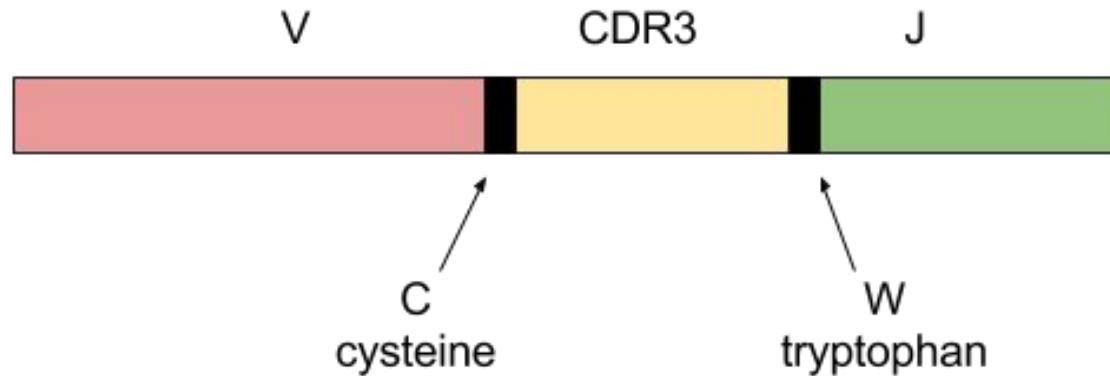
Цель проекта

Научиться размечать CDR3 без использования выравнивания

READ:



Начало: С-W эвристика



Идея: искать крайний правый cysteine и крайний левый tryptophan.

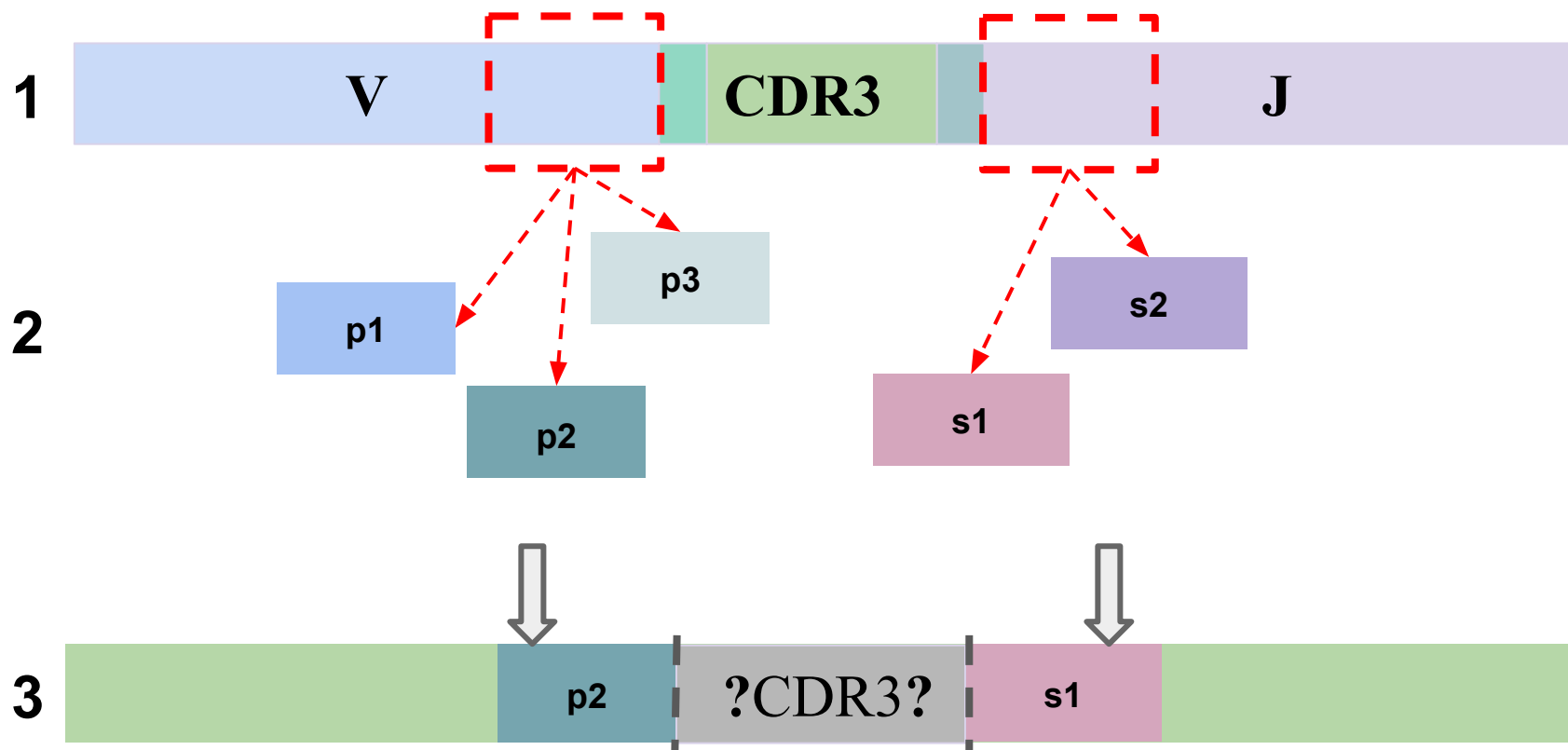
Работает для 70-80% сиквенсов.

Усовершенствованная C-W эвристика

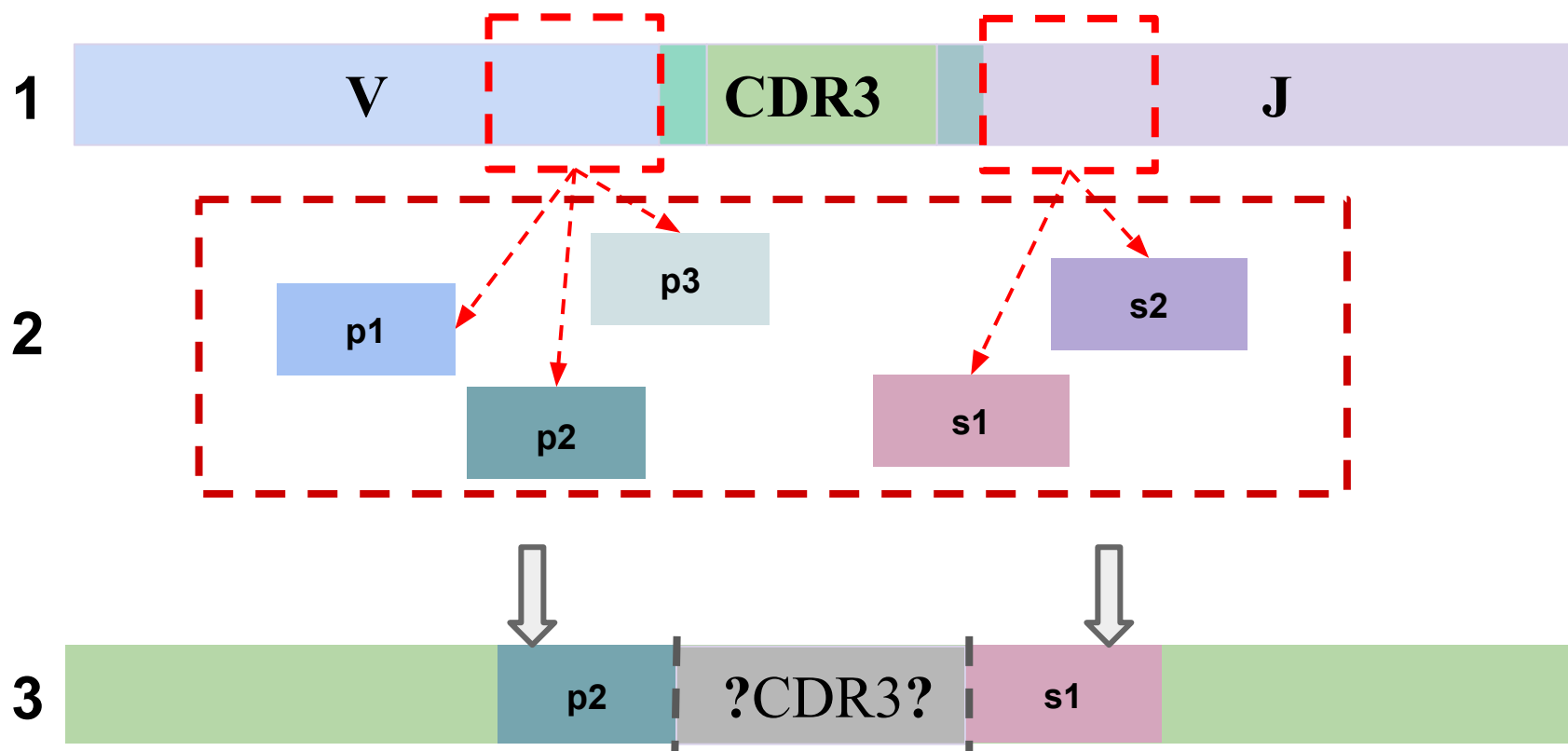


Идея: искать консервативные последовательности на границах CDR3

Принцип работы



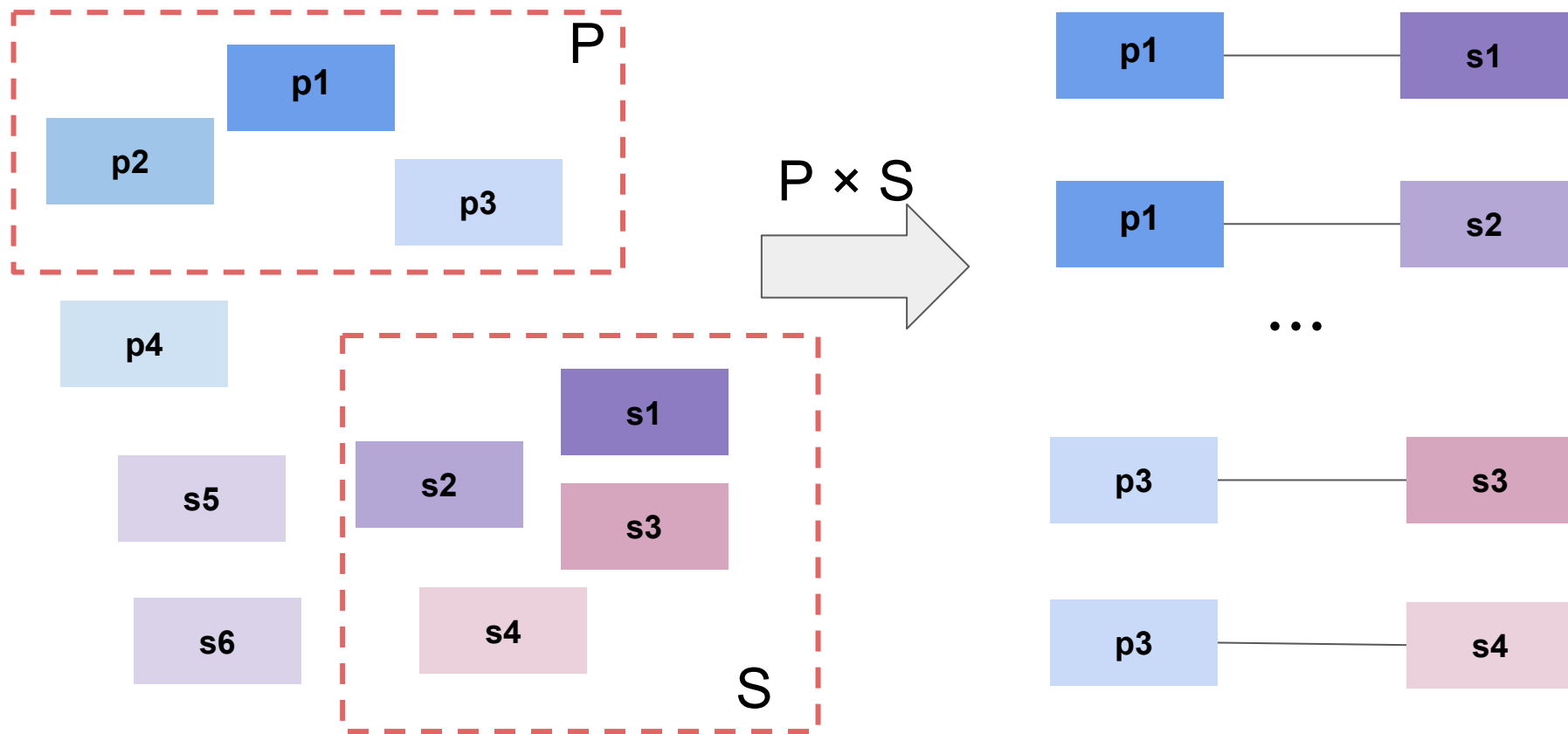
Принцип работы



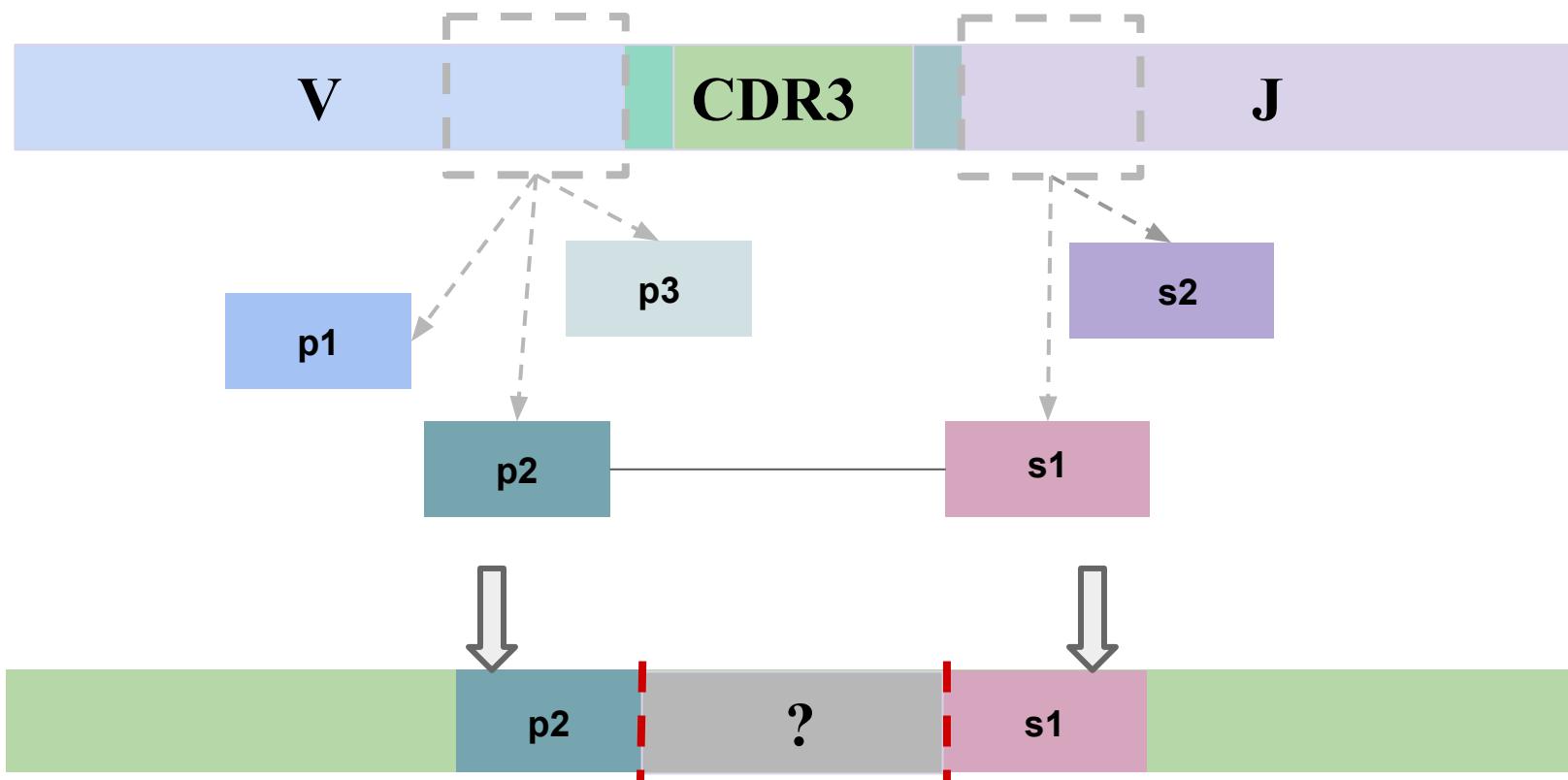
Консервативные Участки

	prefix	frequency	suffix	frequency
1	YYC	0.8452	WGQG	0.7812
2	YFC	0.0833	WGKG	0.0664
3	YYX	0.0092	WGRG	0.046
4	YHC	0.0087	WGPG	0.0151
5	XYC	0.0092	XGQG	0.0096
...				

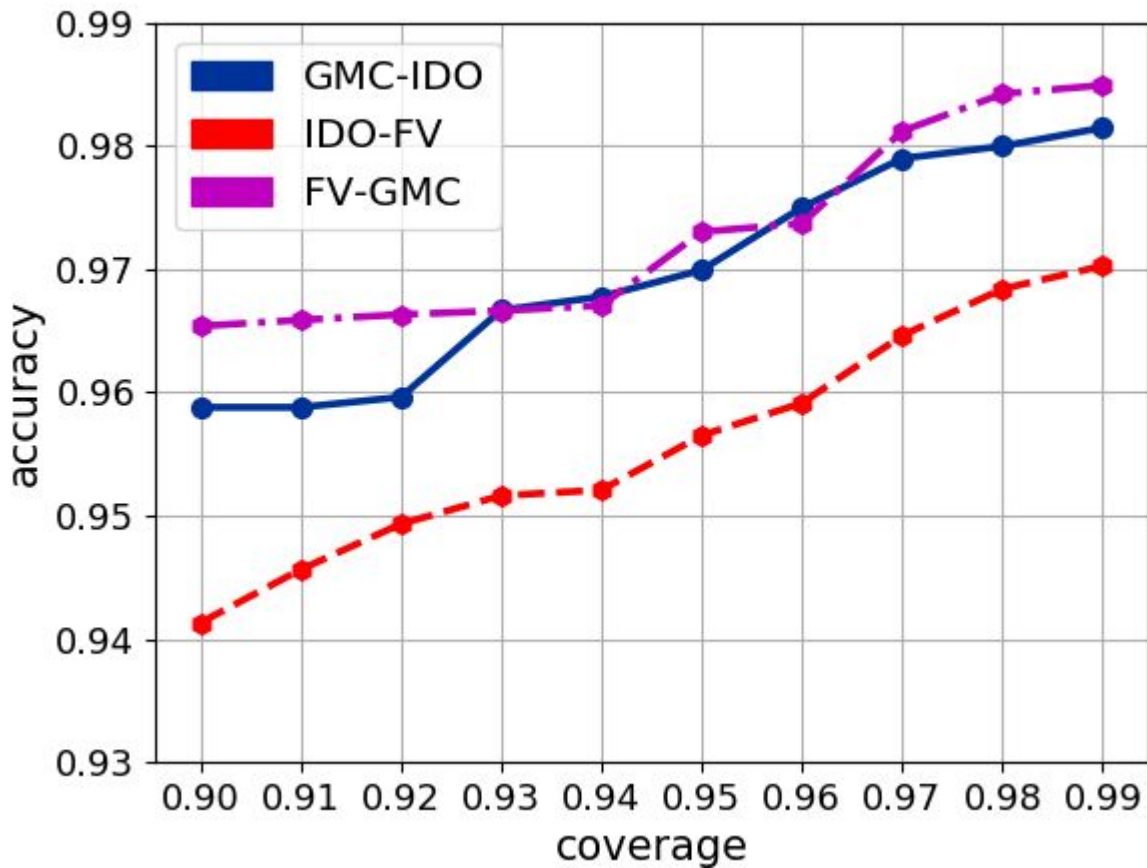
Подготовка шаблонов



Поиск CDR3



Проверка работы эвристики на различных данных



Результаты

- Реализована усовершенствованная C-W эвристика.
- Достигнуты положительные результаты на проверяемых наборах данных.

Спасибо за внимание!