



**Институт
Биоинформатики**

BIOSCAD
Biopharmaceutical Company

Поиск иммуногенных фрагментов на белках



Купрашевич Максим
Сергеева Дарья
Чуприков Павел

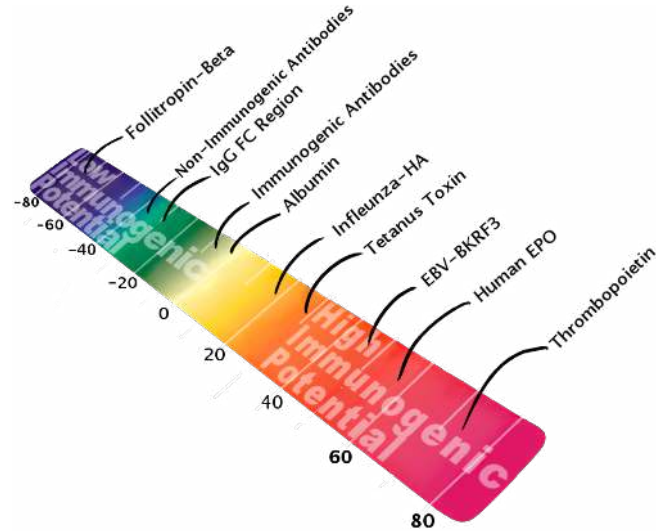
Руководитель: Яковлев Павел

Зачем это нужно?

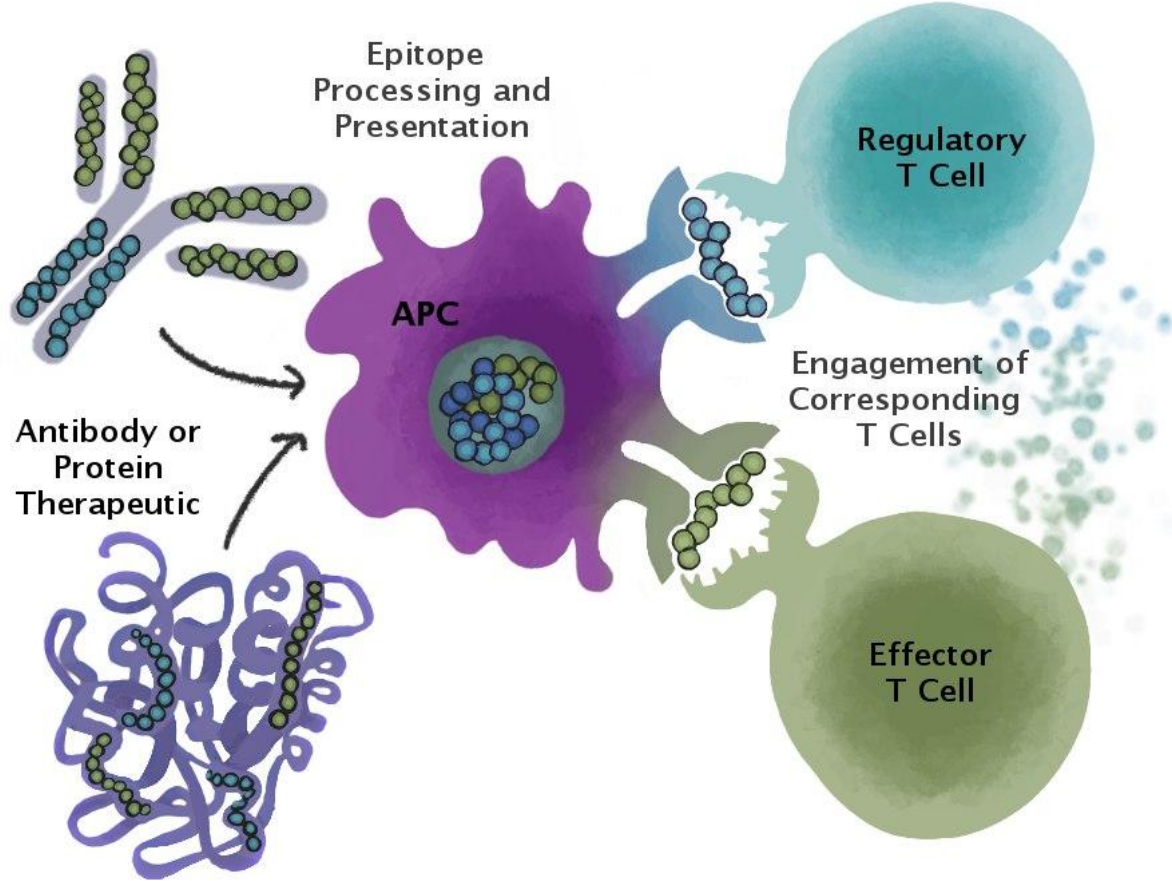
- Предсказание безопасности препаратов, содержащих рекомбинантные белки
- Анализ терапевтического эффекта, который окажут эти препараты
- Для разработки вакцин

Иммуногенность

- **Иммуногенность** — это способность антигена вызывать иммунный ответ.
- **Эпитоп** — часть молекулы антигена, которая распознается иммунной системой (антителами, рецепторами Т-клеток и В-лимфоцитов).
- Иммуногенность зависит от свойств белка и его структуры



Major histocompatibility complex class II



Цели и задачи

Цель:

Предсказание сайтов иммуногенности без сравнения с базой.

Задачи:

- Разобраться в структуре существующих баз.
- Установить взаимосвязь содержимого баз со структурными особенностями.
- Построить классификатор, обучающийся на базах и позволяющий предсказывать сайты иммуногенности без их использования.
- Реализовать инструмент, позволяющий проводить анализ белков на иммуногенность.

Immune Epitope Database (IEDB)

- Информация о Т-клеточных эпитопах, для которых экспериментально измерена аффинность для различных аллелей МНС.
- Инструменты для предсказания связывания эпитопов с МНСI и МНСII.
- Методы для предсказания иммуногенности, основанные на определении гомологичности эпитопов с известными фрагментами, присутствующими в базе.



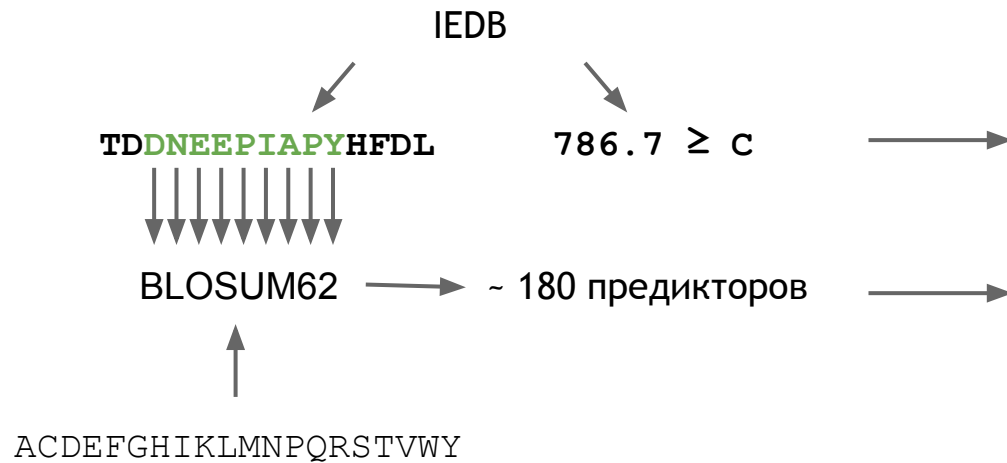
Human Leucocyte Antigens

Распределение HLA аллелей локуса DRB1 в популяции жителей Москвы.

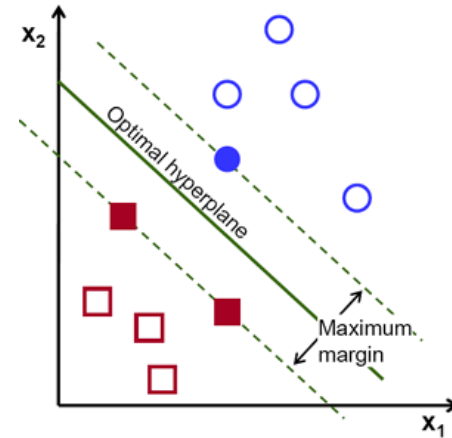


Line	Allele	Population	% of individuals that have the allele	Allele Frequency (in_decimals)	Sample Size
1	DRB1*07	 Russia Moscow		0.1430 	2,650
2	DRB1*13	 Russia Moscow		0.1390 	2,650
3	DRB1*15	 Russia Moscow		0.1330 	2,650
4	DRB1*11	 Russia Moscow		0.1300 	2,650
5	DRB1*01	 Russia Moscow		0.1100 	2,650
6	DRB1*04	 Russia Moscow		0.1100 	2,650
7	DRB1*03	 Russia Moscow		0.0800 	2,650
8	DRB1*16	 Russia Moscow		0.0460 	2,650
9	DRB1*08	 Russia Moscow		0.0350 	2,650
10	DRB1*12	 Russia Moscow		0.0260 	2,650
11	DRB1*14	 Russia Moscow		0.0190 	2,650
12	DRB1*09	 Russia Moscow		0.0140 	2,650
13	DRB1*10	 Russia Moscow		0.0100 	2,650

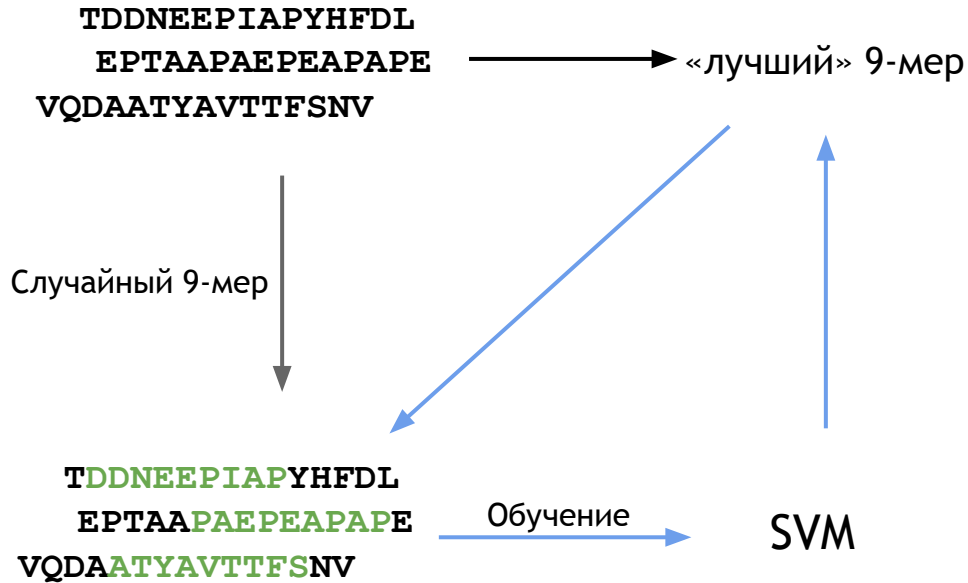
Методика классификации



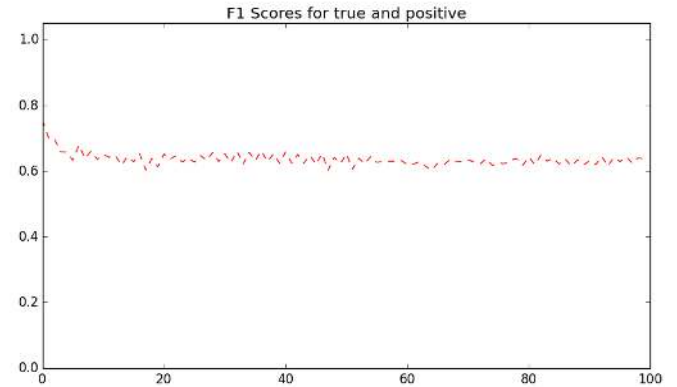
Метод опорных векторов



Поиск core-a



работает хуже,
чем случайный



Метрики классификации

- **Recall** (или Sensitivity, TPR) - отношение числа истинно положительных результатов к сумме истинно положительных и ложно отрицательных результатов:

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

- **Precision** (или PPV, Точность) - отношение числа истинно положительных результатов ко сумме истинно и ложно положительных результатов:

$$\text{PPV} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false positives}} = \frac{\text{number of true positives}}{\text{number of positive calls}}$$

- **F1-score** (F1-мера) - определяется как взвешенное гармоническое среднее точности P и полноты R:

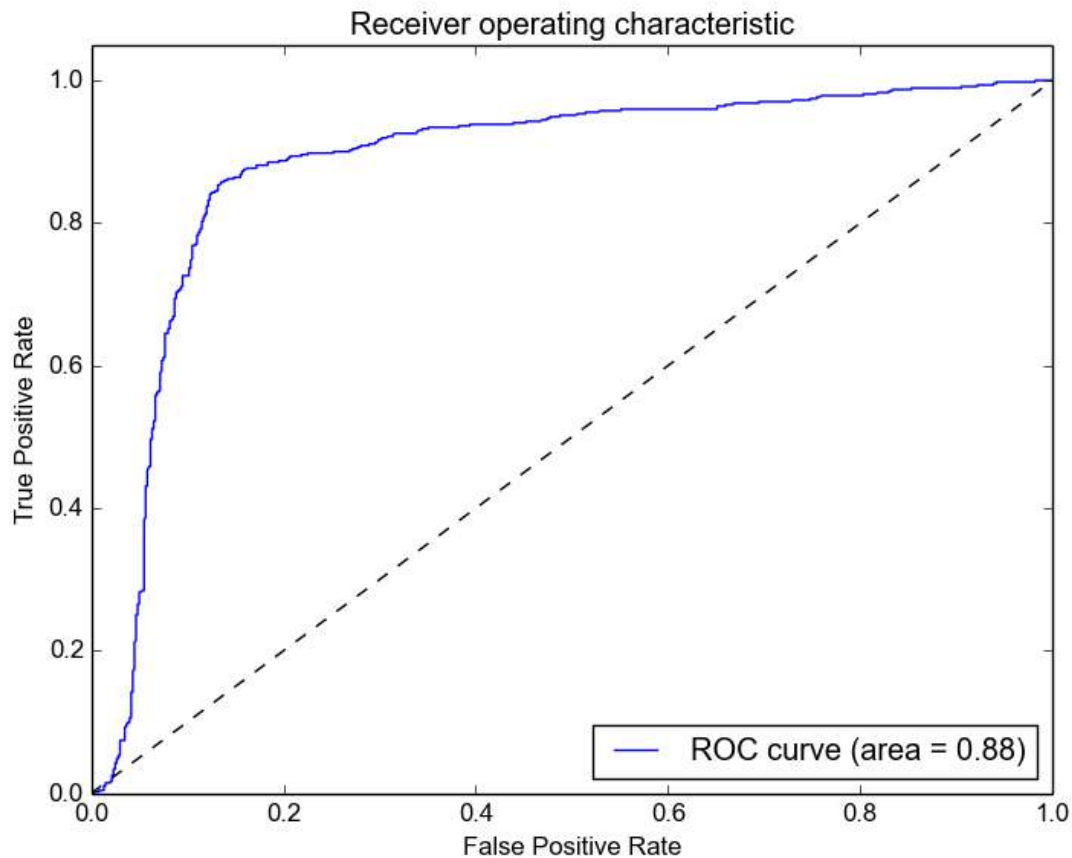
$$F_1 = \frac{2PR}{P + R}$$

- **ROC** (receiver operating characteristic, рабочая характеристика приёмника) - график, позволяющий оценить качество бинарной классификации, отображает соотношение между долей верных положительных классификаций от общего числа положительных значений
- **AUC** (area under curve, площадь под кривой) - количественная интерпретация ROC, площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций

Результаты для некоторых аллелей

Класс аллеля	F1-score	Precision	Recall	AUC
DRB1-0701	0.52	0.74	0.64	0.78
DRB1-1302	0.52	0.74	0.57	0.88
DRB1-1501	0.49	0.74	0.60	0.77
DRB1-1101	0.51	0.73	0.61	0.84
DRB1-0101	0.56	0.74	0.68	0.67

AUC для HLA-DRB1-1302



Что было сделано?

- Проанализированы современные результаты в области предсказания иммуногенности пептидов для МНСII.
- Изучена структура и нынешнее состояние базы IEDB.
- Выбраны методы и инструменты решения задачи классификации.
- Реализован скрипт, выполняющий обучение SVM на данных IEDB.
- Добавлена возможность итеративно искать наилучшего претендента на роль core МНСII.

Что можно сделать в будущем?

- Реализовать поиск coге.
- Построить нейронную сеть.
- Протестировать методы, основанные на регрессии.
- Использовать выбор предикторов и методы уменьшения размерности.
- Установить взаимосвязь со структурными особенностями.

Спасибо за внимание!

Купрашевич Максим (информатик) — undeadblow@gmail.com

Чуприков Павел (информатик) — pschuprikov@gmail.com

Сергеева Дарья (биолог) — daria.sergeeva.bio@gmail.com

Яковлев Павел (Биокад) — yakovlev@biocad.ru



BIOCAD
Biopharmaceutical Company



лама



гуманизированные