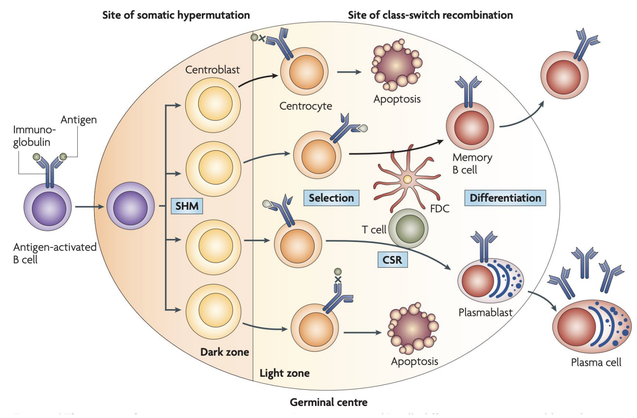
Looking for common sense: modeling B-cell repertoires using machine learning

SMTB



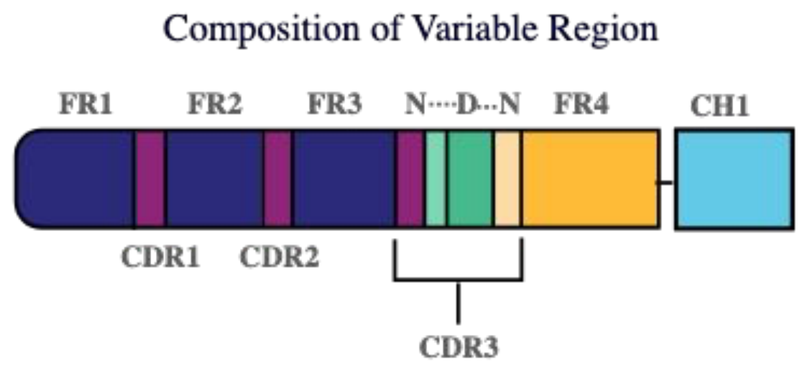
Sabancı Üniversitesi



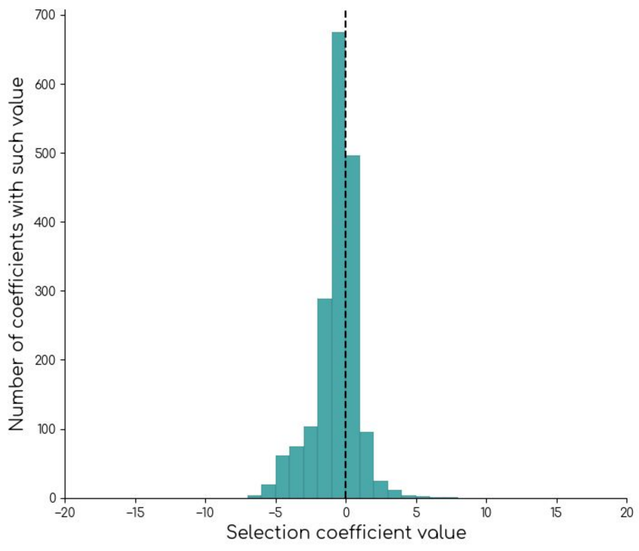
Introduction

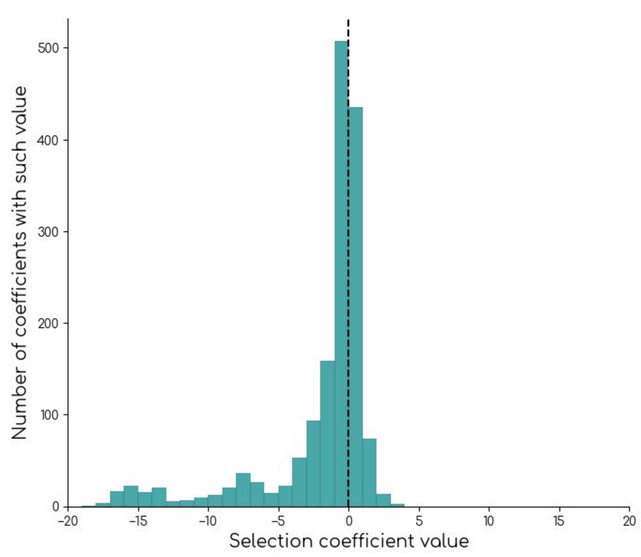
В-клетки адаптивного иммунитета несут рецепторы для распознавания антигенов. При активации наивные клетки проходят созревание аффинности, генерируя улучшенные клоны с усиленным связыванием. Этот процесс уникален для каждой инфекции, но существуют ли общие правила такой генерации и способно ли машинное обучение их выучить?

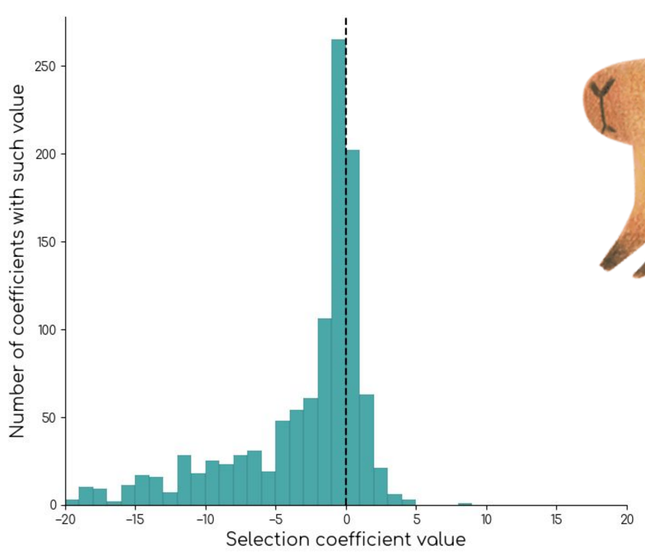
B cells of the adaptive immune system carry receptors for recognising antigens. Upon activation, naïve B-cells undergo affinity maturation, generating improved clones with enhanced binding affinity. This process is unique to each infection, but are there any general rules governing such generation, and can machine learning identify them?



There are a lot of neutral selection factors, and more negative ones than positive ones



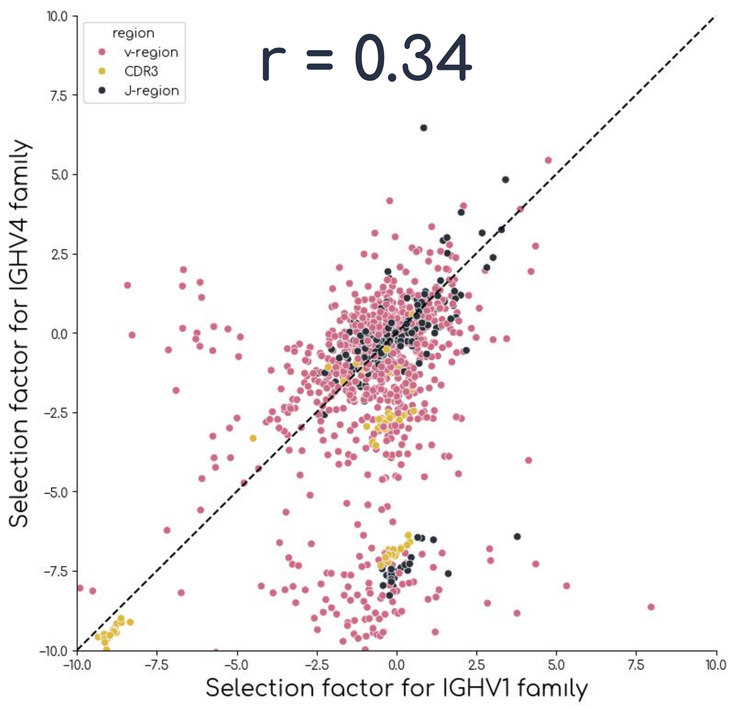




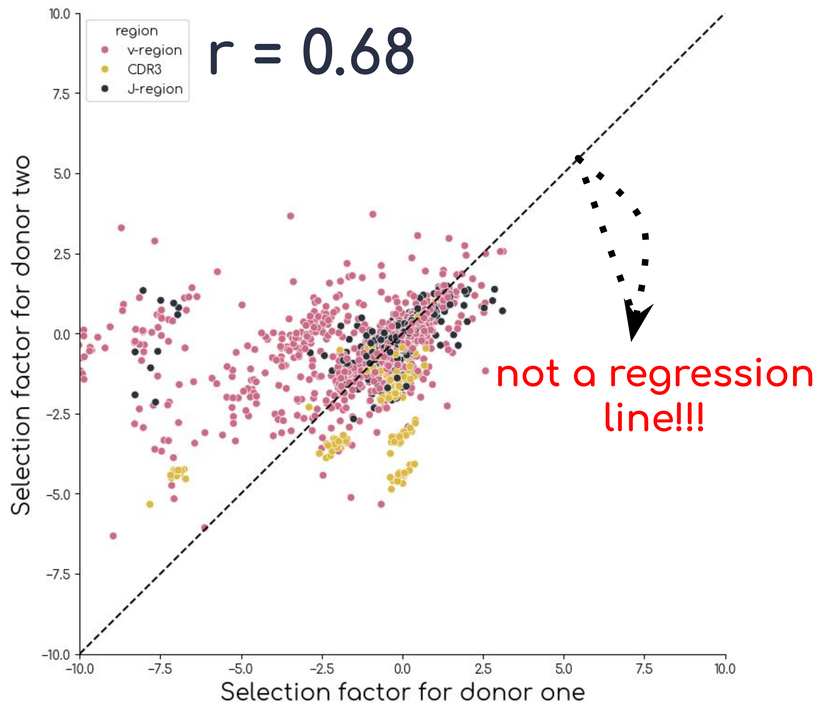


Selection factors from different data are not the same but they correlate

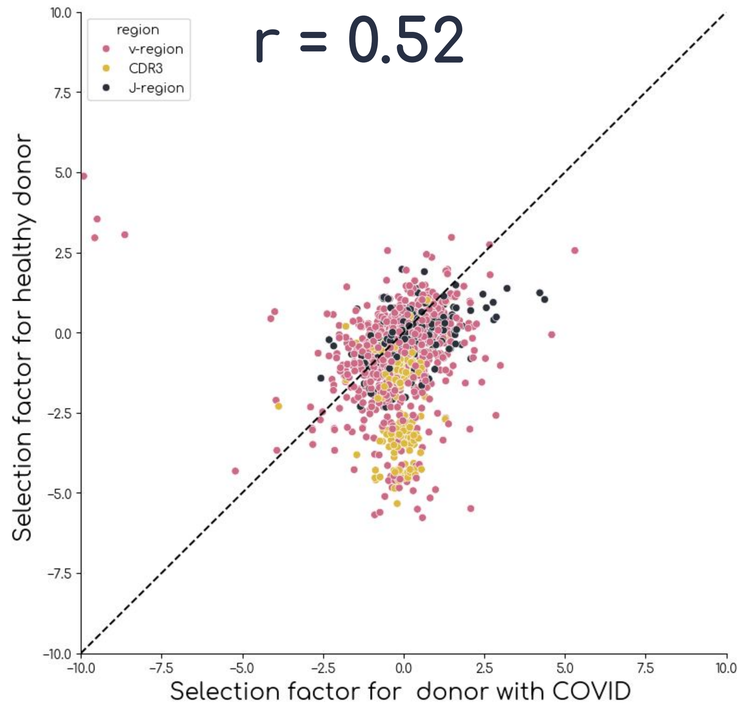
Different V gene families



Different healthy donors

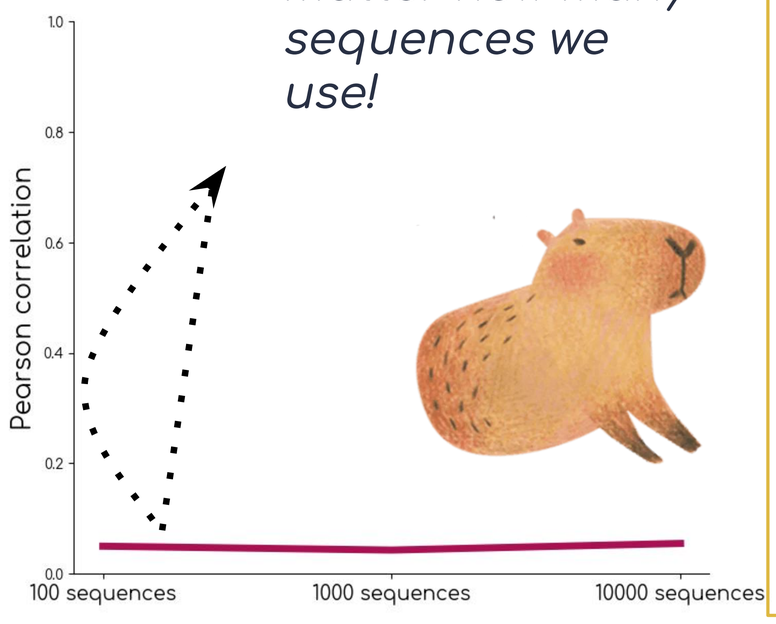


Healthy and COVID-infected donors



Deep learning vs gradient descent-inferred selection factors do not correlate :(

no correlation no matter how many sequences we use!



Conclusions

1. Семейство V гена влияет на значение фактора селекции для конкретной аминокислоты на конкретной позиции/ .The V gene family affects the selection factor value for a specific amino acid at a specific position The V gene family affects the selection factor value for a specific amino acid at a specific position

2. Факторы селекции, выученные на репертуарах разных индивидов коррелируют, но всё равно зависят от индивида у двух здоровых людей. А факторы селекции у здорового и больного различаются сильнее, чем у двух здоровых/ Selection factors learned from the repertoires of different individuals correlate, yet remain individual-specific in two healthy subjects. Moreover, selection factors differ more significantly between a healthy individual and a patient than between two healthy individuals.

3. Выбранный нами метод не позволяет сравнить факторы селекции, полученные с помощью моделей глубокого обучения и полученные с помощью градиентного спуска/ Our method does not allow for a comparison of the selection factors obtained using deep learning models with those obtained via gradient descent.