

BioBootCamp MIPT - 2020

РАЗРАБОТКА МОДЕЛИ ВАКЦИНЫ ПРОТИВ SARS-CoV-2



06 / 12 / 2020

Команда А!

Минск, Молодечно
Беларусь

Обзор

Выявлены наиболее перспективные мишени для создания вакцины.

Проанализирована их консервативность, на основании чего был начат поиск потенциальных эпитопов. Возможные эпитопы проанализированы на сродство к наиболее часто встречающимся в популяции аллелям МНС-I и МНС-II. Предложен способ прямой активации деятельности В-клеток. Полученная модель вакцины имеет низкое сродство к человеческому протеому, что снижает риск возникновения аутоиммунной реакции. Кроме того, используемые эпитопы близки к другим видам зоонозных коронавирусов, что дает возможность предположить некоторый иммунитет в отношении этих видов. Представлена модель вектора для доставки вакцины на основе генома AAV.

Технические подробности

При работе использовались следующие источники и общедоступные сервисы:

1. База данных NCBI: <https://www.ncbi.nlm.nih.gov/nucleotide/>
2. База данных PDB: <https://www.rcsb.org/>
3. Allele frequency net database: <http://www.allelefrequencies.net/>
4. IEDB Analysis Resource: <http://tools.iedb.org/main/>
5. NetChop: <http://www.cbs.dtu.dk/services/NetChop/>
6. The International Union of Biochemistry and Molecular Biology: <https://iubmb.org/>
7. Enzyme Nomenclature Database: <https://www.enzyme-database.org/index.php>
8. Geneious Prime

Для достижения результата использовались некоторые скрипты Python3, написанные самостоятельно.

Ключевые этапы

1. На первом этапе нами была изучена структура вирусной частицы SARS-CoV2. На основании этого было выдвинуто предположение, что главными источниками эпитопов могут являться 4 структурных белка: S, M, N и E. Однако, по нашему мнению, белок N не является хорошим решением для «обучения» иммунной системы человека распознавать SARS-CoV2, т.к. он не является агентом, первостепенно распознаваемым организмом.

Так, на первом этапе мы исключили из потенциальных мишеней для вакцины N-белок.

2. Далее мы использовали базу данных аминокислотных последовательностей белков SARS-CoV2 из NCBI. Здесь были случайным образом отобраны по 500 аминокислотных последовательностей каждого из трех ранее выбранных белков. В каждом случае аминокислотные последовательности были получены в диапазоне от марта до ноября в разных частях света: от Америки до Австралии. После этого было проведено множественное выравнивание аминокислотных последовательностей каждого белка. С помощью встроенного вьювера выравнивание было перекрашено таким образом, чтобы выделить консервативные участки.

S белок



М белок

Е белок

[illegible]

Из S белка были взяты участки 1-257 (A), 524-1273 (B). Из M белка 117-127 и 211-222. Белок E является очень консервативным, поэтому для потенциальных эпитопов был взят полностью (C).

3. После этого мы проверили сходство найденных консервативных участков к протеому человека с помощью BLAST. На этом этапе в связи с высокой гомологией к человеческому протеому были отброшены участки белка M.
4. Далее с помощью сервиса NetChop были предсказаны сайты разрезания отобранных кусочков протеасомой с пороговым значением 75%. Из полученных кусочков были отброшены те, что имеют длину меньше 10. Для этого была написана программа. На данном этапе мы получили потенциальные эпитопы.
5. На пятом этапе мы стали проверять сходство потенциальных линейных эпитопов к MHC-I и MHC-II. В базе данных Allele frequency net database были определены самые распространенные аллели MHC-I и MHC-II у жителей России. Исходя из полученных результатов при проверке на сходство потенциальных эпитопов к самым распространенным аллелям были отобраны самые перспективные эпитопы для аллелей MHC-I. Ни одни из предполагаемых эпитопов не дал удовлетворительных результатов на сходство с MHC-II. Тогда было принято решение посмотреть, какие эпитопы будут определены для MHC-II как наиболее близкие в изначальных консервативных последовательностях. Были выбраны 10 самых перспективных последовательностей, однако, как мы могли заметить раньше, внутри этих последовательностей находятся сайты разрезания протеасомой. Тогда мы начали преобразовывать эти последовательности таким образом, чтобы сократить количество данных сайтов и сохранить изначальные свойства последовательностей. Таким образом мы смогли преобразовать 3 из 10 предполагаемых последовательностей. Таким образом из GINITRFQTLLALHRSYL, KRSFIEDLLFNKVTL, FQFCNDPFLGVYYHK после разрезания протеасомой получим эпитопы GINITRFQTLLAL, KRSFIEDLLFNKVTL, CNDPFLGVYYH.

6. Нами было замечено, что чаще всего протеасома проводит разрез по участкам, содержащим алифатические и ароматические аминокислоты, исходя из этого были предложены разделители, по которым будет происходить разрез протеасомой для эпитопов. Таким образом все определенные эпитопы были “склеены” со вставками из алифатических и ароматических аминокислот между ними:

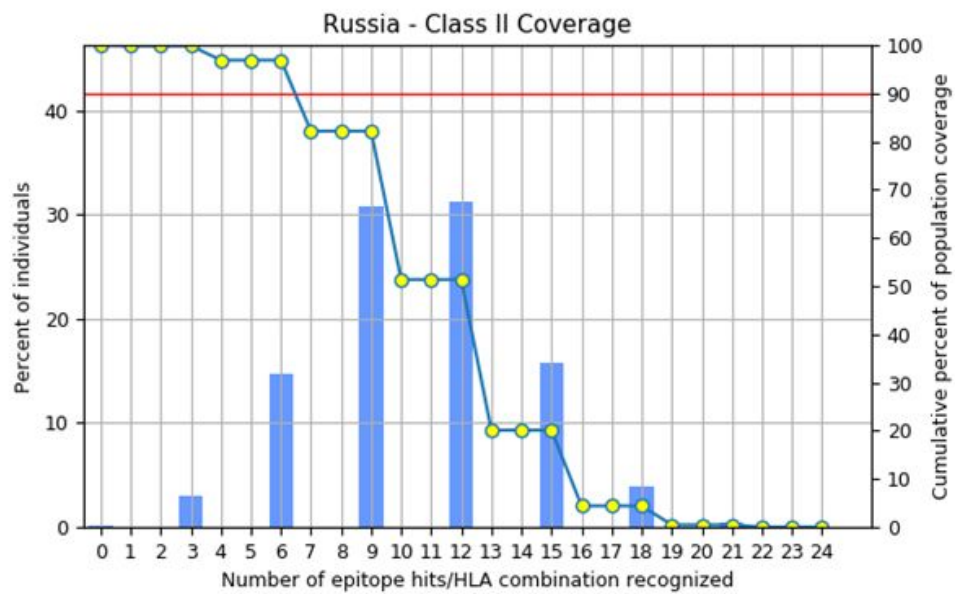
ILLYWVVYTTTRTQLPPAYYYLLWYIVGVYYPDKVFFFWLLVIYVLFFFLPFFSNVTWWWLLWIY
WWLLLIVNNATNVVVKVVVWYLLIFLCNDPFLGVYYLWWILVEGKQGNFKNLREFVFKNIDG
YFLVFYLLWWVWLLQELGKYEQYIKWPWYIWWWILVYVLLMSFPQSAPHGVVFLLLLYVLWAR
SVASQSIHAYYYWLLLVIWAHFPREGVFVSNGTHWWWWLVLYILLAGCLIGAELVNNSSYYLWV
VWYCGPKKSTNLVWVLFYFLLNFYEPQIITDNTFFFFLIYLLDLGDISGINASVNIQKLLIYFFVS
NNSIAIPTNFTIIILYWFFLRSFIEDLLFNKVVVLYVILLWGINITRFQTIILHRSYLYYYKRSFIEDLL
FNKVTLLIIYVVVYVLLLLFQFCNDPFLGVFFHK

7. Полученные эпитопы были пропущены через BLAST среди семейства коронавирусов, исключая SARS-CoV-2. Таким образом мы увидели, что предложенные эпитопы выравниваются на протеомы коронавируса летучей мыши, панголина и свиньи.
8. Полученные эпитопы в сочетании с выбранными аллелями были проанализированы на процент покрытия популяции жителей России с помощью сервиса Population Coverage на платформе IEDB Analysis Resource. Получилось, что аллелями MHC-I было покрыто 84,46% популяции, а аллелями MHC-II - 99,87%. Поскольку количество эффективных эпитопов для MHC-I сильно перевешивает количество эффективных эпитопов для MHC-II и, учитывая, что вероятность связывания эпитопов с Т-клетками далеко не 100%, потенциальная эффективность данной модели будет существенно ниже.

MHC class	Coverage	Average hit	PC90
I	87.46%	8.99	4.78



population/area	Class II		
	coverage ^a	average_hit ^b	pc90 ^c
Russia	99.87%	10.66	7.4



9. Для того, чтобы активировать еще и В-клетки с помощью нашей вакцины, было принято решения создать некоторый эпитоп и для них. Мы решили сделать это на основе Е белка. Для того, чтобы Е белок экспортировался из клетки, на основе чего можно было бы обучить В-клетки распознавать антиген, к N-концу белка должна быть подставлена сигнальная последовательность, которая будет направлять синтез белка в ЭПР. В самом ЭПР данная сигнальная последовательность будет обрезаться. Данная последовательность была найдена в книге “Molecular Biology of The Cell”, Bruce Alberts: MMSFSFVSLLLVGILFWATEQLTKCEVFQ. Согласно нашей идее и требованиям задания, в процессе транскрипции получающийся пептид должен заякориться в мембране ЭР: Е белок и сигнальная последовательность уйдут на экспорт, а остальная часть пептида, отвечающая эпитопам, останется в цитоплазме. Такой трансмембранный домен был заимствован у природы - белок 3BJ4 является трансмембранным белком и содержит альфа-спирали, пронизывающие мембрану. Последовательность этой альфа-спирали мы и взяли в качестве последовательности, которая обеспечит заякоривание полипептида в мембране ЭР. Далее идет последовательность эпитопов, разделенная участками, богатыми ароматическими и алифатическими аминокислотами. Для того, чтобы наша полипептидная цепочка в итоге разделилась на несколько частей и частично осталась в цитоплазме, а частично ушла на экспорт, необходимы сайты разрезания этой цепочки на стыке модифицированного Е белка и заякоривающей последовательности и на стыке заякоривающей последовательности и эпитопов. Нами были найдены протеазы, которые потенциально могут находиться в пределах ЭР и в цитоплазматической мембране. Источником этой информации послужили следующие ссылки:

<https://www.enzyme-database.org/query.php?name=Site-1+protease> (ссылка на ЭР протеазу) и <https://www.enzyme-database.org/query.php?ec=3.4.22.34>

(ссылка на цитозольную протеазу). На основании специфичности данных протеаз были сконструированы сайты разрезания этими протеазами нашей цепочки. Конечно, этот вариант является “сырым” и не очень тонко продуманным, однако в пределах ограничений задания, мы не нашли другого способа “расширить” действие иммунной системы и на другие типы иммунных

клеток. Предлагаем Вашему вниманию кусочек рабочего момента нашей команды ;)



10. Далее мы решили попробовать создать вектор на основе генома AAV (https://www.ncbi.nlm.nih.gov/nuccore/NC_040671).

Нами была взята последовательность генома, после чего в нем были определены сайты рестрикции. Среди сайтов были отобраны 2, которые бы соответствовали следующим требованиям:

1. В геноме AAV каждый из них повторяется только 1 раз
2. В ДНК-последовательности, кодирующей выбранный нами пептид, данные сайты не встречаются

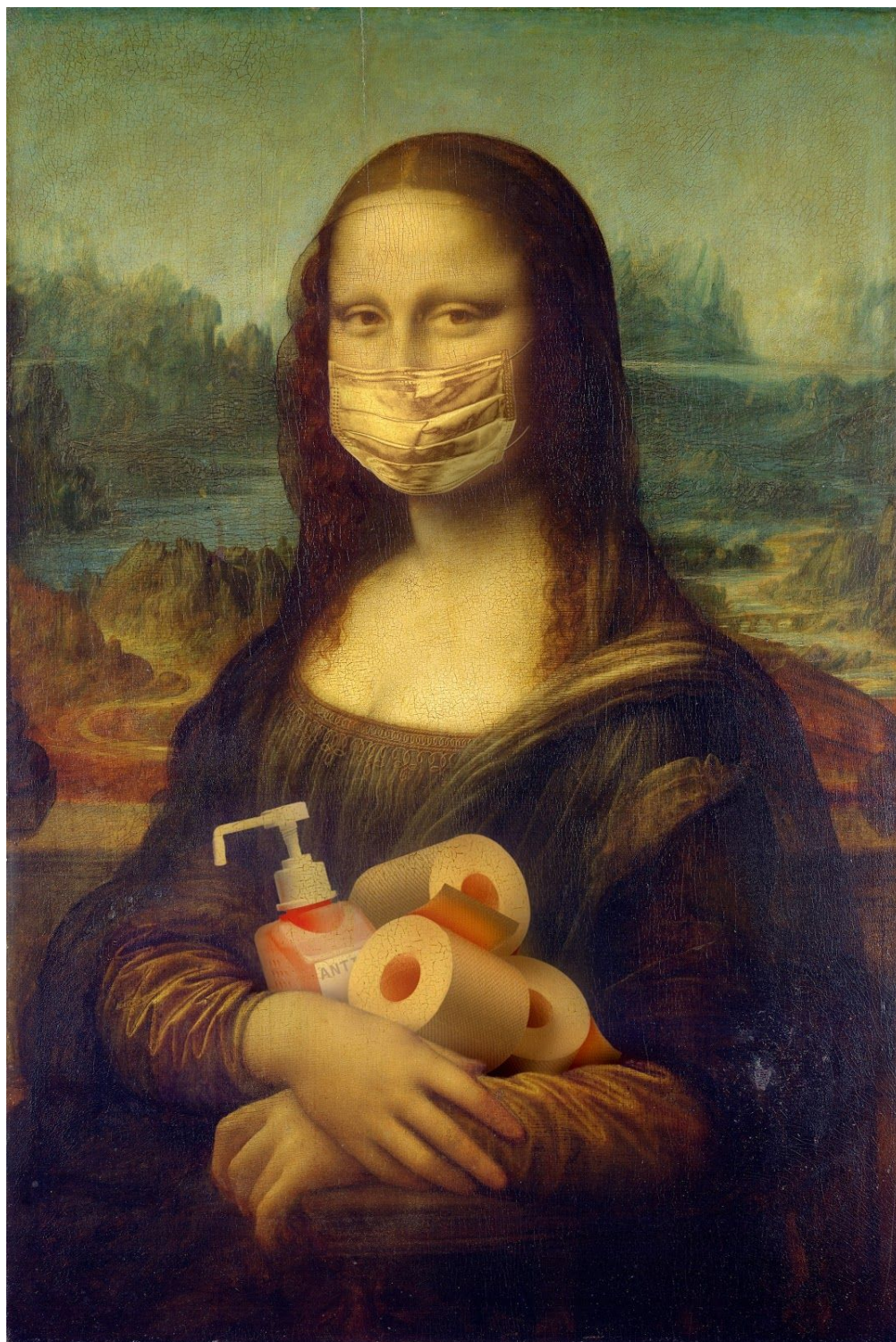
По таким критериям были отобраны сайты AatII и EcoRV, а участок между ними – вырезан.



Полученную ранее пептидную последовательность мы перевели в последовательность ДНК с помощью

https://www.bioinformatics.org/sms2/rev_trans.html

Для удобства амплификации и повышения точности встраивания созданной нами последовательности в вектор на нее были «навешаны» праймеры в виде сайтов AatII и EcoR5 (комплементарный ему) на 5' и 3' концы, соответственно. Поскольку вставленный нами фрагмент стоит под вирусными регуляторными генетическими элементами, то его экспрессия должна, как нам кажется, идти.



Извините за внимание!