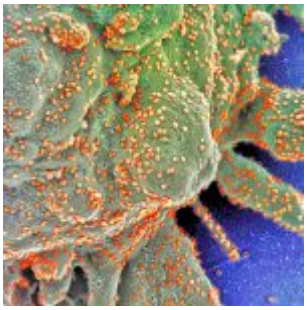
Reverse engineering kodu źródłowego szczepionki BioNTech / Pfizer SARS-CoV-2



Witaj! W tym poście będziemy przyglądali się znak po znaku kodowi źródłowemu szczepionki BioNTech / Pfizer SARS-CoV-2 mRNA.

Poniższy materiał jest tłumaczeniem z języka angielskiego analizy **Berta Huberta** pt. „Reverse Engineering the source code of the BioNTech/Pfizer SARS-CoV-2 Vaccine”. Publikowany jest za zgodą autora, jako jeden z oficjalnych przekładów.

> *Chciałbym podziękować licznej grupie osób, które spędziły czas na przeglądaniu tego artykułu pod względem czytelności i poprawności. Wszystkie błędy pozostają moimi, lecz pragnę szybko o nich usłyszeć czytając e-maile wysyłane pod adres bert@hubertnet.nl lub wiadomości Twittera kierowane do @PowerDNS_Bert.*

Tytułowe słowa mogą brzmieć nieco zgrzytliwie – szczepionka to płyn, który wstrzykuje się w ramię. Jak możemy mówić o kodzie źródłowym?

To dobre pytanie, więc zacznijmy od małej części samego kodu źródłowego szczepionki **BioNTechu / Pfizera**, znanej również jako **BNT162b2**, a także jako **Tozinameran** bądź **Comirnaty**.

WHO 9/2020
International Nonproprietary Names Programme

Sequence / Séquence / Secuencia

GAGAAΨAAAC	ΨAGΨAΨΨCΨΨ	CΨGGΨCCCCA	CAGACΨCAGA	GAGAACCCGC	50
CACCAΨGΨΨC	GΨGΨΨCCΨGG	ΨGCΨGCΨGCC	ΨCΨGGΨGΨCC	AGCCAGΨGΨG	100
ΨGAACCΨGAC	CACCAGAACA	CAGCΨGCCΨC	CAGCCΨACAC	CAACAGCΨΨΨ	150
ACCAGAGGCG	ΨGΨACΨACCC	CGACAAGGΨG	ΨΨCAGAΨCCA	GCGΨGCΨGCA	200
CΨCΨACCCAG	GACCΨGΨΨCC	ΨGCCΨΨΨCΨΨ	CAGCAACGΨG	ACCΨGGΨΨCC	250
ACGCCAΨCCA	CGΨGΨCCGGC	ACCAAΨGGCA	CCAAGAGAΨΨ	CGACAACCCC	300
GΨGCΨGCCCΨ	ΨCAACGACGG	GGΨGΨACΨΨΨ	GCCAGCACCG	AGAAGΨCCAA	350
CAΨCAΨCAGA	GGCΨGGAΨCΨ	ΨCGGCACCAC	ACΨGGACAGC	AAGACCCAGA	400
GCCΨGCΨGAΨ	CGΨGAACAAC	GCCACCAACG	ΨGGΨCAΨCAA	AGΨGΨGCGAG	450
ΨΨCCAGΨΨCΨ	GCAACGACCC	CΨΨCCΨGGGC	GΨCΨACΨACC	ACAAGAACAA	500

Pierwsze 500 znaków sekwencji mRNA szczepionki BNT162b2

Szczepionka BNT162b mRNA ma w sercu ten cyfrowy kod. Ma on długość **4284** znaki, więc zmieściłby się w kilku tweetach. Na samym początku procesu produkcji szczepionki ktoś wysłał ten kod do drukarki DNA (tak), która następnie przekształciła bajty na dysku w prawdziwe cząsteczki DNA.



Drukarka Codex DNA BioXp 3200

Z maszyny takiej wychodzą małe ilości DNA, które po długim biologicznym i chemicznym przetwarzaniu staje się RNA (o czym później) w fiolce ze szczepionką. Okazuje się, że dawka 30 mikrogramów naprawdę zawiera 30 mikrogramów RNA. Poza tym istnieje sprytny system opakowywania w lipidy (tłuszcze),

który dostarcza mRNA do naszych komórek.

RNA to ulotna wersja DNA, będąca jej „pamięcią podręczną". DNA jest w biologii napędem typu flash. Bardzo trwałe, redundantne wewnętrznie i niezawodne. Jednak podobnie do komputerów, kod nie jest wykonywany bezpośrednio z dysku flash; zanim cokolwiek się wydarzy zostanie skopiowany do szybszego, bardziej uniwersalnego, ale o wiele bardziej delikatnego systemu.

W przypadku komputerów będzie to pamięć RAM, a w biologii RNA. Podobieństwo jest uderzające. W przeciwieństwie do pamięci flash, RAM ulega degradacji bardzo szybko, chyba że miłościwie będziemy jej doglądać. Powód, dla którego szczepionka mRNA firmy Pfizer / BioNTech musi być przechowywana w najgłębszych z głębokich zamrażarek jest taki sam: RNA to delikatny kwiatek.

Każdy znak RNA waży około $0{,}53 \cdot 10^{-21}$ grama, co oznacza, że w pojedynczej 30-mikrogramowej dawce szczepionki znajduje się $6 \cdot 10^{16}$ znaków. Wyrażone w bajtach będzie to około 25 petabajtów, jednak trzeba dodać, że składa się z około 2000 miliardów powtórzeń tych samych 4284 znaków. Rzeczywista zawartość informacyjna szczepionki to nieco ponad kilobajt. Sam SARS-CoV-2 „waży" około 7,5 kilobajta.

Najkrótszy bit tła

DNA to cyfrowy kod. W przeciwieństwie do komputerów, które używają 0 i 1, życie wykorzystuje A, C, G oraz U/T (nukleotydy, nukleozydy bądź zasady).

W komputerach przechowujemy 0 i 1 jako (nie-)obecność ładunku elektrycznego, zamknięcie obwodu prądu, przejście dipolu magnetycznego, różnicę potencjałów, modulację sygnału, a także jako zmianę refleksyjności. Krótko mówiąc, 0 i 1 nie są jakimś abstrakcyjnym pojęciem – „żyją" jako elektrony, a także

w wielu innych, fizycznych postaciach.

W naturze A, C, G i U/T to cząsteczki przechowywane jako łańcuchy wewnątrz DNA (lub RNA).

W komputerach grupujemy 8 bitów w bajt, zaś bajt jest typową jednostką przetwarzanych danych.

Natura grupuje 3 nukleotydy w kodon i ów kodon jest typową jednostką przetwarzania. Kodon zawiera 6 bitów informacji. (2 bity na każdy znak DNA, 3 znaki = 6 bitów. Oznacza to 2^6 = 64 różne wartości kodonu).

Do tej pory jest całkiem cyfrowo. W razie wątpliwości przejdź do dokumentu WHO z kodem cyfrowym, aby przekonać się na własne oczy.

> *Więcej informacji można znaleźć tutaj – odnośnik ten (pt. „Czym jest życie”) może pomóc w zrozumieniu pozostałej części strony. Jeżeli lubisz wideo, mogę zająć ci dwie godziny.*

Co więc robi ten kod?

Ideą szczepionki jest nauczenie naszego układu odpornościowego sposobu walki z patogenem bez sprawiania, że naprawdę się rozchorujemy. W przeszłości dokonywano tego przez wstrzykiwanie osłabionego lub unieczynnionego (atenuowanego) wirusa wraz z adiuwantem, aby „przestraszyć” nasz układ odpornościowy, wywołując jego działanie. Była to zdecydowanie analogowa technika, bazująca na użyciu miliardów jaj (lub owadów). Wymagało to również sporego szczęścia i mnóstwa czasu. Czasami używano też innego (niespokrewnionego) wirusa.

Szczepionka mRNA osiąga ten sam efekt (uczy nasz układ odpornościowy), ale w sposób podobny do lasera. Mam na myśli oba znaczenia – bardzo wąskie, ale też bardzo mocne działanie.

Oto jak to działa: Zastrzyk zawiera ulotny materiał

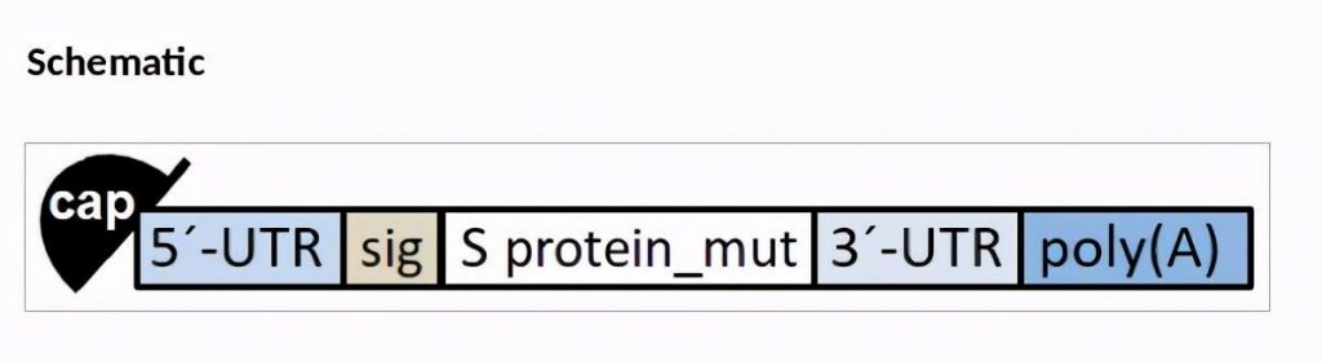
genetyczny, który opisuje słynną białkową wypustkę (ang. spike) wirusa SARS-CoV-2. Dzięki przebiegłym chemicznym sposobom szczepionce udaje się przenieść ten materiał genetyczny do naszych niektórych komórek.

Następnie komórki posłusznie zaczynają wytwarzać białko wypustki SARS-CoV-2 w wystarczająco dużych ilościach, żeby nasz układ odpornościowy zaczął działać. W konfrontacji z białkami wypustki i charakterystycznymi oznakami świadczącymi o tym, że komórki zostały przejęte, nasz układ odpornościowy rozwija potężną odpowiedź przeciwko wielu aspektom białkowej wypustki ORAZ procesowi jej produkcji.

I to właśnie prowadzi nas do szczepionki skutecznej w 95%.

Kod źródłowy!

Zacznijmy od samego początku, od końca byłoby na wspak. Dokument WHO zawiera taki pomocny obraz:



To rodzaj spisu treści. Zaczniemy od „czapeczki” (ang. cap), przedstawionej właściwie jako kapelusik.

Podobnie jak nie można zwyczajnie wrzucić kodów operacyjnych do pliku na komputerze i go uruchomić, również biologiczny system operacyjny wymaga nagłówków, ma konsolidatory i odpowiednie konwencje wywoływania.

Kod szczepionki zaczyna się od następujących dwóch nukleotydów:

GA

Można to porównać do praktycznie każdego pliku wykonywalnego systemów DOS i Windows zaczynającego się od MZ, albo do skryptów systemów typu Unix zaczynających się od #!. Zarówno w „systemie życia”, jak i w systemie operacyjnym te dwa znaki nie są w żaden sposób wykonywane. Muszą tam jednak być, bo inaczej nic się nie stanie.

Czapeczka mRNA pełni wiele funkcji. Po pierwsze oznacza kod pochodzący z jądra komórkowego. W naszym przypadku oczywiście tak nie jest, bo nasz kod pochodzi ze szczepienia, lecz nie musimy tego mówić komórce. „Kapelusz” sprawia, że kod wygląda w porządku, co chroni go przed zniszczeniem.

Początkowe dwa nukleotydy GA również nieznacznie różnią się chemicznie od reszty RNA. W tym sensie GA ma funkcję sygnalizowania pozapasmowego.

„Rejon 5 prim niepodlegający translacji”

Mamy tu pewien żargon. Cząsteczki RNA można odczytywać wyłącznie w jednym kierunku. Część, w której zaczyna się odczyt, myląco nazywano 5′ lub „pięć prim”. Odczyt kończy się przy 3′ (lub „trzy prim”).

Życie składa się z białek (lub rzeczy zbudowanych na bazie białek). Białka te opisane są w RNA. Kiedy RNA zostaje przekształcone w białka, nazywamy to **translacją**.

Mamy tu region niepodlegający translacji 5′ (5′ UTR), więc ten fragment nie znajdzie się w białku:

GAAΨAAACΨAGΨAΨΨCΨΨCΨGGΨCCCCACAGACΨCAGAGAGAACCCGCCACC

Napotykamy tu pierwszą niespodziankę. Zazwyczaj znaki RNA to A, C, G i U. Znak U jest też znany jako T w DNA. Tutaj jednak znajdujemy Ψ. Co się dzieje?

To jeden z wyjątkowo sprytnych fragmentów szczepionki. W naszym ciele działa potężny („oryginalny”) system antywirusowy. Z tego powodu komórki są wybitnie

nieentuzjastycznie nastawione do obcego RNA i bardzo starają się je zniszczyć, zanim cokolwiek uczyni.

Stanowi to pewnego rodzaju problem dla naszej szczepionki – musi ona przedrzeć się przez układ odpornościowy. W trakcie wielu lat eksperymentowania odkryto, że gdy U w RNA zostanie zastąpiona przez nieznacznie zmodyfikowaną cząsteczkę, nasz układ odpornościowy traci zainteresowanie takim tworem. Naprawdę.

Tak więc w szczepionce marki BioNTech / Pfizer każde wystąpienie U zostało zastąpione 1-metylo-3′-pseudourydyną oznaczoną Ψ. Naprawdę przebiegłe jest to, że chociaż wymiana na Ψ łagodzi (uspokaja) nasz układ odpornościowy, będzie zaakceptowana jako normalna U przez odpowiednie części komórki.

W bezpieczeństwie komputerowym znamy również taką sztuczkę – czasem można przesłać nieco uszkodzoną wersję wiadomości, która zmyli zapory sieciowe i narzędzia zabezpieczające, ale zostanie zaakceptowana przez serwery zaplecza – a te mogą następnie zostać zhackowane.

Podobnie jak w przypadku innych, fundamentalnych badań naukowych, z których obecnie czerpiemy korzyści, odkrywcy tej techniki musieli walczyć, żeby ich praca była finansowana i zaakceptowana. Powinniśmy być im wszyscy bardzo wdzięczni i jestem pewien, że w swoim czasie pojawią się również nagrody Nobla.

> *Wielu ludzi pytało, czy wirusy również mogą skorzystać z techniki przemycania z użyciem Ψ, aby pokonać nasz system odpornościowy. W skrócie: jest to niezmiernie mało prawdopodobne. Życie zwyczajnie nie ma mechanizmu, aby budować 1-metylo-3′-pseudourydynę. Wirusy polegają na mechanizmach życia, żeby się replikować, a tego typu zdolności po prostu tam nie ma. Szczepionki mRNA szybko ulegają degradacji w ludzkim ciele i nie ma możliwości,*

żeby Ψ-modyfikowane RNA replikowało się z wciąż obecną Ψ. Warto też poczytać: „Nie, naprawdę, szczepionki mRNA nie zmienią twojego DNA”.

OK, wracając do 5′ UTR. Co robi te 51 znaków? Podobnie jak wszystko w naturze, prawie nic nie ma jednej, wyraźnej funkcji.

Kiedy nasze komórki muszą dokonać translacji RNA na białka, odbywa się to z użyciem maszyny zwanej rybosomem. Rybosom jest jak drukarka 3D dla białek. Zjada nić RNA i na tej podstawie emituje łańcuch aminokwasów, z których następnie składane jest białko.

Translacja białka

Oto, co widzimy powyżej. Czarna wstęga na dole to RNA. Wstęga pojawiająca się w zielonym fragmencie to formowane białko. Wlatujące i wylatujące obiekty to aminokwasy i pomagające im w dopasowywaniu się do RNA adaptory.

Rybosom musi fizycznie „usiąść” na nitce RNA, aby to zadziałało. Gdy już się do niej przyczepi, może rozpocząć formowanie białek, bazując na dalej konsumowanym RNA. Na tej podstawie łatwo sobie wyobrazić, że części, na których rybosom spoczywa na początku nie mogą być odczytane. To tylko jedna z funkcji UTR: strefa lądowania dla rybosomów. UTR umożliwia „wprowadzenie”.

Oprócz tego UTR zawiera też metadane: kiedy powinno dojść do translacji? jak powinna być duża? W przypadku szczepionki zastosowano najbardziej „natychmiastowy” UTR, jaki tylko można było znaleźć, wzięty z genu alfa-globiny. Gen ten jest znany z intensywnego wytwarzania dużej liczby białek. W poprzednich latach naukowcy znaleźli już sposoby na jeszcze większe zoptymalizowanie tego konkretnego UTR (zgodnie z dokumentem WHO), więc nie jest to zwyczajny UTR bazujący na alfa-globinie. Jest lepszy.

Peptyd sygnałowy glikoproteiny S

Jak wspomniano, celem szczepionki jest spowodowanie, żeby komórka wyprodukowała obfite ilości białka wypustki SARS-CoV-2. Do tego momentu w kodzie źródłowym szczepionki najczęściej spotykaliśmy się z metadanymi i „konwencją wywoływania". Teraz jednak wchodzimy na prawdziwe terytorium białek wirusowych.

Wciąż mamy jednak do przejścia jedną warstwę metadanych. Gdy rybosom (ze wspaniałej animacji wcześniej) utworzy białko, to musi ono gdzieś się udać. Zostanie to zakodowane w peptydzie sygnałowym glikoproteiny S (wydłużona sekwencja liderowa).

Można to widzieć w ten sposób, że na początku białka znajduje się rodzaj etykiety adresowej – zakodowanej jako część samego białka. W tym konkretnym przypadku peptyd sygnałowy mówi, że białko to powinno opuścić komórkę przez endoplazmatyczne retikulum. Nawet żargon ze „Star Treka" nie jest równie wyszukany!

Białko sygnałowe nie jest zbyt długie, lecz kiedy spojrzymy na jego kod, zauważymy, że istnieją różnice między RNA wirusa i szczepionki:

(Zwróć uwagę, że dla celów porównawczych zamieniłem fantazyjny, zmodyfikowany Ψ na zwyczajny RNA U).

```
                3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3
      Wirus: AUG UUU GUU UUU CUU GUU UUA UUG CCA CUA GUC UCU AGU CAG UGU GUU
Szczepionka: AUG UUC GUG UUC CUG GUG CUG CUG CCU CUG GUG UCC AGC CAG UGU GUU
                 !   !   !   !   ! ! ! !     !   !   !   !   !
```

Co się więc dzieje? Nieprzypadkowo przedstawiłem RNA w grupach po 3 litery. Trzy znaki RNA tworzą kodon, a każdy kodon koduje określony aminokwas. Peptyd sygnałowy w szczepionce składa się z dokładnie takich samych aminokwasów, co w samym wirusie.

Więc czemu RNA jest inne?

Istnieją 4^3 = 64 różne kodony, ponieważ są 4 znaki RNA, a trzy z nich są w kodonie. Jednak istnieje tylko 20 różnych aminokwasów. Oznacza to, że wiele kodonów koduje ten sam aminokwas.

Życie używa następującej, niemal uniwersalnej tabeli do mapowania kodonów RNA na aminokwasy:

Kodon	Aminokwas	Kodon	Aminokwas	Kodon	Aminokwas	Kodon	Aminokwas
UUU	fenyloalanina	UCU	seryna	UAU	tyrozyna	UGU	cysteina
UUC	fenyloalanina	UCC	seryna	UAC	tyrozyna	UGC	cysteina
UUA	leucyna	UCA	seryna	UAA	**Ochre** (*Stop*)	UGA2	**Opal** (*Stop*)
UUG	leucyna	UCG	seryna	UAG	**Amber** (*Stop*)	UGG	tryptofan
CUU	leucyna	CCU	prolina	CAU	histydyna	CGU	arginina
CUC	leucyna	CCC	prolina	CAC	histydyna	CGC	arginina
CUA	leucyna	CCA	prolina	CAA	glutamina	CGA	arginina
CUG	leucyna	CCG	prolina	CAG	glutamina	CGG	arginina
AUU	izoleucyna	ACU	treonina	AAU	asparagina	AGU	seryna
AUC	izoleucyna	ACC	treonina	AAC	asparagina	AGC	seryna
AUA	izoleucyna	ACA	treonina	AAA	lizyna	AGA	arginina
AUG	metionina (*Start*)	ACG	treonina	AAG	lizyna	AGG	arginina
GUU	walina	GCU	alanina	GAU	asparaginian	GGU	glicyna
GUC	walina	GCC	alanina	GAC	asparaginian	GGC	glicyna
GUA	walina	GCA	alanina	GAA	glutaminian	GGA	glicyna
GUG1	walina	GCG	alanina	GAG	glutaminian	GGG	glicyna

Tabela kodonów mRNA
W tabeli tej widzimy, że wszystkie modyfikacje w szczepionce (UUU → UUC) są synonimiczne. Kod RNA szczepionki jest inny, ale powstają takie same aminokwasy i to samo białko.

Jeżeli przyjrzymy się uważnie, zobaczymy, że większość zmian zachodzi na trzeciej pozycji kodonu, oznaczonej wyżej przez 3. Gdy sprawdzimy uniwersalną tablicę kodonów, zauważymy, że ta trzecia pozycja często faktycznie nie ma znaczenia

w kontekście tego, jaki aminokwas będzie wytworzony.

Zmiany są więc synonimiczne, ale w takim razie dlaczego w ogóle tam są? Uważnie się przyglądając, widzimy, że wszystkie oprócz jednej prowadzą do większej liczby C i G.

Czemu więc mielibyśmy zrobić coś takiego? Jak wspomniano powyżej, nasz układ odpornościowy bardzo słabo patrzy na „egzogenny" RNA, pochodzący spoza komórki. Aby uniknąć wykrycia, U w RNA został zastąpiony przez Ψ.

Okazuje się jednak, że kod RNA o większej liczbie G i C jest też wydajniej przekształcany w białka, a udało się to osiągnąć w szczepionce RNA, zastępując wiele znaków G i C, gdziekolwiek było to możliwe.

> *Jestem nieco zafascynowany jedną zmianą, która nie doprowadziła do dodatkowego C lub G, czyli modyfikacją CCA CCU. Jeżeli ktoś zna powód, niech da mi znać! Proszę zauważyć, że jestem świadom faktu, iż niektóre kodony występują w ludzkim genomie częściej niż inne, ale wyczytałem też, że nie wpływa to zbytnio na prędkość translacji.*

Prawdziwa wypustka białkowa

Kolejne 3777 znaków szczepionkowego RNA zostało w podobny sposób „zoptymalizowane pod względem kodonów", aby dodać wiele C i G. Z uwagi na obszar analizy nie będę tutaj przedstawiać całego kodu, ale przybliżymy jeden wyjątkowo szczególny fragment. Będzie nim część, która sprawia, że to wszystko działa; część, która faktycznie pomoże nam wrócić do normalnego życia:

```
                  *   *
              L   D   K   V   E   A   E   V   Q   I   D   R   L   I   T   G
      Wirus: CUU GAC AAA GUU GAG GCU GAA GUG CAA AUU GAU AGG UUG AUC ACA GGC
Szczepionka: CUG GAC CCU CCU GAG GCC GAG GUG CAG AUC GAC AGA CUG AUC ACA GGC
              L   D   P   P   E   A   E   V   Q   I   D   R   L   I   T   G
               !     !!! !!       !   !       !   !   !   ! !
```

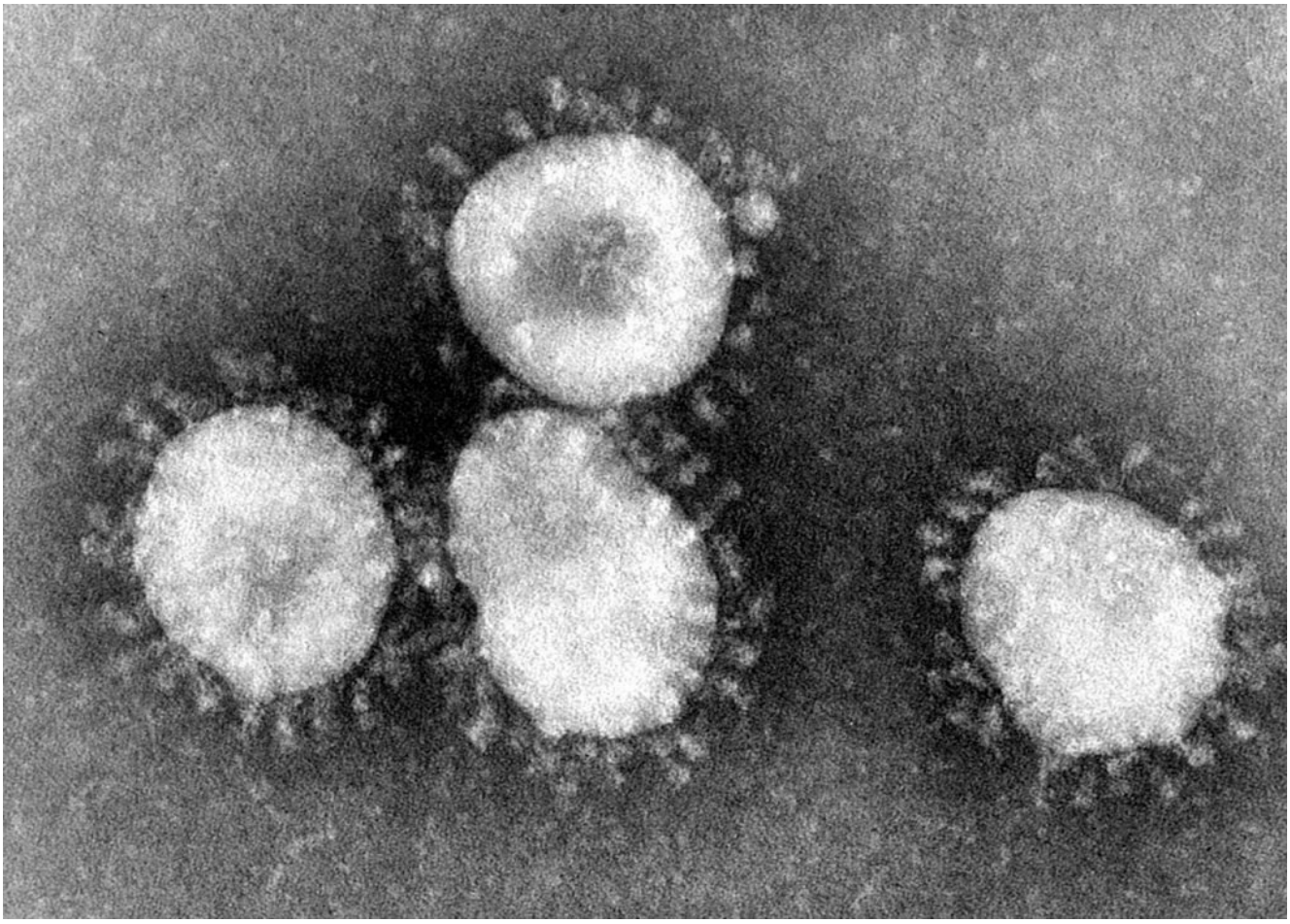
Widzimy tu zwykłe, synonimiczne zmiany RNA. Na przykład w pierwszym kodonie możemy zaobserwować, że CUU jest zamieniane na CUG. To dodaje do szczepionki kolejną G, która, jak wiemy, pomaga zwiększyć produkcję białka. Zarówno CUU, jak i CUG kodują aminokwas L (leucynę), więc w samym białku nic się nie zmieniło.

Kiedy porównamy całe białko wypustki obecne w szczepionce, zauważymy, że wszystkie zmiany są analogiczne… z wyjątkiem dwóch, i właśnie to w tym miejscu widzimy.

Trzeci i czwarty kodon powyżej reprezentują praktyczne różnice. Aminokwasy K i V są tam zastępowane przez P (prolinę). Dla K wymagało to trzech zmian (!!!), a dla V jedynie dwóch (!!).

Okazuje się, że te dwie zmiany ogromnie zwiększają skuteczność szczepionki.

Co się więc dzieje? Gdy spojrzysz na prawdziwą cząsteczkę SARS-CoV-2, ujrzysz białka wypustki właśnie jako… wypustki:



Cząsteczki wirusa SARS

Wypustki są osadzone na ciele wirusa (białko nukleokapsydowe). Chodzi jednak o to, że nasza szczepionka sprawia, iż generowane są tylko te „kolce", nieprzytwierdzone do żadnego ciała wirusa.

Okazuje się, że niezmodyfikowane białko wypustki samoczynnie układa się w całkiem inną strukturę. Wstrzyknięcie go jako szczepionki spowodowałoby, że nasz organizm rozwinąłby odporność… jednak tylko względem białka wypustki.

Prawdziwy SARS-CoV-2 pojawia się jednak z kolczastą wypustką. W takim przypadku szczepionka nie zadziałałaby zbyt dobrze.

Więc co robić? W 2017 roku opisano, jak umieszczenie podwójnej substytucji proliny we właściwym miejscu sprawia, że białka SARS-CoV-1 i MERS S przyjmują konformację (układ) „sprzed fuzji", nawet bez stawania się częścią całego wirusa. To działa, ponieważ prolina to bardzo sztywny aminokwas. Działa jak pewnego rodzaju szyna, stabilizując białko w stanie, który

musimy pokazać układowi odpornościowemu.

Osoby, które dokonały tego odkrycia, powinny nieustannie chodzić i przybijać sobie piątkę. Powinny z nich emanować nieznośne ilości samozadowolenia. I byłoby to w pełni zasłużone.

> *Aktualizacja! Skontaktowało się ze mną Laboratorium McLellana, a właściwie jedna z grup odpowiedzialnych za odkrycie kwestii proliny. Przekazali mi, że przybijanie piątek zostało nieco powściągnięte ze względu na trwającą pandemię, ale są zadowoleni ze swojego wkładu w szczepionki. Podkreślają też znaczenie wielu innych grup, pracowników i ochotników.*

Koniec białka, kolejne kroki

Gdy przewiniemy zapis reszty kodu źródłowego, napotkamy pewne małe modyfikacje na końcu wypustki białkowej:

```
                V   L   K   G   V   K   L   H   Y   T   s
      Wirus: GUG CUC AAA GGA GUC AAA UUA CAU UAC ACA UAA
Szczepionka: GUG CUG AAG GGC GUG AAA CUG CAC UAC ACA UGA UGA
                V   L   K   G   V   K   L   H   Y   T   s   s
                    !   !   !   !     ! !   !         !
```

Na końcu białka znajdujemy kodon „stop”, oznaczony tu małą literą s. To grzeczny sposób powiedzenia, że białko powinno się w tym miejscu zakończyć. Oryginalny wirus używa kodonu zatrzymującego UAA, szczepionka zaś dwóch kodonów UGA, prawdopodobnie dla pewności.

Rejon 3′ niepodlegający translacji

Podobnie jak rybosom potrzebował wprowadzenia na końcu 5′, gdzie znaleźliśmy rejon niepodlegający translacji 5 prim, na końcu białka znajdujemy podobny konstrukt zwany 3′ UTR.

Można by pisać wiele o UTR 3′, lecz tu zacytuję, co mówi

Wikipedia: „Niepodlegający translacji rejon 3′ spełnia kluczową rolę w ekspresji genów, wpływając na lokalizację, stabilność, eksport i wydajność translacji mRNA […] **pomimo naszego obecnego zrozumienia różnych UTR 3′, stanowią one wciąż pewną tajemnicę**”.

Wiemy natomiast, że niektóre UTR 3′ są bardzo skuteczne w promowaniu ekspresji białek. Zgodnie z dokumentem WHO, UTR 3′ szczepionki firm BioNTech / Pfizer został wybrany z „końca aminowego (N-końca) wzmacniacza podziału (AES) mRNA i kodowanego mitochondrialnie rybosomalnego RNA 12S w celu ustabilizowania RNA i zwiększenia całkowitej ekspresji białka”. Wobec tego pragnę powiedzieć: dobra robota.



AAAAAAAAAAAAAAAAA koniec wszystkiego

Sam koniec mRNA jest poliadenylowany. To fantazyjny sposób powiedzenia, że kończy się na wielu AAAAAAAAAAAAAAAAAA. Wydaje się, że nawet mRNA ma już dosyć roku 2020.

mRNA może być używane wielokrotnie, ale gdy do tego dochodzi, traci też część A na końcu. Po wyczerpaniu A mRNA nie jest już funkcjonalne i zostaje odrzucone. W ten sposób ogon „wielo-A” chroni przed degradacją.

Przeprowadzono badania, aby dowiedzieć się, jaka jest optymalna liczba A na końcu dla szczepionek mRNA. W otwartej literaturze wyczytałem, że wartością maksymalną było około 120.

Szczepionka BNT162b2 kończy się tak:

```
 ****** ****
UAGCAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGCAUAU GACUAAAAAA AAAAAAAAAA
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA
```

Mamy 30 A, następnie „10 nukleotydów łączących” (GCAUAUGACU), zaś dalej kolejne 70 A.

Podejrzewam, że to, co tutaj widzimy, jest efektem dalszej, zastrzeżonej optymalizacji, aby jeszcze bardziej zwiększyć ekspresję białek.

Podsumowując

Znamy teraz dokładną zawartość mRNA szczepionki BNT162b2 i rozumiemy powody istnienia większości jej części:

- czapeczki (CAP), aby upewnić się, że RNA wygląda jak zwyczajne mRNA;
- znanego, udanego i zoptymalizowanego rejonu 5′ niepodlegającego translacji (UTR);
- białka sygnałowego zoptymalizowanego pod kątem kodonów do wysyłania wypustki we właściwe miejsce (w 100% skopiowanego z oryginalnego wirusa);
- wersji oryginalnej wypustki zoptymalizowanej pod kątem kodonów, z dwoma podstawieniami proliny, aby upewnić się, że białko pojawia się we właściwej postaci;
- znanego, udanego i zoptymalizowanego rejonu 3′ niepodlegającego translacji;
- nieco tajemniczego ogona wielo-A z niewyjaśnionym „łącznikiem”.

Optymalizacja kodonu dodaje dużo G i C do mRNA. W tym samym czasie użycie Ψ (1-metylo-3′-pseudourydyny) zamiast U pomaga zmylić nasz układ odpornościowy, więc mRNA pozostaje wystarczająco długo, abyśmy mogli faktycznie pomóc w jego wytrenowaniu.

Do dalszego przeczytania / oglądania

W roku **2017** prowadziłem dwugodzinną prezentację poświęconą DNA, którą możesz obejrzeć w tym miejscu. Podobnie jak ta strona, jest ona przeznaczona dla komputerowców.

Dodatkowo od **2001** prowadzę stronę „DNA dla programistów”.

Może ci się również spodobać to wprowadzenie do niesamowitego układu odpornościowego.

W końcu to zestawienie moich wspisów w blogu zawiera trochę materiałów związanych z DNA, SARS-CoV-2 i COVID-em.

Tłumaczenie: PW

Źródło: berthub.eu