

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

I Big Data ai tempi del Coronavirus

Rimini – 01/12/2021

Valentina Pellicioni

Studente di dottorato in Scienza e Cultura del Benessere
e degli Stili di Vita

Dipartimento di Scienze per la Qualità della Vita

Le malattie infettive si diffondono più rapidamente

- Il 31 dicembre 2019 sono stati segnalati all'ufficio cinese dell'OMS diversi casi di polmonite di eziologia sconosciuta a Wuhan.
- Il 9 gennaio 2020, il nuovo focolaio di coronavirus è stato dichiarato dall'OMS un'emergenza sanitaria pubblica di interesse internazionale.
- L'11 marzo 2020, con oltre 118.000 casi confermati e 4.291 decessi in 114 paesi, la malattia da coronavirus (COVID-19) è stata dichiarata pandemia
- La pandemia COVID-19 è ciò che viene definito Mass Casualty Incident (MCI).



Le malattie infettive si diffondono più rapidamente

Un modo in cui le epidemie di questa portata possono essere gestite è un approccio collaborativo:

- Accademici che si uniscono per la ricerca di misure preventive o curative
- industrie che uniscono le forze per fornire risorse adeguate e tempestive
- Organizzazioni senza scopo di lucro che soccorrono la popolazione vulnerabile
- Governi che assicurano la diffusione delle informazioni, la trasparenza delle loro azioni e la condivisione delle risorse.



L'incertezza e il poco tempo a disposizione ha spinto i governi dei vari paesi a usare metodi basati sui big data.

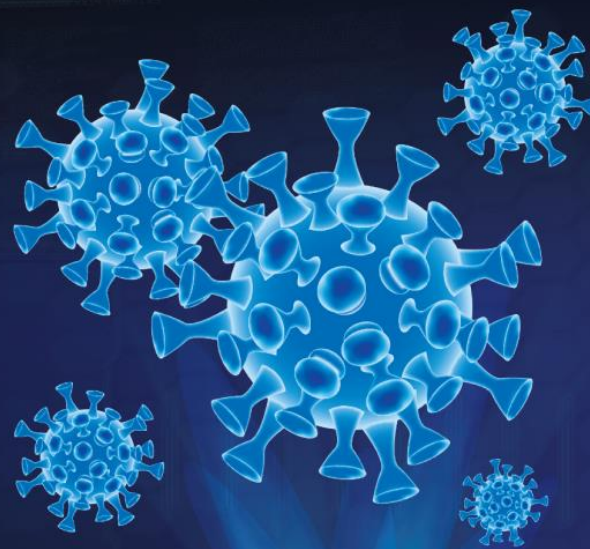
Utilizzare strumenti analitici sui big data, per esaminare l'evoluzione della malattia e comprendere lo stato globale della pandemia può essere definito **Pandemic analytics**.



Bluedot, l'algoritmo canadese che per primo ha previsto la pandemia

Il 9 gennaio, l'Organizzazione Mondiale della Sanità (OMS) ha notificato al mondo un'epidemia influenzale in Cina: un gruppo di casi di polmonite era stato segnalato a Wuhan, i centri statunitensi per il controllo e la prevenzione delle malattie avevano avuto la notizia il 6 gennaio. Ma una piattaforma canadese di monitoraggio della salute li ha battuti sul tempo, inviando la notizia dell'epidemia ai suoi clienti il **31 dicembre**.

How **BlueDot**
Predicted
Coronavirus Using
**Artificial
Intelligence (AI)?**



Bluedot, l'algoritmo canadese che per primo ha previsto la pandemia

Bluedot nasce nel 2014.

È un algoritmo che utilizza l'elaborazione del linguaggio naturale e le tecniche di machine learning per setacciare le notizie in 65 lingue, insieme ai dati delle compagnie aeree e le segnalazioni di focolai di malattie animali e vegetali e le proclamazioni ufficiali.

Una volta che il setacciamento automatico dei dati è completo, subentra l'analisi umana. Gli epidemiologi controllano che le conclusioni abbiano senso da un punto di vista scientifico, e poi viene inviato un rapporto al governo, ai funzionari della sanità pubblica di vari paesi, alle compagnie aeree e agli ospedali di prima linea dove potrebbero finire i pazienti infetti.



Google Flu Trends (GFT)

Ogni giorno milioni di persone usano Google per trovare informazioni che li riguardano.

Nel 2008 è nato Google Flu Trends.

Aveva l'obiettivo di prevedere le epidemie di influenza in base ai termini di ricerca delle persone.

google.org

Flu Trends



Così i big data possono anticipare nuovi focolai epidemici di Covid-19

Lo studio di Bologna

Uno studio del Rizzoli di Bologna e dell'Alma Mater dimostra che a una impennata delle ricerche sul web, con parole chiave come “febbre” e “tosse”, corrisponde un proporzionale aumento dei contagiati.

La ricerca si basa sui termini di ricerca di Google Trends, le parole chiave sono parole utilizzate nel linguaggio comune e si riferiscono ai sintomi dell'infezione da Coronavirus.

L'obiettivo era vedere se esiste una correlazione tra le ricerche sul web sui sintomi più comuni della malattia e l'andamento dell'epidemia.



Così i big data possono anticipare nuovi focolai epidemici di Covid-19

Lo studio di Pavia

Lo studio considera il periodo dall'1 gennaio all'11 maggio 2020 e si è basato sui trend nelle ricerche di Google e le domande sui sintomi da Covid-19 che le persone hanno fatto al chatbot di Pagine Mediche.

L'analisi su Google Trends è stata condotta usando molte parole chiave e frasi riconducibili ai sintomi del coronavirus: febbre, tosse, dispnea, difficoltà a respirare, diarrea, tachicardia e palpitazioni.

- Febbre + 33%
- Tosse + 28%
- Palpitazioni + 86%

Rispetto alla media dei quattro anni precedenti (2016-19),

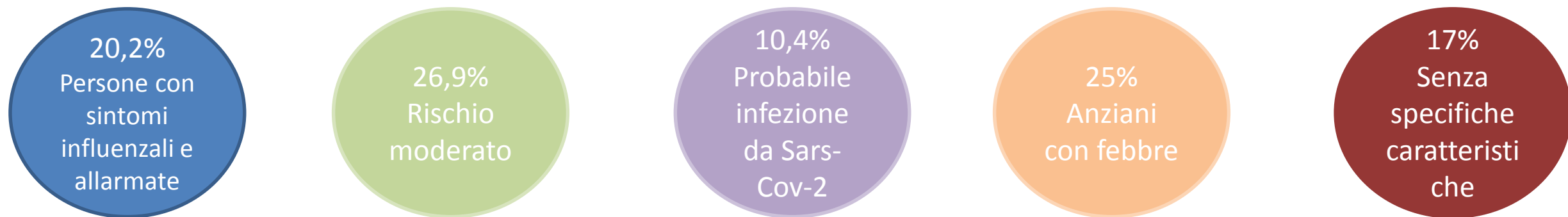


Così i big data possono anticipare nuovi focolai epidemici di Covid-19

Lo studio di Pavia

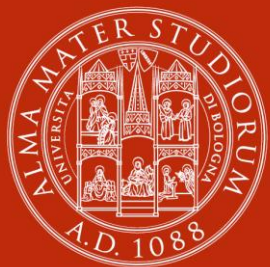
Nel periodo di studio si sono registrate 150.110 interazioni con il chatbot, di cui oltre la metà (65.912) dalla Lombardia.

Clusters:



Conclusioni: il caso della ricerca con la keyword dispnea su Google e di alcune interazioni su Pagine Mediche, sembrerebbe davvero anticipare di una settimana la curva dei contagi.





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Come le nazione dell'UE hanno impiegato i Big Data contro il Coronavirus

Rimini – 01/12/2021

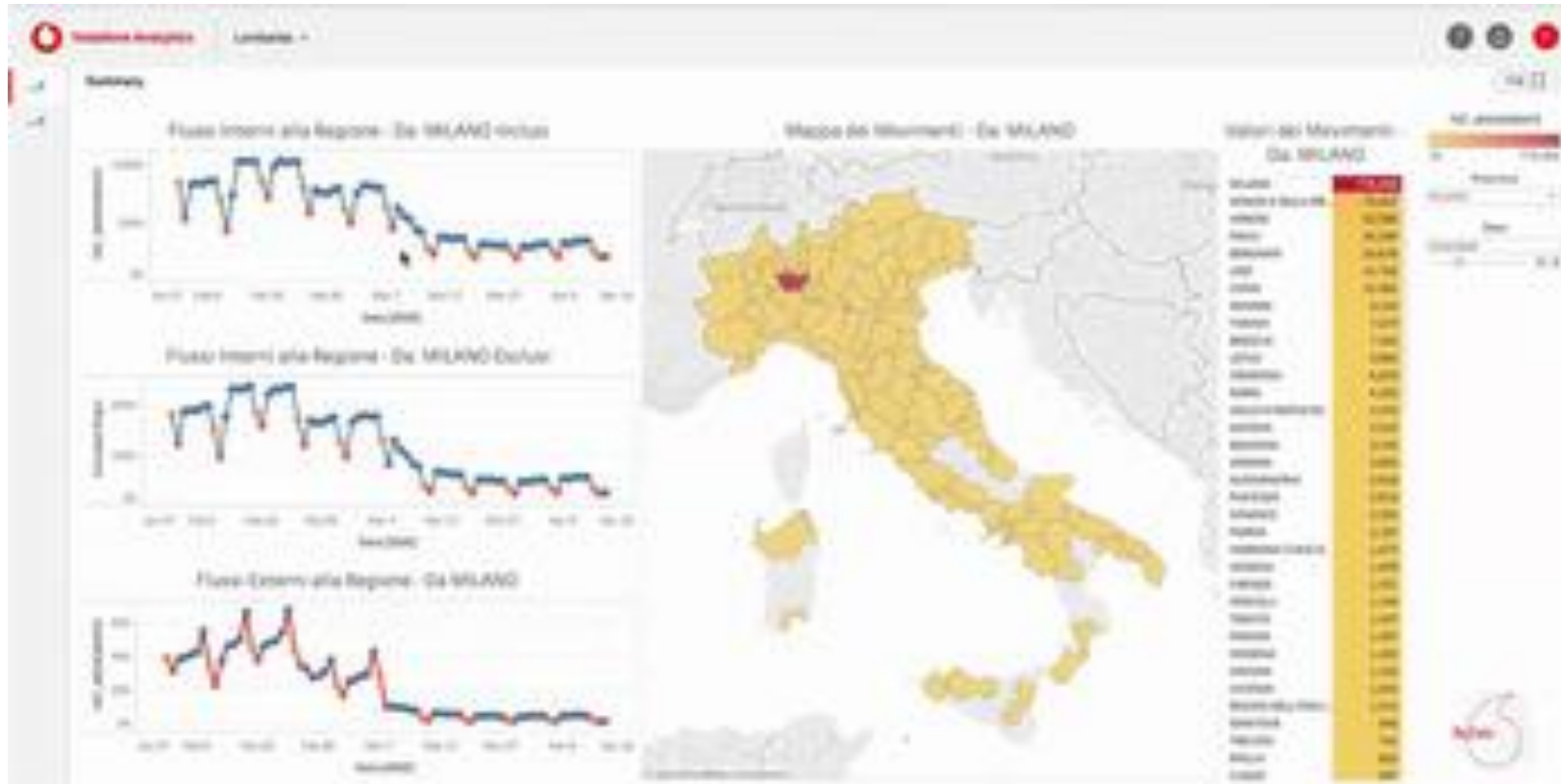
Valentina Pellicioni

Studente di dottorato in Scienza e Cultura del Benessere
e degli Stili di Vita

Dipartimento di Scienze per la Qualità della Vita

La heatmap di Vodafone

Vodafone ha creato una heat map per la regione Lombardia in Italia, permettendo alle autorità governative di tracciare meglio il flusso della popolazione.



INSIGHT CXR - La soluzione di analisi dei raggi X del torace basata sull'AI

L'algoritmo originale di Lunit INSIGHT CXR è in grado di rilevare 10 patologie diverse dalle immagini radiografiche del torace, tra cui la polmonite, il cancro ai polmoni, pneumotorace e altro.

Alla fine di marzo 2020 i produttori hanno rilasciato un algoritmo in versione aggiornata integrando la funzione che rileva i segni dell'infezione da Coronavirus.

Con questo algoritmo hanno fatto un'analisi retrospettiva.

Come lavora questo algoritmo?

Sull'input di un'immagine radiografica vengono generate le previsioni di classificazione dell'immagine. L'output è un punteggio di probabilità della presenza di lesioni che viene restituito con una heatmap.

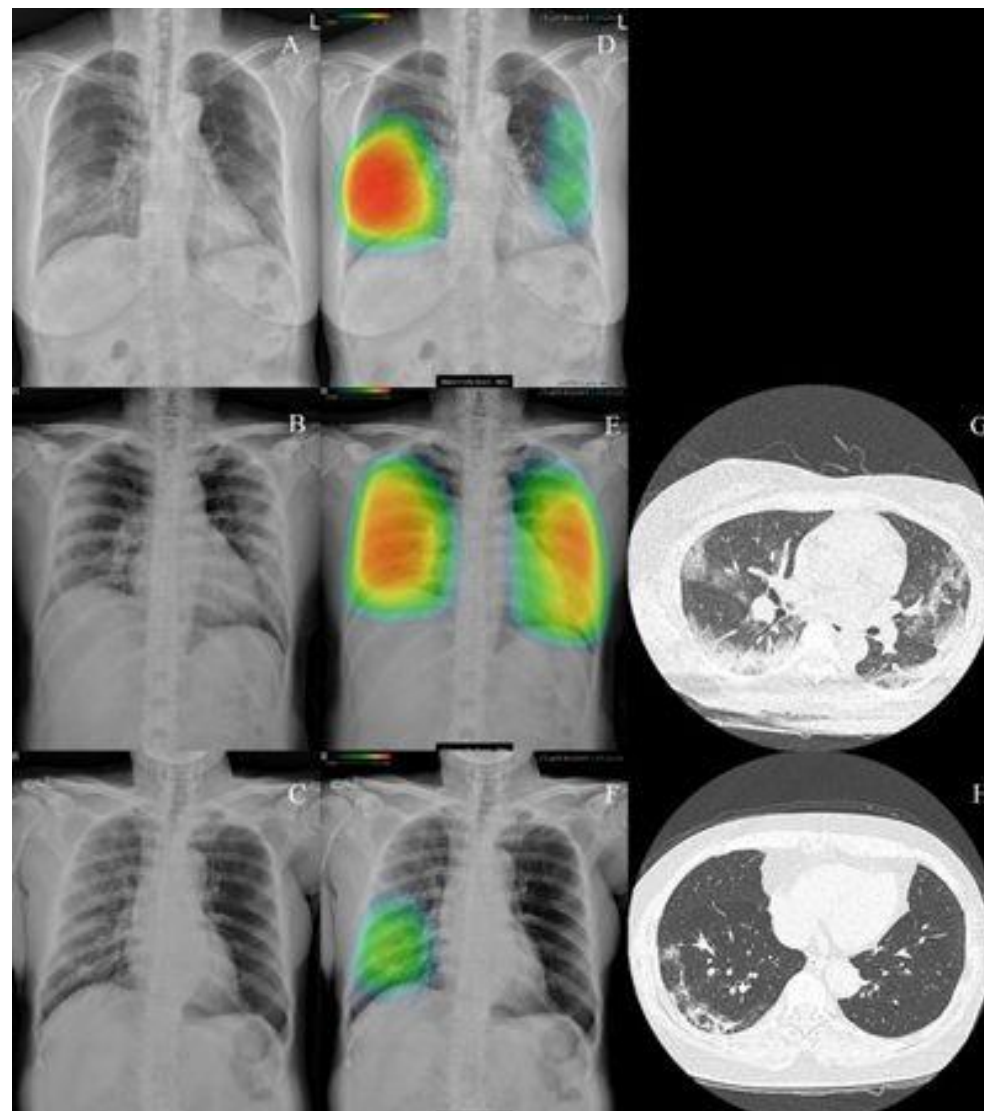


INSIGHT CXR - La soluzione di analisi dei raggi X del torace basata sull'AI

Check Out Results from Our Latest AI Algorithms within Seconds.



INSIGHT CXR - La soluzione di analisi dei raggi X del torace basata sull'AI



Reg Covid-19 – uno strumento statistico per prevedere la progressione della malattia

All'inizio di marzo 2020 in Spagna è stato sviluppato RegCov19, un registro di coorte prospettico con la capacità di un'analisi dei dati continua e immediata che mira a fornire informazioni online anonime alle unità di cura.

Più di 80 centri ospedalieri hanno aderito a questo progetto.

I dati:

dati demografici e comorbidità, sintomi prima dell'ammissione in pronto soccorso, segni vitali, pannello chimico, supporto respiratorio, trattamenti farmacologici, complicazioni mediche e infettive durante l'ammissione, cronologia della malattia e durata della permanenza in ospedale. Punteggi di gravità, numero di pazienti, pazienti dimessi, pazienti morti o ancora in corso di ricovero, mortalità a 28 giorni.



Hispabot-COVID19

Hispabot è un chatbot integrato in whatsapp per consentire la trasmissione automatica e accurata delle informazioni ufficiali ai cittadini spagnoli.

Il sistema è stato allenato con più di 200 domande che possono essere poste in 1.000 modi diversi, per le risposte utilizza informazioni provenienti da fonti del Ministero della Salute e altri organismi ufficiali relative ai sintomi della malattia, misure preventive, informazioni sulla protezione, cifre dei contagi, numeri di contatto, ecc.

Hispabot-Covid19 contribuisce a ridurre la pressione sulle hotline sanitarie, offrendo un'alternativa con informazioni semplici e disponibili in qualsiasi momento.



I Big Data delle telcom italiane per rintracciare i contatti

A Marzo 2020 società di telecomunicazioni italiane tra cui Tim, Vodafone, Wind Tre e FastWeb, hanno offerto set di dati anonimi che aggregano il movimento degli utenti per aiutare a rintracciare i contatti tra persone o consentire altre forme di monitoraggio.



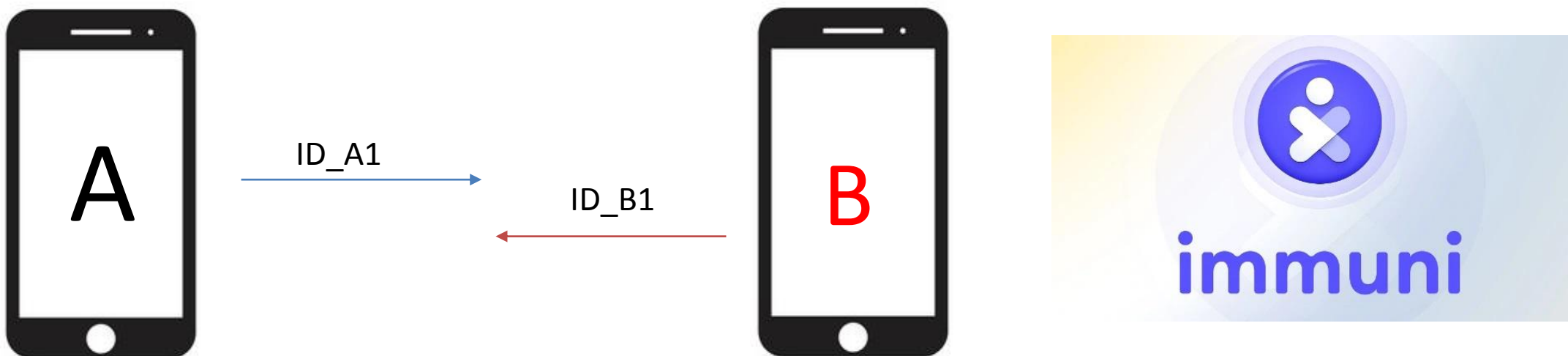
L'app Immuni



L'app Immuni

Il suo sistema è basato su Bluetooth Low Energy e non utilizza alcun dato di geolocalizzazione, compresi i dati GPS.

Così, mentre l'app sa che il contatto con un utente infetto è avvenuto, quanto tempo è durato, e può stimare la distanza che ha separato i due utenti, non può dire dove il contatto ha avuto luogo, né le identità delle persone coinvolte. L'app raccomanda agli utenti a rischio cosa fare.



City Analytics Mobility Map

Stima le variazioni negli spostamenti delle persone e i chilometri percorsi in tutto il territorio nazionale, regionale e comunale.

La soluzione visualizza dati di localizzazione anonimizzati e aggregati provenienti da sensori di veicoli connessi, sistemi di navigazione, applicazioni mobili e agenzie governative.

L'elaborazione dei dati genera quattro indicatori chiave

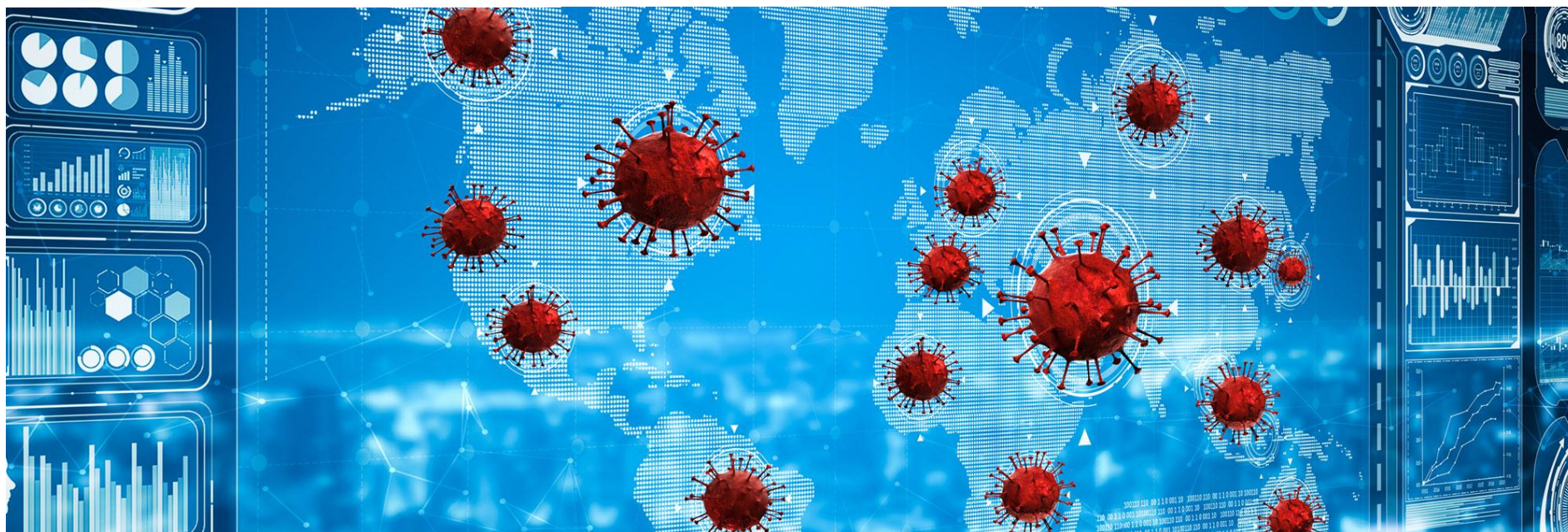
1. la variazione percentuale giornaliera del numero di viaggi in un'area
2. la variazione percentuale giornaliera del totale dei chilometri percorsi
3. la percentuale di viaggi giornalieri in entrata e in uscita secondo l'origine o la destinazione
4. la proporzione di viaggi settimanali in entrata e in uscita in base all'origine o alla destinazione

offrendo una soluzione efficace per valutare l'andamento dei flussi di mobilità sul territorio.



L'azienda Expert System e il software che prevede focolai di Covid

Questo software analizza i dati dai post dei social media, i dati degli ospedali, i registri delle chiamate di emergenza ecc. per allertare le autorità italiane in caso di tendenze allarmanti



Il software francese che riconosce chi non indossa la mascherina

Il software, integrato negli autobus, genera statistiche su quanti individui sono stati visti indossare maschere in intervalli di 15 minuti elaborando le immagini in tempo reale.

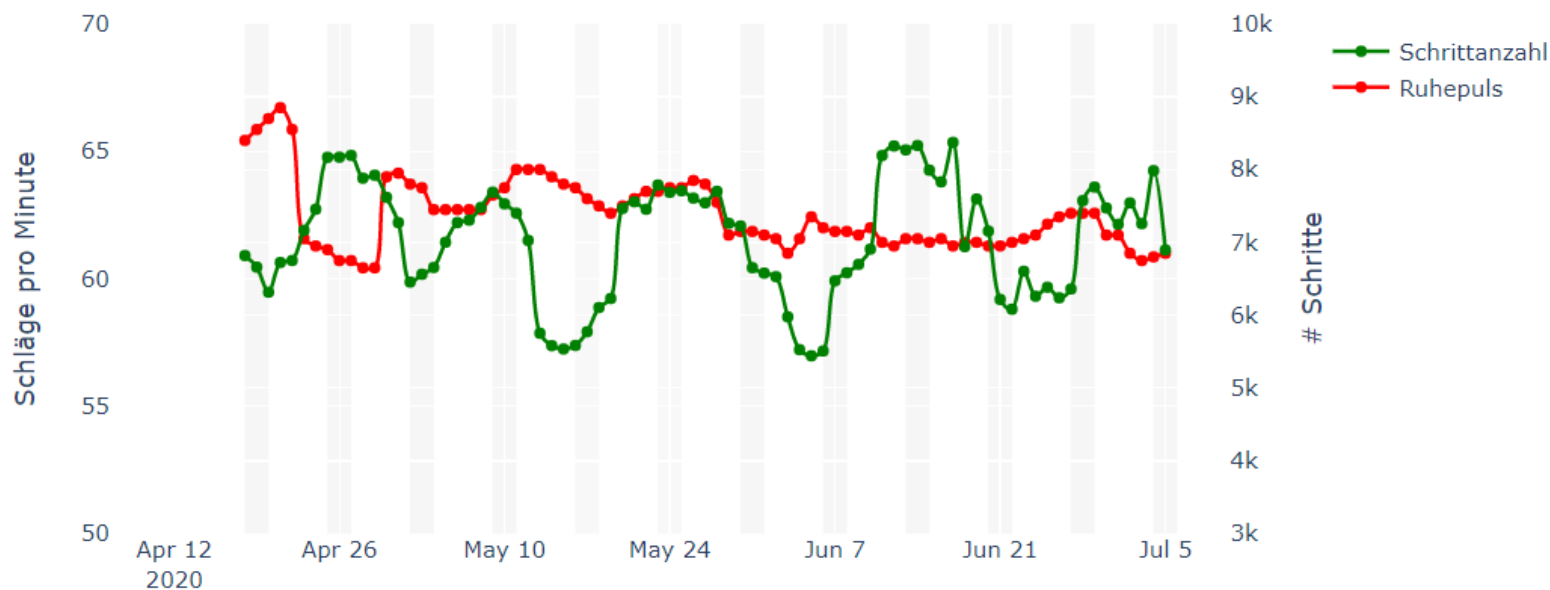
Quando l'autobus torna al deposito di notte, si connette al Wi-Fi e invia i dati alle autorità locali dei trasporti.



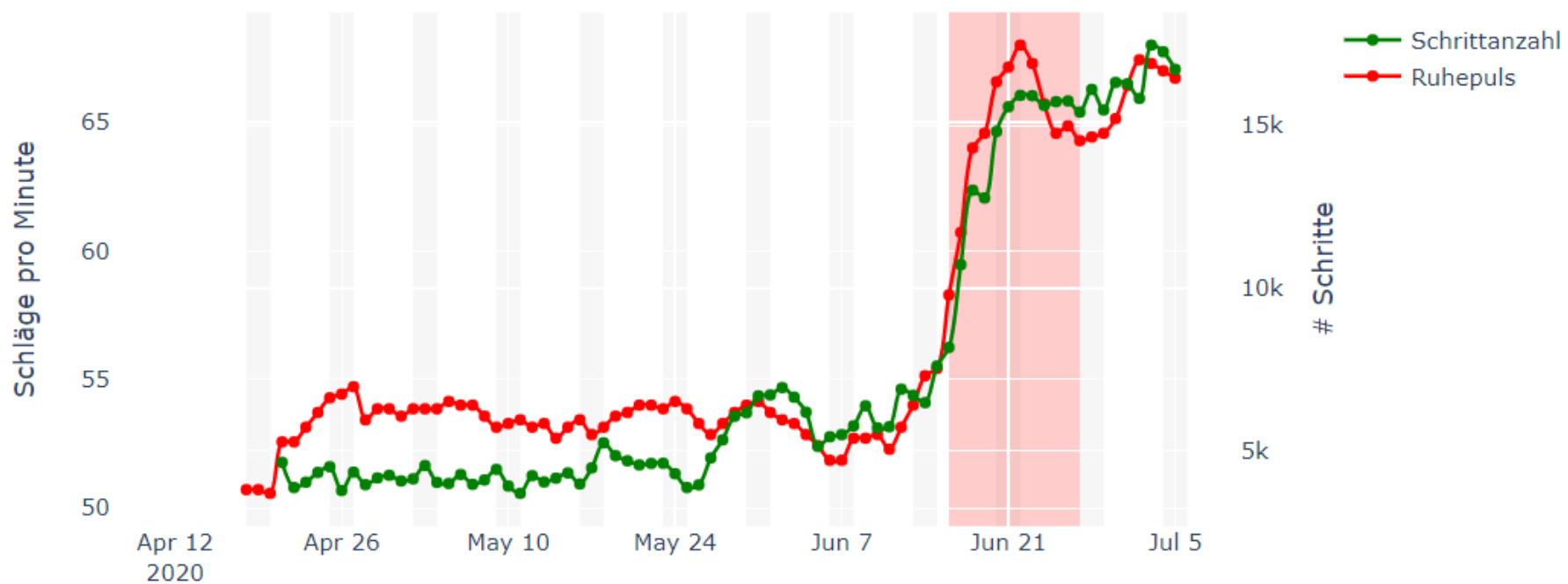
Corona Datenspende – l'app che ricava i dati dai dispositivi wearables

I dati dei braccialetti per il fitness e degli smartwatch, i cosiddetti wearables, possono fornire indizi sui sintomi dell'infezione da COVID-19. Insieme ai dati provenienti da altre fonti, ad esempio i dati di segnalazione ufficiale, questi dati aiutano gli scienziati a registrare e comprendere meglio la diffusione del coronavirus.

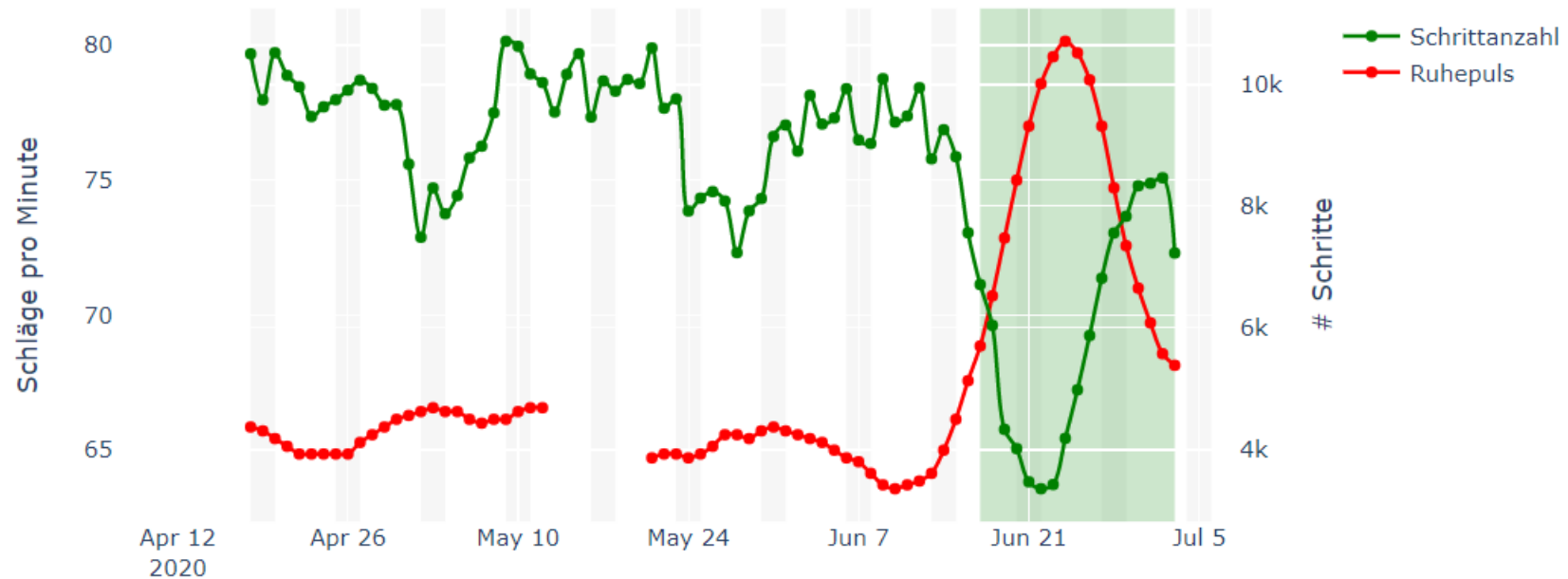
Il risultato è **la prima curva di temperatura** (o febbre).



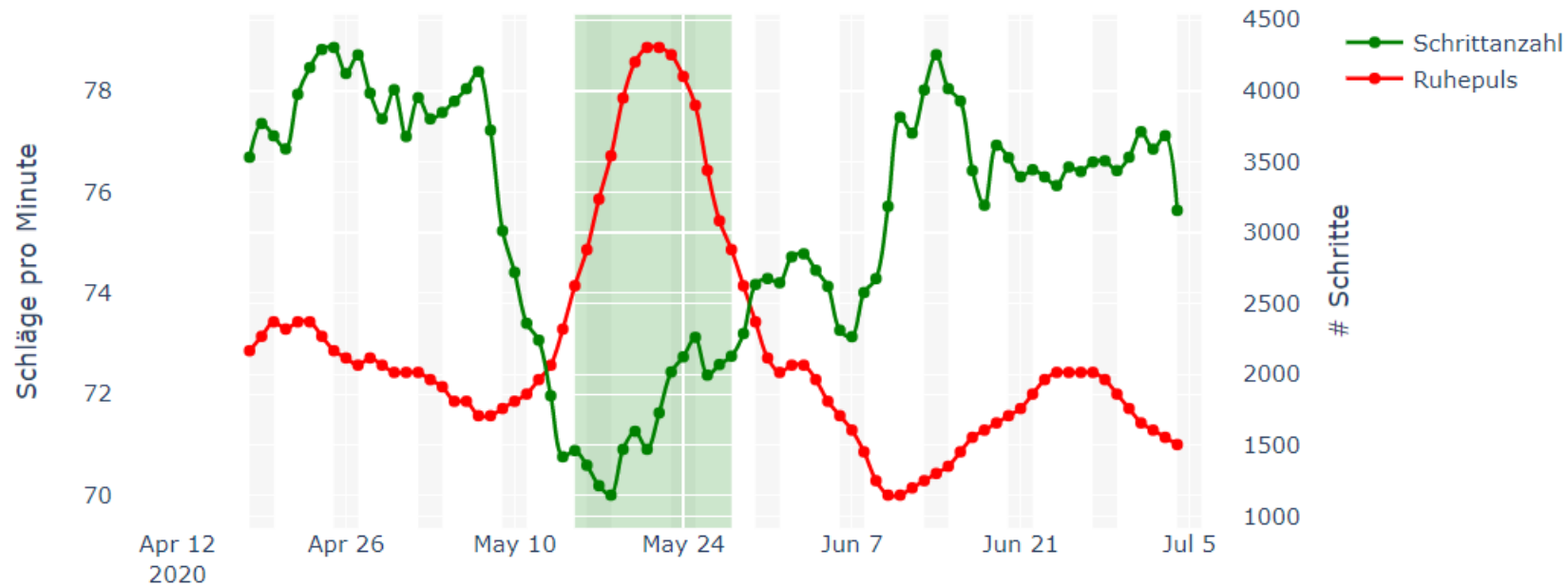
Probabilmente il segnale sbagliato



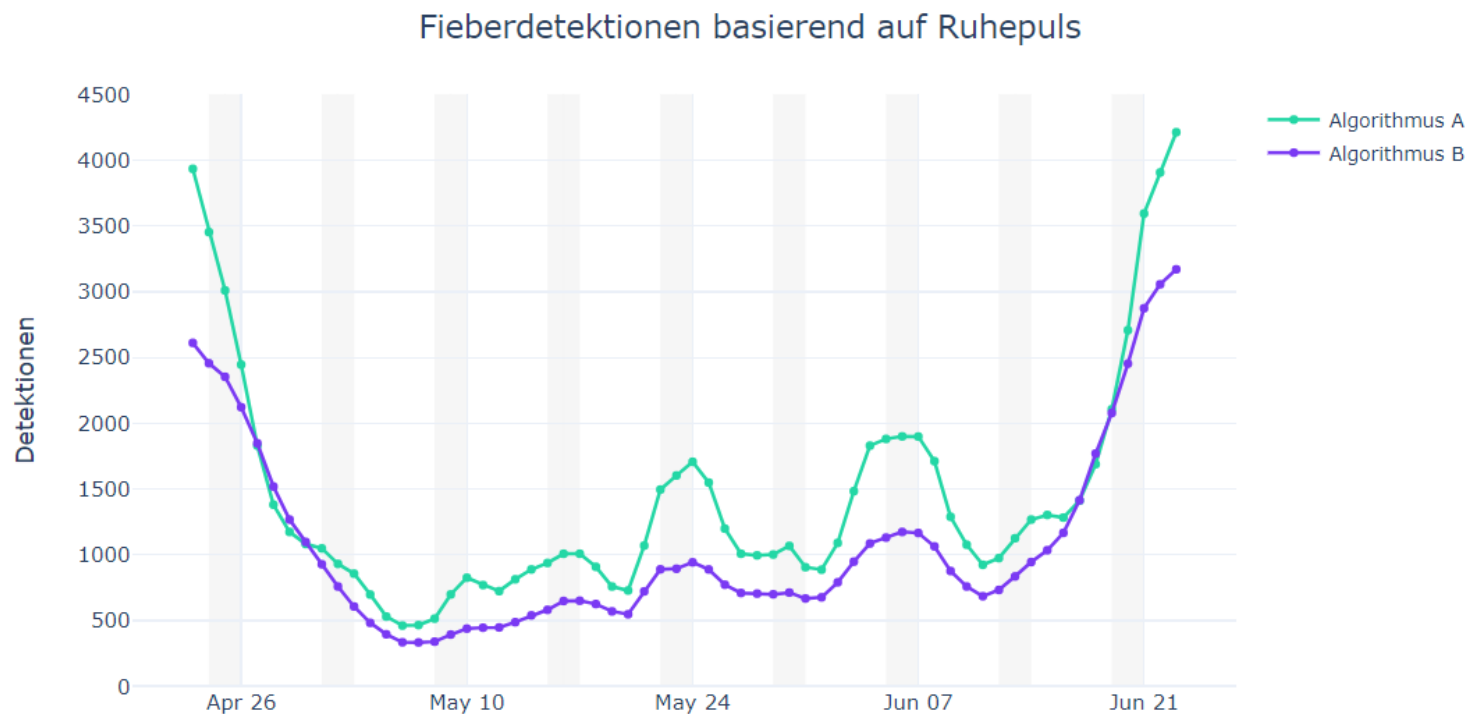
Questo donatore ha un polso a riposo di circa 65 battiti al minuto e un numero di passi giornalieri di circa 10.000. Al termine del periodo di osservazione, la frequenza cardiaca a riposo aumenta, raggiunge un picco e poi diminuisce nuovamente nell'arco di alcuni giorni. Allo stesso tempo, i passi giornalieri sono ridotti a circa 4000.



Questo donatore ha una frequenza cardiaca a riposo di circa 73 battiti al minuto e un numero di passi giornalieri di circa 3.500. L'impulso di riposo «picco» e il numero di passo «valle» coincidono.



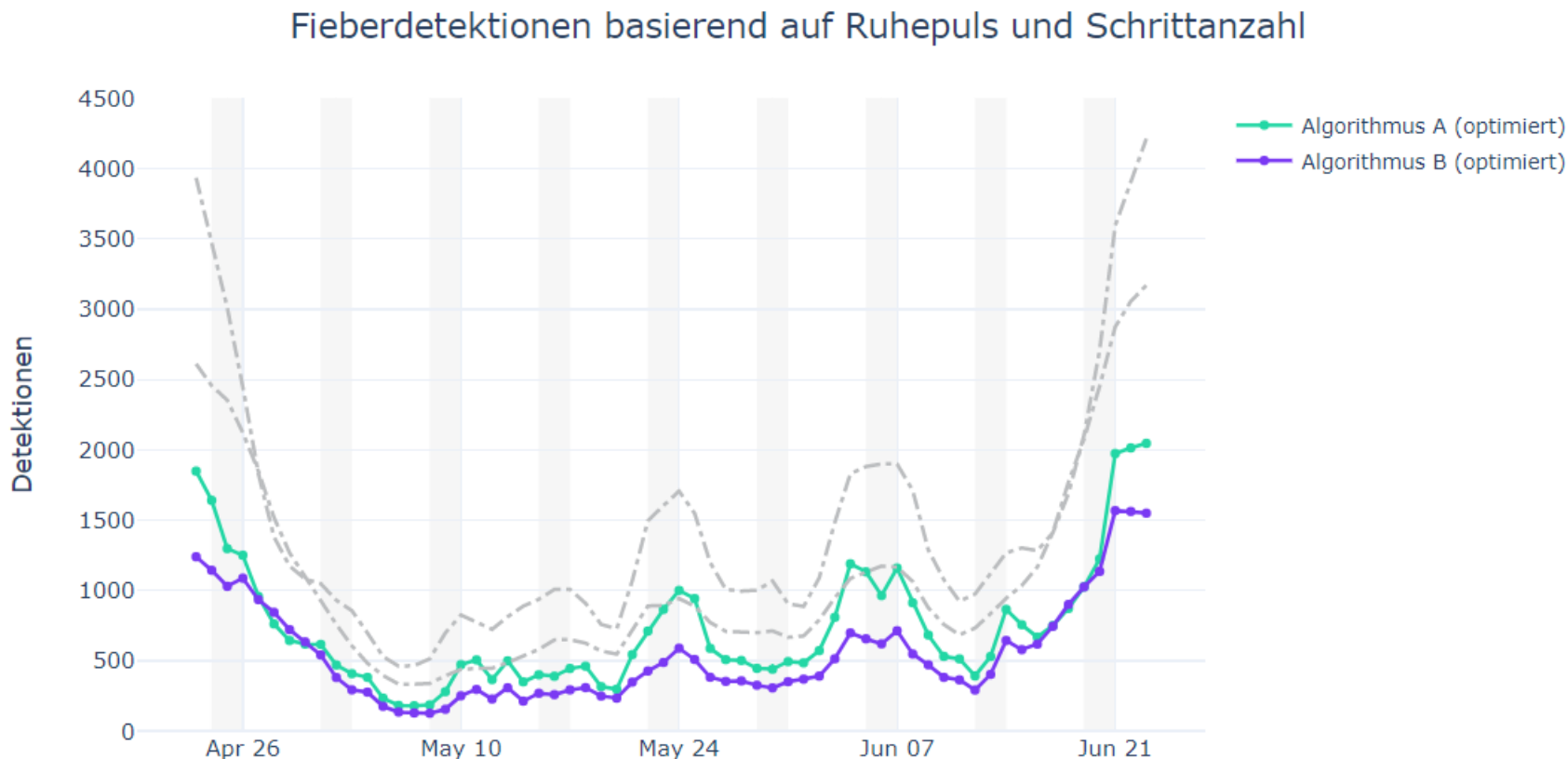
Applicando l'algoritmo solo al segnale del polso a riposo e calcolando i potenziali "rilevamenti di febbre" giornalieri per ogni donatore in serie temporali sufficientemente lunghe, hanno ottenuto il primo grafico della febbre.



La curva nazionale della febbre basata sui dati della frequenza cardiaca a riposo: per ogni giorno hanno calcolato il numero di rilevazioni in base alle anomalie della frequenza cardiaca a riposo. Entrambe le curve mostrano una diminuzione iniziale dei rilevamenti positivi, un plateau a metà del periodo di osservazione e un aumento a metà giugno.



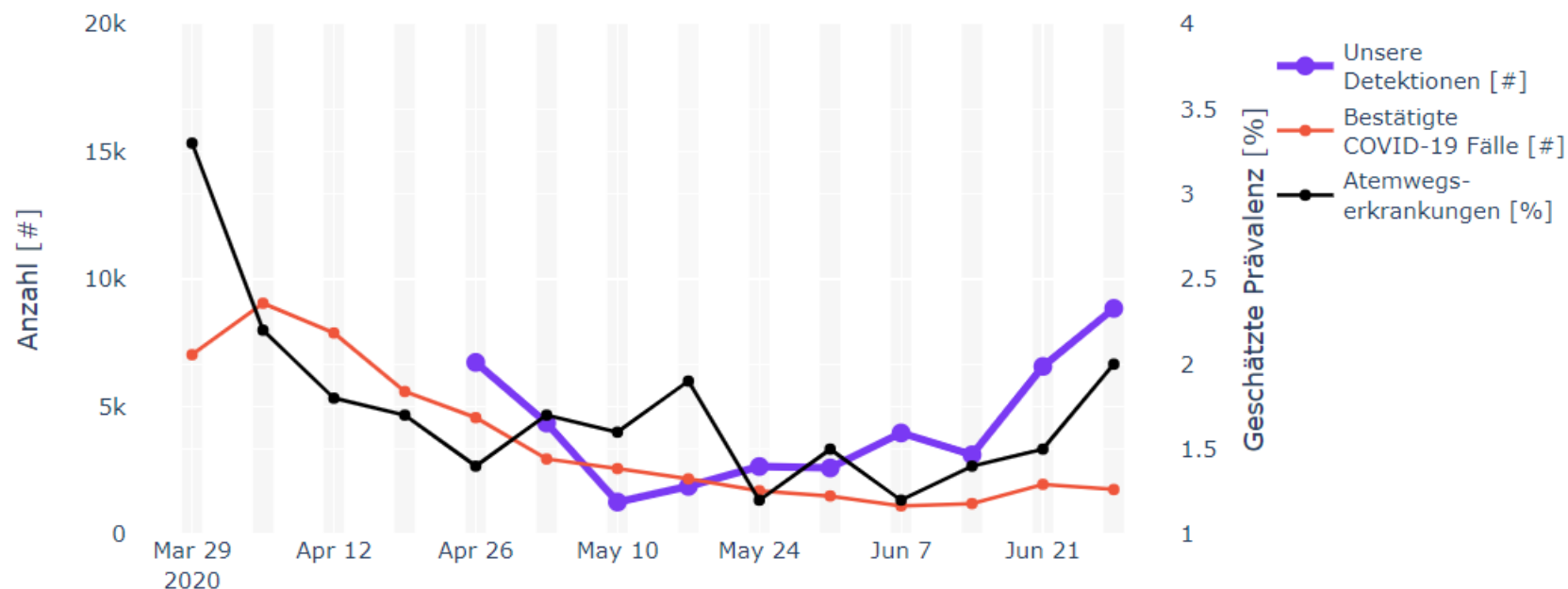
Inclusione del numero giornaliero di passi



La curva della febbre nazionale basata sui dati della frequenza cardiaca a riposo in combinazione con il numero di passi giornalieri: questo metodo riduce il numero di rilevamenti positivi. Per confronto, i risultati dell'analisi della frequenza cardiaca a riposo puro sono mostrati in grigio.



Confronto con altre fonti di dati

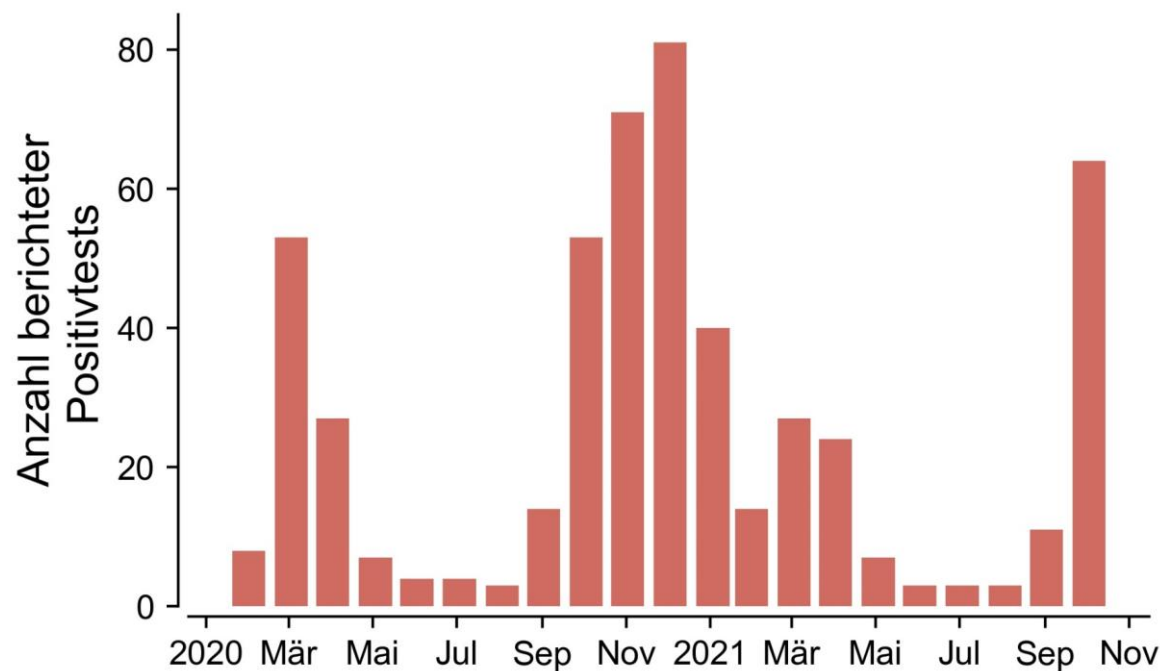


Rilevazioni settimanali rispetto ad altre fonti di dati : la figura confronta le curve di febbre previste (aggregate per settimane) con i numeri dei casi COVID-19 e anche con l'incidenza settimanale delle malattie respiratorie acute (FluWeb).



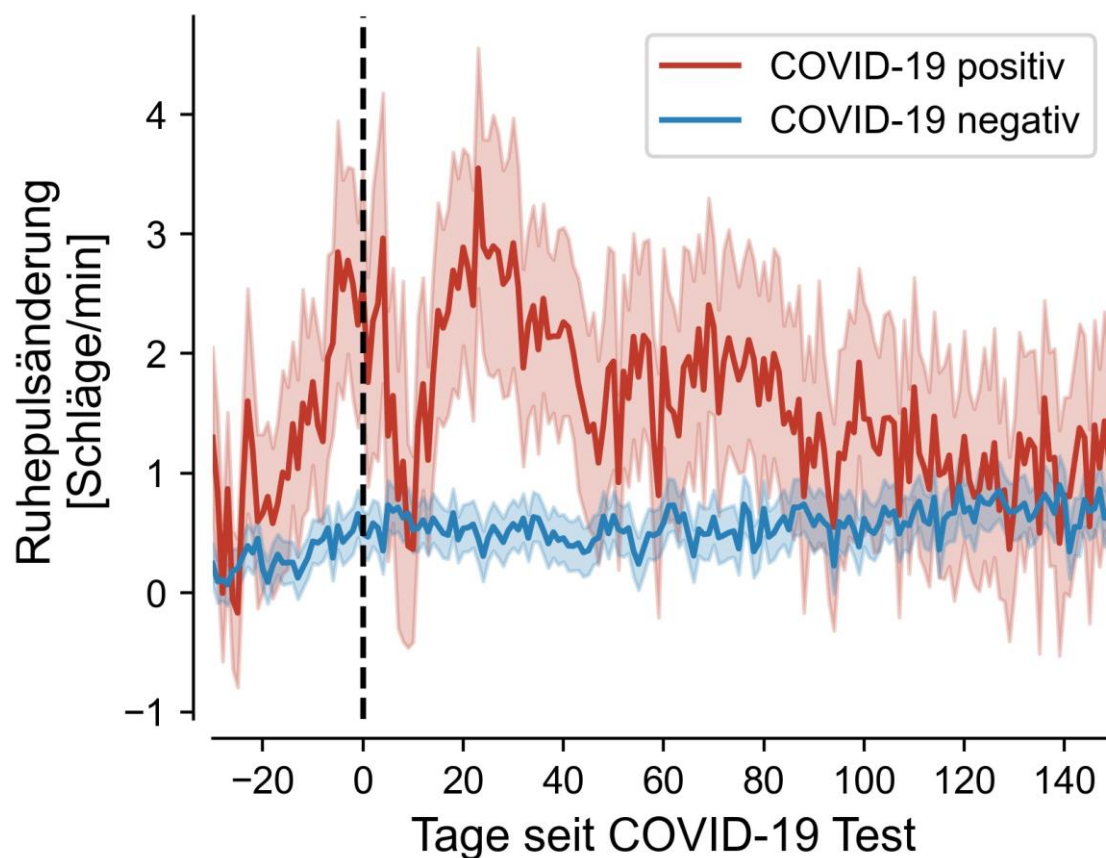
Il miglioramento della curva della febbre

Da metà ottobre è possibile aiutare il progetto e a migliorare ulteriormente la comprensione dei dati vitali donati rispondendo a brevi questionari nell'app.

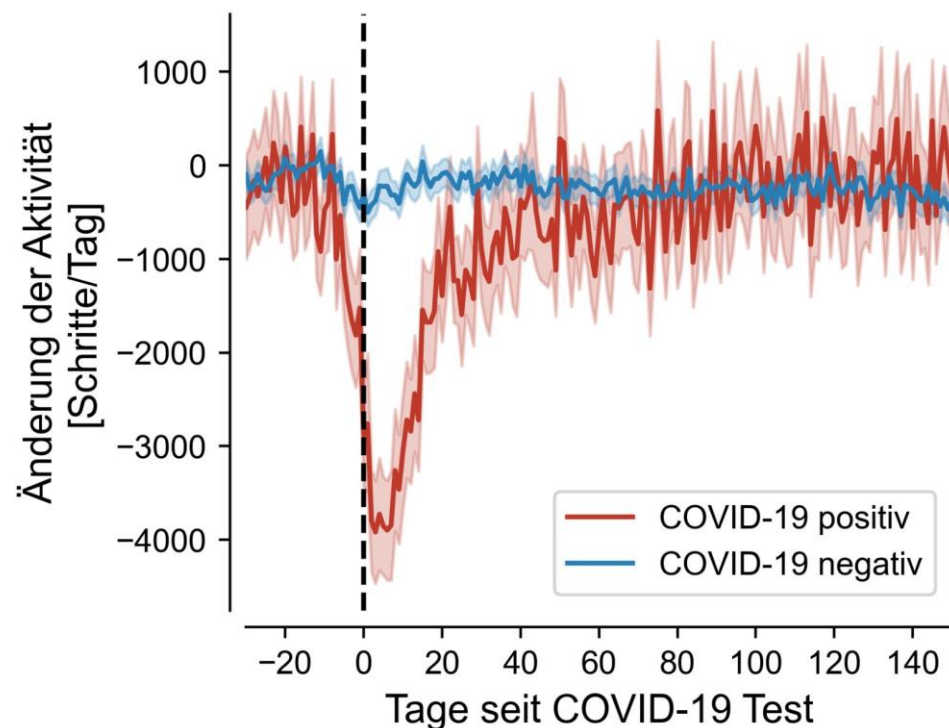


Cambiamenti a lungo termine nei dati vitali nei donatori risultati positivi

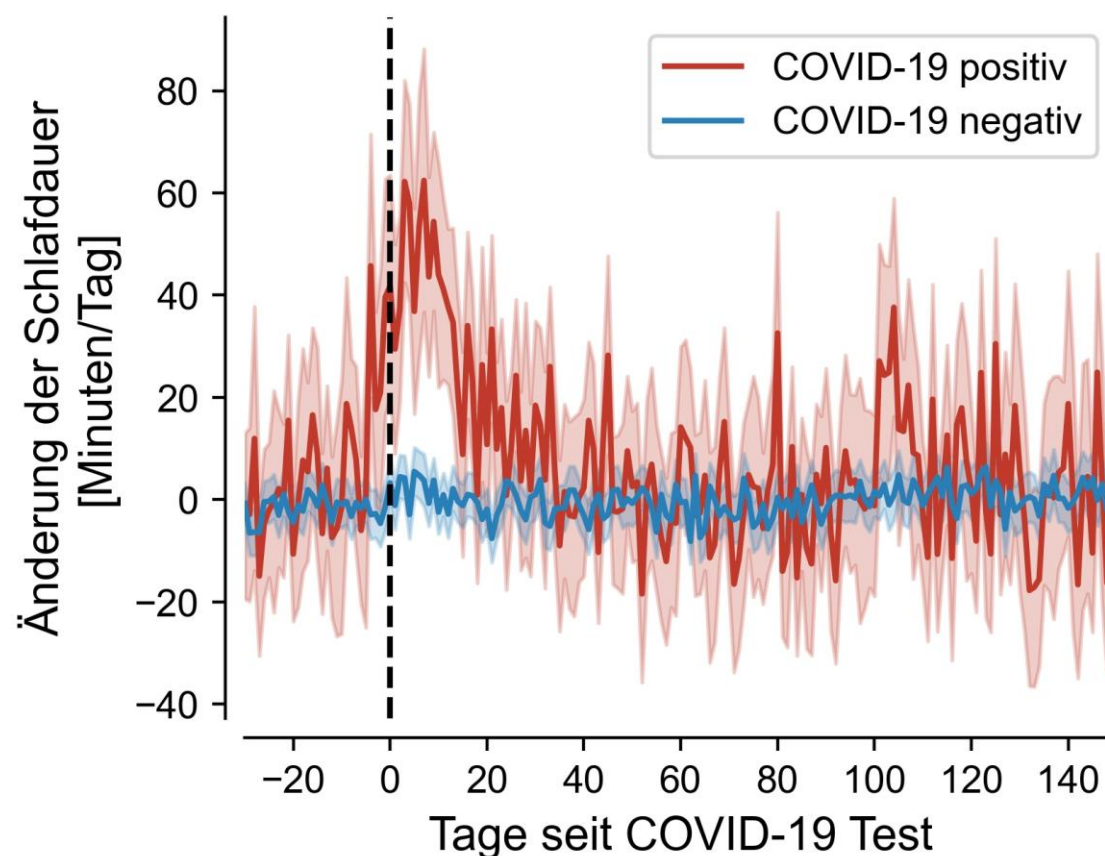
Da metà ottobre è possibile aiutare il progetto e a migliorare ulteriormente la comprensione dei dati vitali donati rispondendo a brevi questionari nell'app.



Con l'attività media (misurata in passi al giorno), l'influenza di una malattia COVID-19 può essere vista ancora più chiaramente. Alla vigilia di un test positivo, i passi al giorno diminuiscono fino a 3500, riduzione di oltre il 50% dell'attività abituale. Questa modifica dura fino a un mese.



Anche la durata del sonno è influenzata dal COVID-19. I nostri donatori dormono intorno al momento della diagnosi in media, un'ora in più. Solo dopo circa un mese questo valore torna alla normalità.



Conclusioni dello studio Corona-Datenspende

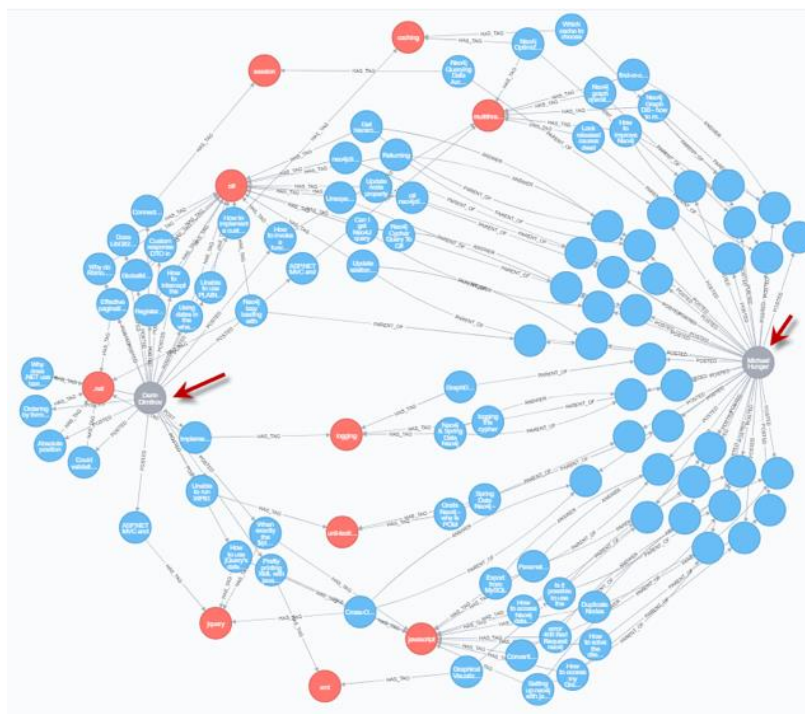
I dati donati suggeriscono che la malattia da COVID-19 può causare cambiamenti duraturi nel comportamento e dati vitali . Soprattutto nel caso della frequenza cardiaca a riposo, questo cambiamento può richiedere diversi mesi.



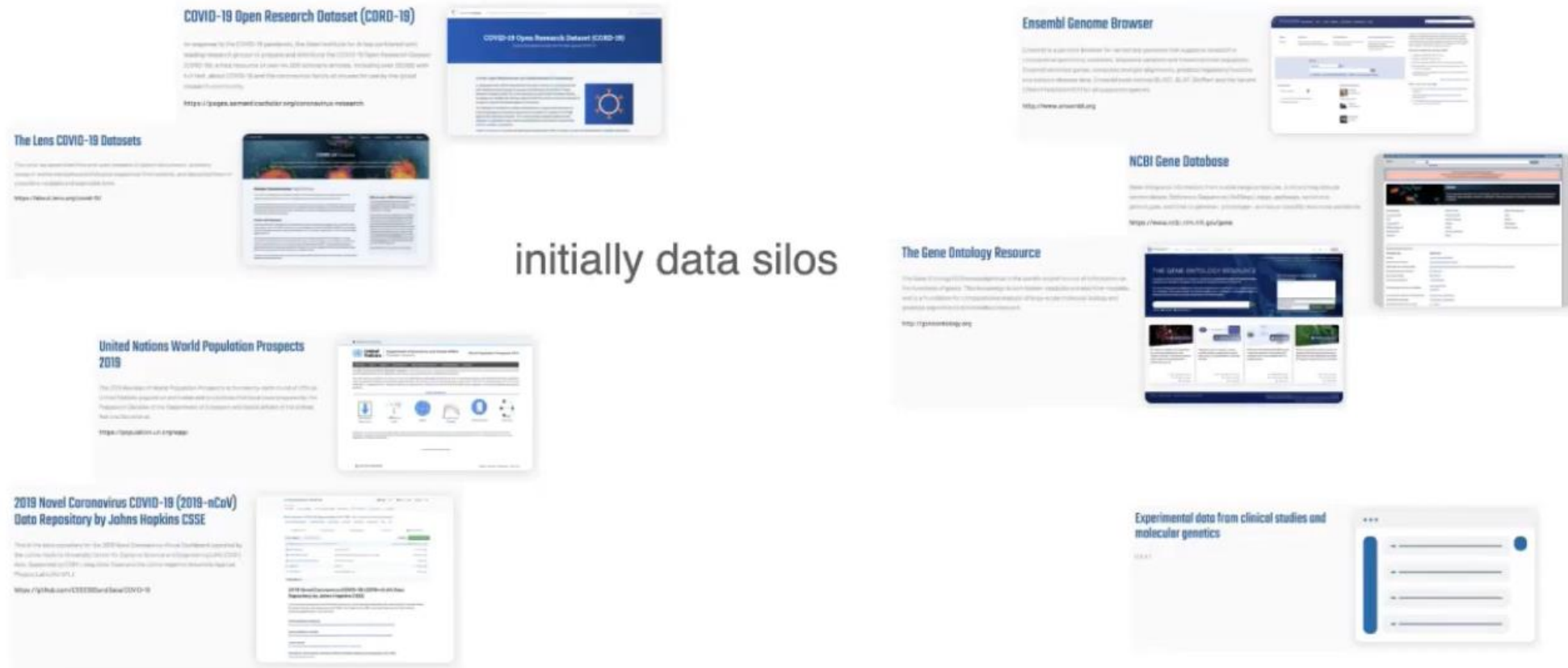
Covidgraph

Per rispondere alle domande biomediche davvero difficili quindi oggi dobbiamo collegare i dati di molti studi diversi. Quando ci avviciniamo alle malattie complesse abbiamo bisogno di una tecnologia per riunire i silos di dati e scoprire le connessioni, i grafi possono aiutarci a fare ciò.

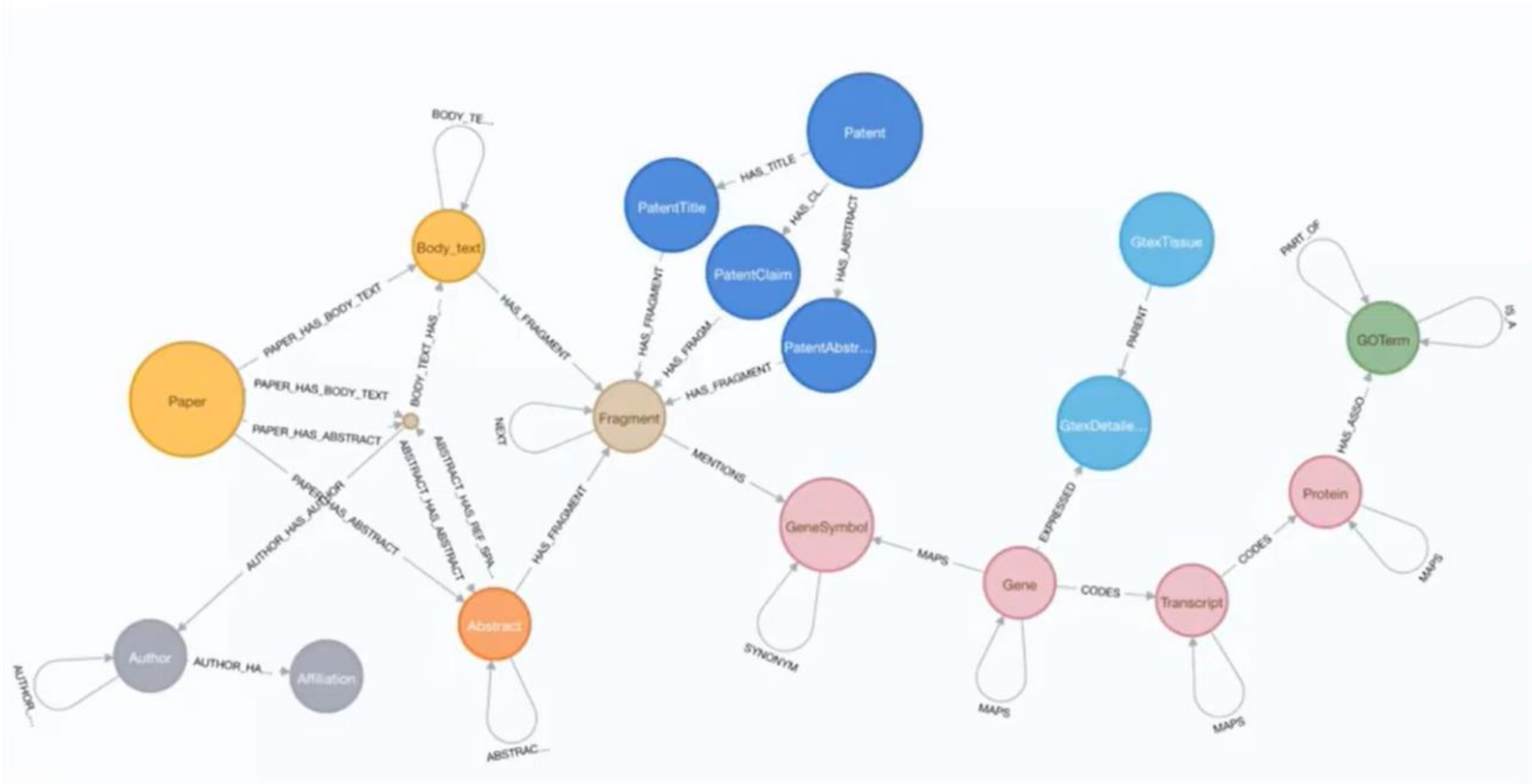
Una prima scoperta è stata riguardo ai recettori per l'angiotensina 2 chiamati recettori ACE2, i recettori delle nostre cellule che media l'infezione da SARS-CoV-2.



Datasets

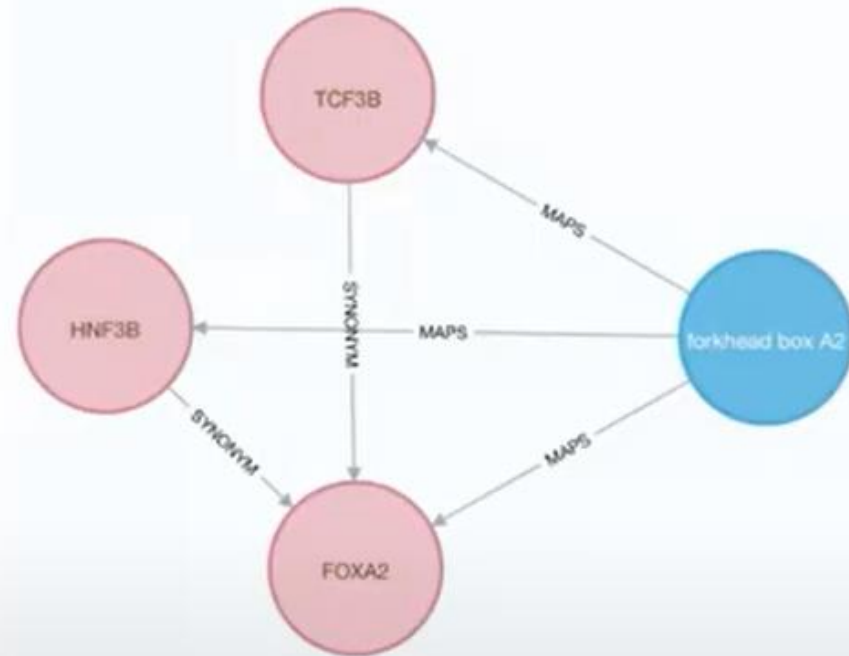


Covidgraph



Covidgraph

Example synonyms for gene „FOXA2“



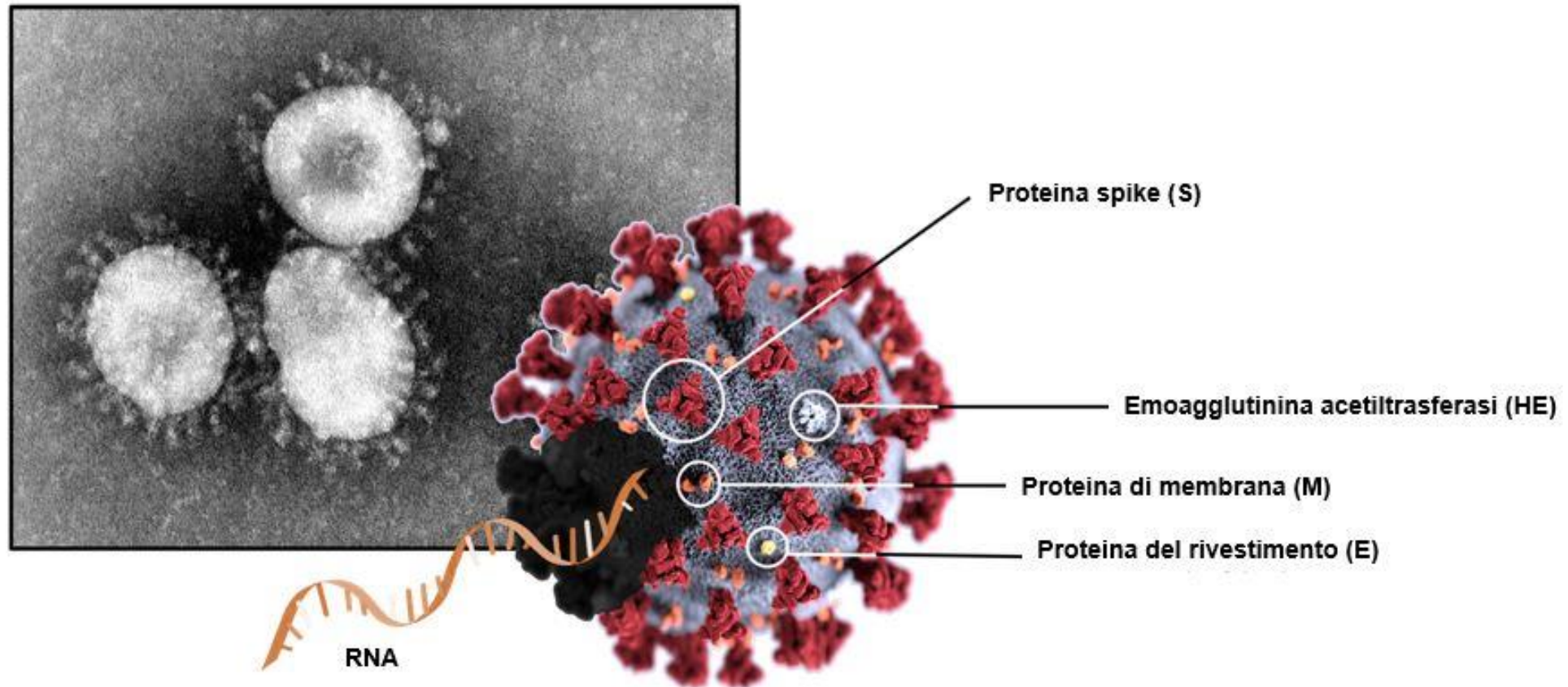
the synonyms of FOXA2 are not that obvious

in our graph database query include all synonyms on
gene, transcript and on protein level

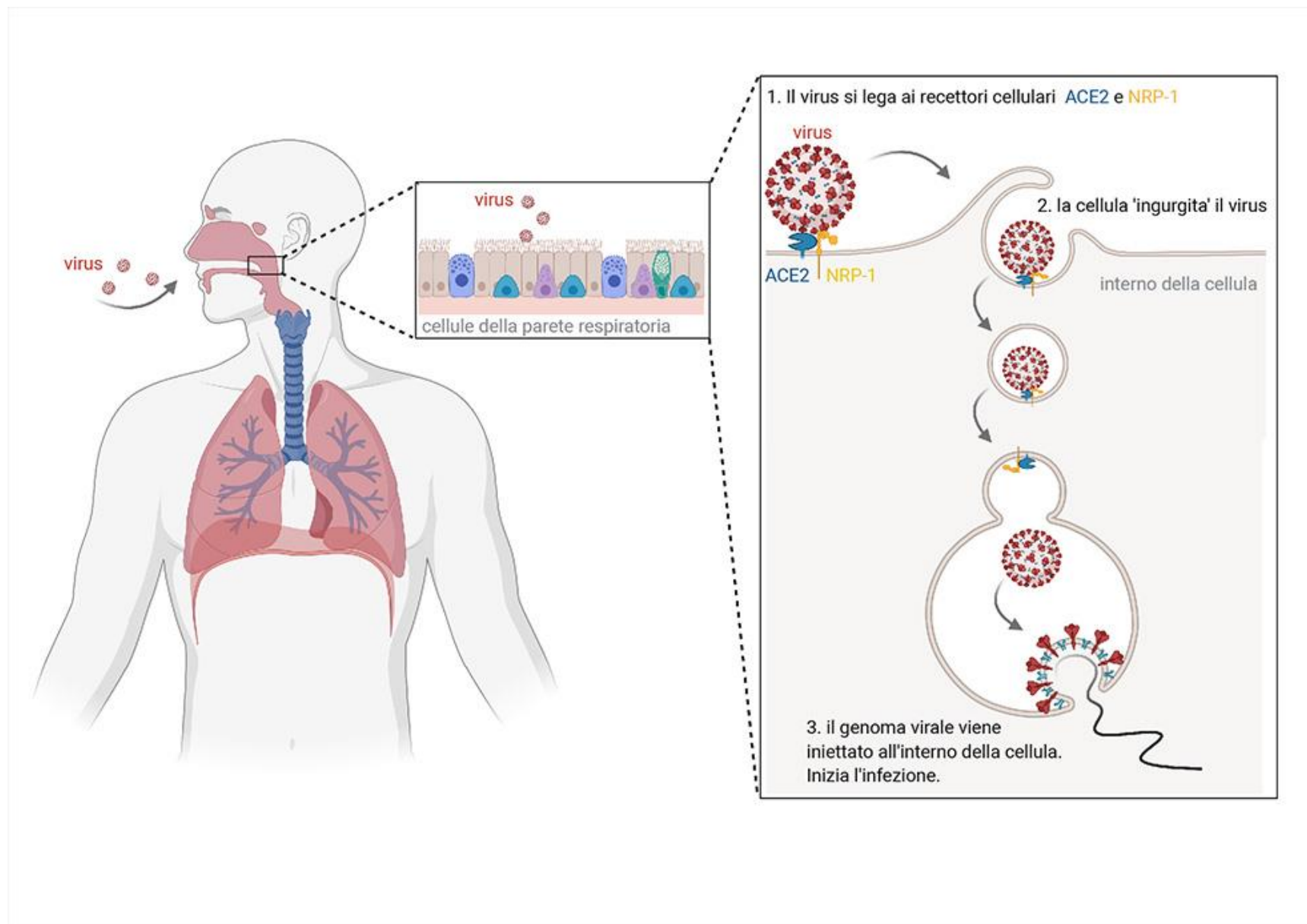


Com'è fatto un virus

Un virus è un organismo estremamente semplice che ha bisogno di infettare un altro organismo per potersi riprodurre e propagare. I virus possono essere pensati come delle scatole organiche che contengono del materiale genetico sotto forma di DNA o RNA.



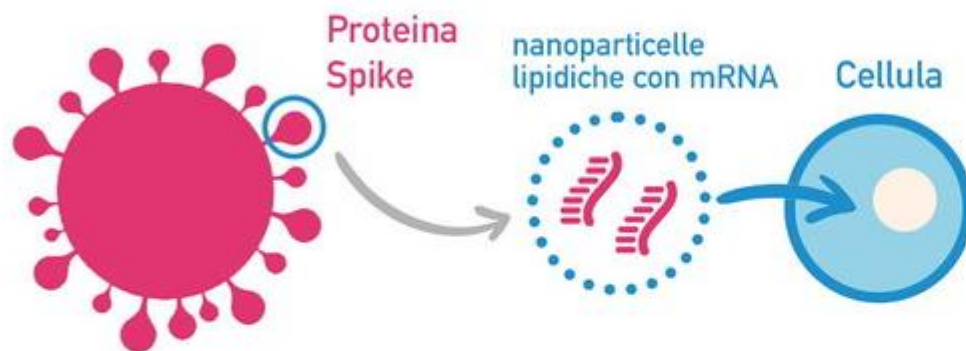
Come il Coronavirus entra nelle nostre cellule



Come funziona il vaccino a mRNA?

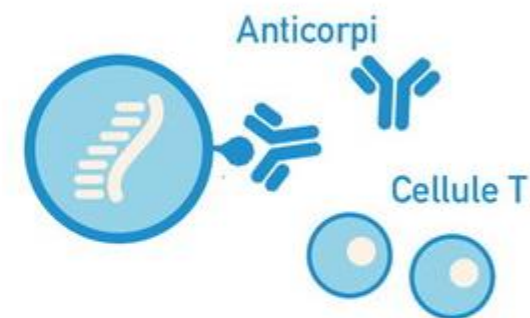
Come funziona il vaccino Pfizer BioNTech

I vaccini a mRNA danno al sistema immunitario le istruzioni genetiche per riconoscere una particolare proteina del virus



Gli scienziati si sono concentrati sulla sequenza genetica della proteina virale Spike. Hanno sintetizzato una sequenza di mRNA con le istruzioni che la cellula userà per creare altre proteine Spike.

L'mRNA di sintesi viene impacchettato in una nanoparticella lipidica che consegna le istruzioni alla cellula umana



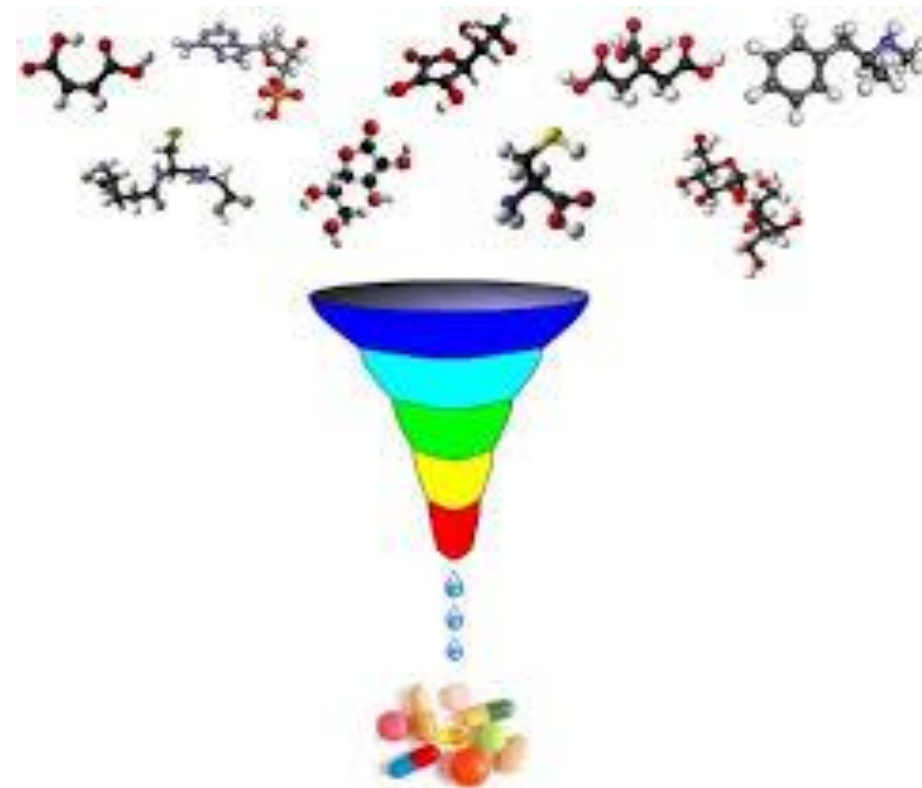
Il citoplasma cellulare seguirà le istruzioni per costruire proteine virali Spike che compariranno sulla superficie della cellula umana e stimoleranno la risposta del sistema immunitario

Exscalate4Cov

Il consorzio E4C è coordinato da Dompé Farmaceutici ed è composto da 18 istituzioni di sette paesi europei. Al centro di E4C c'è Exscalate (EXaSCale smArt pLatform Against paThogEns), una piattaforma di supercalcolo intelligente.

Attraverso la piattaforma EXSCALATE e i protocolli di screening virtuale, E4C seleziona

1. I farmaci più promettenti contro il coronavirus presi dai farmaci commercializzati e in via di sviluppo sicuri per l'uomo (> 10000 farmaci, SIM)
2. seleziona il Tangible Chemical Database (TCDb) che comprende > 500 miliardi di molecole per identificare nuovi potenziali farmaci da testare contro il coronavirus, migliorando così il tasso di successo della fase di screening virtuale.



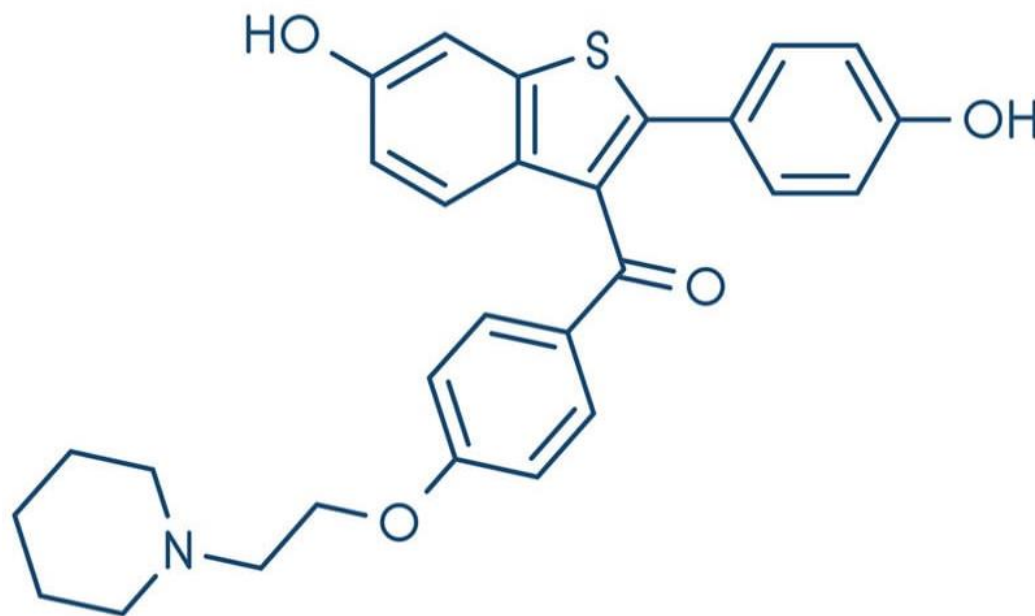
Exscalate4Cov

Dopo aver individuato le molecole promettenti attraverso le tecniche CADD, c'è una seconda fase che prevede L'high throughput screening biochimico e fenotipico che permette la rapida valutazione delle molecole che sono risultate interessanti e la riduzione dei tempi per la scoperta di nuovi farmaci.



Exscalate4Cov – il raloxifene

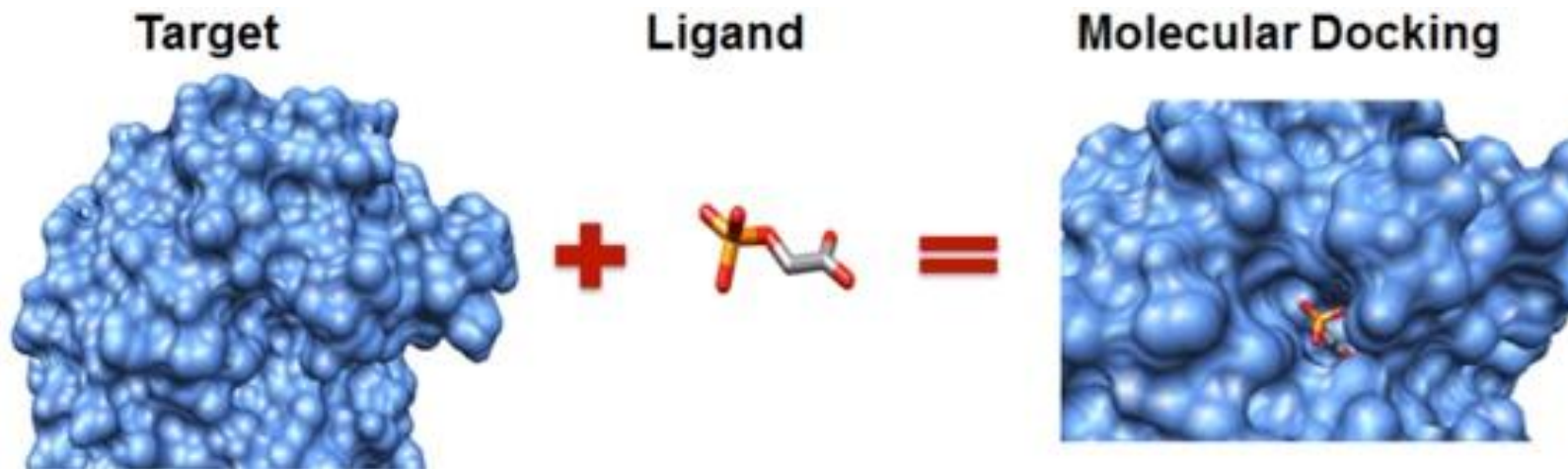
La prima fase del progetto si è basata su uno screening virtuale su oltre **400 mila** molecole messe a disposizione da Dompé Farmaceutici e dal Fraunhofer Institute, con priorità per i prodotti in fase clinica o sul mercato. Da questo test sono emerse **7 mila** molecole dalle caratteristiche promettenti che sono state testate e selezionate ulteriormente fino ad arrivare a **100** molecole attive in vitro e **40** con capacità di contrastare il virus nelle cellule animali. Alla fine si è arrivati ad **1** sola molecola promettente, come potenziale cura contro il covid-19: il **raloxifene**.



Exscalate4Cov – la simulazione da record

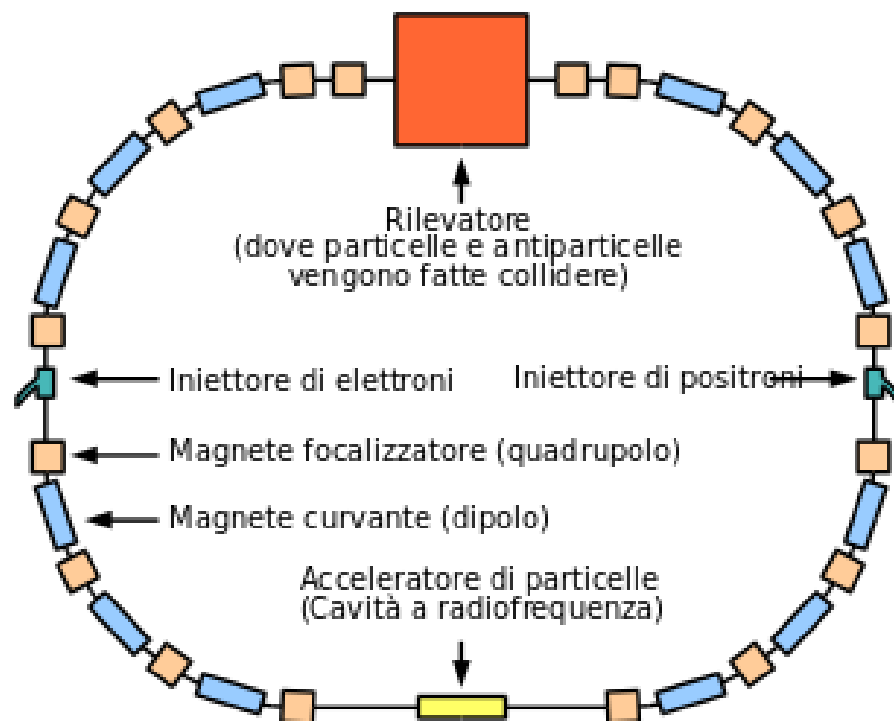
in Italia è stato eseguito l'esperimento di supercalcolo più complesso mai realizzato al mondo per studiare l'interazione di oltre 70 miliardi di molecole antivirali con le proteine del virus Sars-Cov-2.

La simulazione ha impiegato 60 ore di elaborazione, sono state testate 71,6 miliardi di molecole su 15 "siti attivi" del virus. In totale sono state elaborate oltre un miliardo di interazioni, l'equivalente di 5 milioni di simulazioni al secondo.



Exscalate4Cov – Il sincrotrone di Trieste

In questa catena di esperimenti il sincrotrone ha il compito di dimostrare che i farmaci identificati all'interno del progetto si leghino veramente ad alcune proteine del virus. La luce di sincrotrone irraggia i cristalli di proteine che deviano i raggi e da questa informazione riescono a capire com'è fatta la struttura della proteina al suo interno".





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Valentina Pellicioni

Dipartimento di Scienze per la Qualità della Vita

valentina.pellicion2@unibo.it

www.unibo.it