

15. Gennaio

L'apprendimento automatico nei processi diagnostici

Parte Prima

Il più grande pericolo dell'Intelligenza Artificiale è che le persone concludano troppo presto di averla compresa.

Premessa

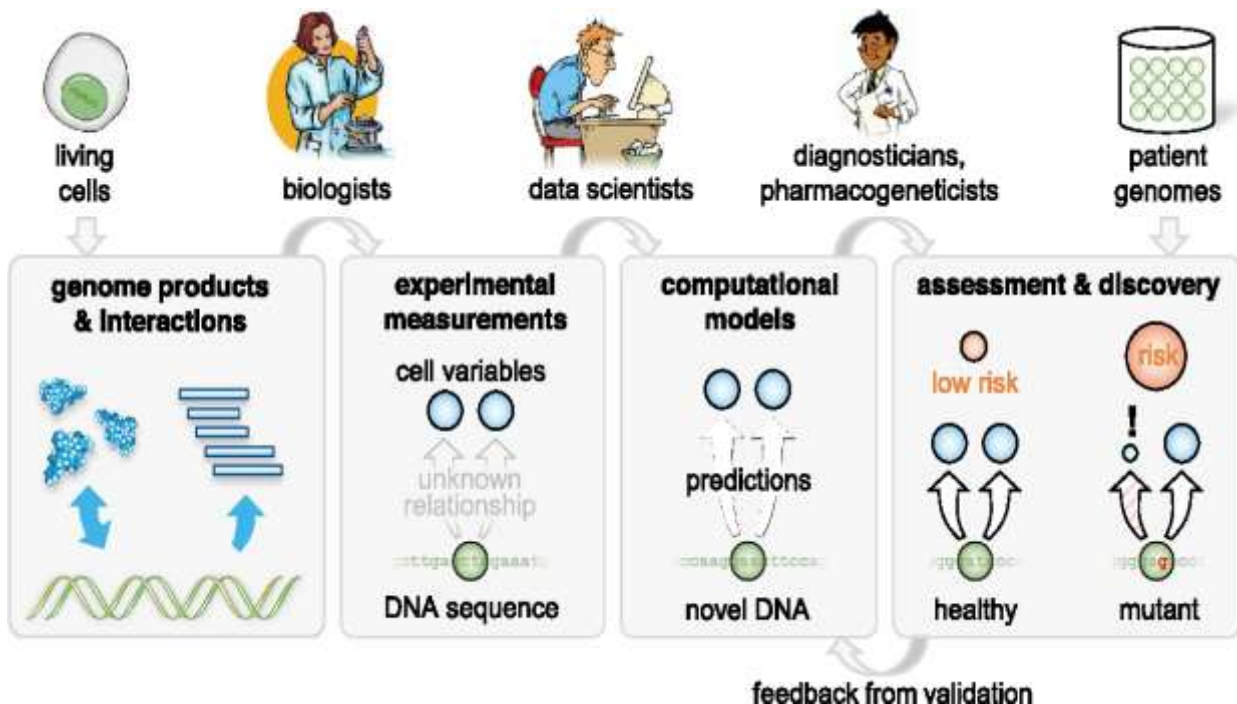
La rilevazione mediante **RT-PCR** è, a oggi, il metodo di riferimento diagnosticare un'infezione da CoV-2, ma può essere un processo lungo, in quanto può richiedere test multipli, per escludere i falsi negativi.

Le tempistiche necessarie per questo tipo di diagnosi non sempre si conciliano con le esigenze di una situazione di emergenza.

L'applicazione di approcci diagnostici complementari o alternativi può, perciò, costituire un molto strumento utile per il triage prima dell'ottenimento della diagnosi definitiva. La diagnosi rapida dell'infezione da Cov-2 nelle strutture ospedaliere o in altri contesti critici per il contenimento della pandemia (per esempio, porti e aeroporti) è fondamentale per minimizzare il rischio di trasmissione virale ad altri soggetti e anche perché un esito falso negativo potrebbe ritardare il trattamento di un soggetto sintomatico che potrebbe vedere peggiorare velocemente il suo quadro clinico.

La **tomografia computerizzata (TC)** non può essere applicata per diagnosticare i pazienti con una forma lieve della malattia, che non sempre manifestano i tipici segni radiologici. Inoltre, non tutti gli istituti ospedalieri potrebbero avere a disposizione dei radiologi con esperienza nell'imaging toracico. Infine come tutte le attività umane, la diagnosi basata sulla TC è un processo con un certo tasso di errore, soggetto all'interpretazione, non sempre riproducibile, e quindi non è completamente affidabile.

Proprio per questo motivo negli ultimi anni vari autori hanno suggerito l'uso di **strumenti di intelligenza artificiale**, allo scopo di migliorare i processi diagnostici basati sulle analisi delle immagini.



L'intelligenza artificiale (IA o AI, Artificial Intelligence) e le tecnologie correlate trovano sempre maggiore applicazione nella nostra società, e hanno certamente il potenziale per migliorare anche alcuni aspetti legati alla diagnosi dei pazienti. Con il termine **AI** non si intende una singola tecnologia, ma piuttosto un insieme di tecniche computazionali. Le metodiche di questo tipo maggiormente usate per la clinica e la diagnostica sono le cosiddette tecniche di **machine learning o apprendimento automatico**.

L'apprendimento automatico è una tecnica computazionale in grado di generare modelli per la soluzione di problemi che sono in grado di apprendere direttamente dai dati. L'apprendimento automatico è una tecnica ampia, alla base di molti approcci all'intelligenza artificiale. Esistono numerose implementazioni diverse dei metodi di apprendimento automatico. L'applicazione più comune in ambito clinico è l'implementazione di metodi per la medicina di precisione, una disciplina che richiede la capacità di prevedere quali protocolli di trattamento hanno maggiore probabilità di avere successo su determinati pazienti in base ai vari attributi dei pazienti e al contesto del trattamento.

La grande maggioranza delle applicazioni di questo tipo richiede, per addestrare un modello di apprendimento, un insieme di dati già classificati (per esempio, caratteristiche dei pazienti a seguito di insorgenza della malattia) da cui il metodo computazionale è in grado di derivare le caratteristiche salienti; questo approccio è chiamato apprendimento supervisionato in quanto l'algoritmo impara dagli esempi noti.

Uno degli algoritmi più di successo applicato ai problemi di apprendimento automatico è la rete neurale: una tecnologia disponibile dagli anni '60 del Novecento che è stata utilizzata nella ricerca sanitaria per diversi decenni, per esempio per classificare i pazienti o predire possibili fattori di rischio per una malattia. Nella sua formulazione più semplice, la rete neurale visualizza i problemi in termini di relazioni tra l'input e l'output.

Queste relazioni vengono descritte da determinate caratteristiche, cui la rete associa un peso ottimale. Più una caratteristica dell'input è fondamentale per descrivere/ottenere l'output, maggiore è il peso a essa assegnato.

Per esempio, in un caso di classificazione di pazienti, l'input potrebbero essere i soggetti dell'indagine (per esempio, pazienti e controlli sani), da cui il sistema determina automaticamente la diagnosi (per esempio, malato o sano) sulla base delle loro caratteristiche (per esempio, analisi ematiche o altri parametri clinici).

Le forme più avanzate di machine learning sono oggi note come: modelli di rete neurale con un numero elevatissimo di strati, che analizzano diversi sottoinsiemi di problemi e che mostrano una topologia molto articolata. Un sistema di **deep learning** potrebbe contenere e integrare migliaia di reti neurali che si occupano di risolvere diversi problemi specifici, per poi integrare i risultati in una singola predizione.

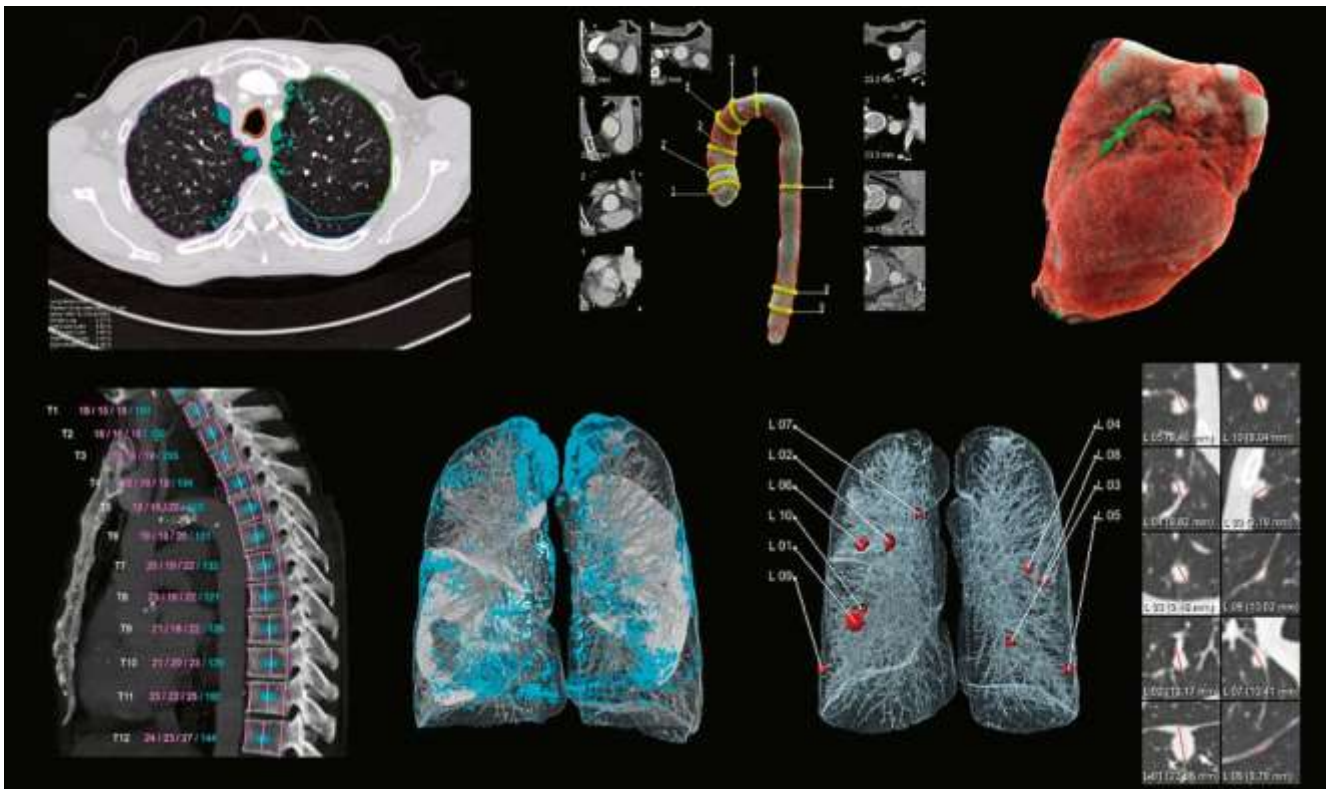
Questi approcci vengono sempre più applicati nella diagnostica medica, poiché sono più efficienti dell'occhio umano nel valutare in maniera esaustiva le complesse immagini radiologiche. L'uso di questi approcci sembra promettere una maggiore precisione nella diagnosi rispetto alla precedente generazione di strumenti e metodi. Alcune applicazioni basate su metodi di intelligenza artificiale sono state recentemente usate anche per la diagnosi della COVID-19.

Un gruppo di ricercatori *dell'Icahn School of Medicine di Manhattan* è stato in grado di sviluppare un algoritmo di apprendimento automatico, basato su una rete neurale, che applica principi di intelligenza artificiale a sostegno delle metodiche molecolari. Questo algoritmo è stato progettato per favorire/velocizzare la diagnosi in tempi rapidi di pazienti che si sospetta possano avere la COVID-19.

Il software, oltre ad un elaborato modello per l'analisi dei dati radiologici dei pazienti, integra nel processo di decisione la loro storia clinica, i risultati di tutti i test di laboratorio eseguiti, i diversi parametri fisiologici, i sintomi e, quando possibile, o dati riguardo a episodi di esposizione ad altri soggetti affetti, o informazioni sulle catene di contagio locali

Lo studio ha dimostrato che la rete neurale è più sensibile e specifica (84,3% e 82,8%) dei modelli basati solo sulle immagini radiologiche (83,6% e 75,9%) o sui dati clinici (80,6% e 68,3%)
Tenendo conto di questi dati sperimentali, l'algoritmo potrebbe essere usato per identificare rapidamente le persone con il CoV-2 prima dei risultati dei test molecolari.

To be continued... a domani





Il gene perduto ed il volo del colibrì

I colibrì mostrano un vero volo sospeso, un'attività incredibilmente ad alta intensità energetica. La velocità con cui il colibrì batte le sue piccole ali non ha eguali: 50-80 volte al secondo muovendole secondo una complicata figura a 8 che gli consente di compiere manovre precise e rimanere sospeso in aria nello stesso punto ronzando come un'ape.

In questo modo può nutrirsi del nettare dei fiori di piante e alberi senza doversi posare, passando da un fiore all'altro in breve tempo.

Di conseguenza, trascorrendo la maggior parte del tempo in volo, i colibrì hanno zampe corte e piccole.

Di notte, quando finalmente si fermano, essendo troppo piccoli per mantenere il calore, entrano in uno stato di torpore diminuendo i battiti cardiaci e, con esso, il fabbisogno di energia. Il segreto di questa potenza energetica è identificabile nella storia della sua evoluzione

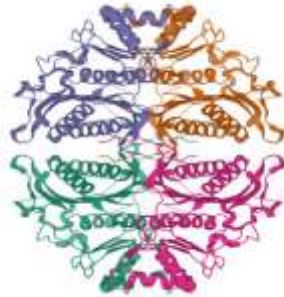
(Finestra 1)

Sebbene si sappia molto sulla fisiologia di questo movimento, poco è stato scoperto sulla genetica alla base della sua evoluzione. Il team di **Ekaterina Osipova** del *Max Planck Institute of Molecular Cell Biology and Genetics di Dresden*



hanno esaminato i genomi degli uccelli appena generati e precedentemente sequenziati per cercare i cambiamenti chiave che facilitano questa locomozione ad alta energia

Hanno scoperto che un **enzima muscolare gluconeogenico FBP2** è stato perso durante l'evoluzione del volo in bilico.



Il **knockout** di questo gene analizzato nelle linee di cellule aviarie ha portato ad un aumento della glicolisi, della produzione di mitocondri e della respirazione mitocondriale, determinando così le condizioni per una maggiore efficienza energetica.

Osipova E et al. Loss of a gluconeogenic muscle enzyme contributed to adaptive metabolic traits in hummingbirds. Science. 2023 Jan 13;379(6628):185-190.

Questi risultati suggeriscono come la perdita di un gene possa essere adattativi ed indispensabile ai fini evolutivi.

Identificare i cambiamenti genomici che sono alla base degli adattamenti fenotipici è una sfida chiave nella biologia evolutiva e nella genomica.

La perdita di geni codificanti proteine è un tipo di cambiamento genomico con il potenziale per influenzare l'evoluzione fenotipica.

Una serie di perdite geniche hanno contribuito agli adattamenti morfologici, fisiologici e metabolici nei mammiferi acquatici e volanti.

Queste perdite geniche fanno luce sui possibili meccanismi molecolari e cellulari che sono alla base di questi fenotipi adattivi.

La perdita genica è un meccanismo evolutivo per l'adattamento che potrebbe essere più diffuso di quanto previsto e costituisce un grande potenziale per rivelare le basi genomiche alla base dei cambiamenti macroevolutivi.

Sharma V et al. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. Nat Commun. 2018 Mar 23;9(1):1215.

Un aspetto intrigante del silenziamento del **FBP2** è presente in un ampio spettro di sottotipi di sarcoma come potrai approfondire, se lo desideri, nelle **finestre 2 e 3** allegate

Finestra 1: A proposito dei Colibri

Esistono 300 specie diverse di **colibri**, tutti di piccole dimensioni e con becchi adatti a impollinare particolari fiori. La specie più minuta è il **colibri ape** (*Mellisuga helenae*), che vive a Cuba ed è il più piccolo uccello conosciuto al mondo: i maschi misurano in media 5,5 centimetri e pesano 1,95 grammi, mentre le femmine raggiungono 6,12 centimetri e un peso di 2,6 grammi.

Il ritmo della diversificazione delle specie nei grandi cladi può rivelare meccanismi evolutivi fondamentali che operano su grandi scale temporali e spaziali. I colibri si sono irradiati in un variegato assemblaggio di nettariatori specializzati comprendenti 338 specie, ma la loro storia evolutiva non è stata, fino ad ora, completamente esplorata.

Uno studio della filogenesi calibrata nel tempo per 284 per 284 specie di colibrì, dimostra che i colibrì hanno invaso il Sud America circa 22 milioni di anni fa e successivamente si sono diversificati in nove cladi principali

McGuire JA et al .*Molecular phylogenetics and the diversification of hummingbirds*. *Curr Biol*. 2014 Apr 14;24(8):910-6.

Finestra 3 : A proposito del FBP2

Il fruttosio 2,6-bisfosfato è un derivato del fruttosio che, pur non essendo un composto intermedio del metabolismo dei carboidrati, ne regola finemente le dinamiche.

Il metabolismo dei carboidrati (glicolisi/gluconeogenesi) ha 3 cicli del substrato e quindi 3 potenziali punti di controllo del flusso glicolitico nei confronti di quello gluconeogenico e viceversa. Il flusso netto attraverso il ciclo del substrato, creato dalle reazioni opposte della PFK (fosfofruttochinasi 1) e della FBPasi (fruttosio bisfosfatasi), è determinato dalla concentrazione di fruttosio 2,6-bisfosfato **FBP2**

La concentrazione del **FBP2** dipende dalla sua velocità di sintesi e degradazione da parte degli enzimi PFK-2 (sintesi) e FBPasi-2 (degradazione), che si trovano in domini differenti sulla stessa proteina omodimerica con peso molecolare di circa 100 kD.

Questo enzima bifunzionale è regolato da un processo di fosforilazione/defosforilazione catalizzato da protein-chinasi A (PKA) e dalla fosfoproteina fosfatasi (PP1) rispettivamente. L'equilibrio tra glicolisi e gluconeogenesi è quindi sottoposto a controllo ormonale.

Per esempio, se nel sangue c'è poco glucosio interviene l'ormone peptidico glucagone, che fa aumentare la concentrazione del cAMP e quindi la PKA, con conseguente fosforilazione e quindi inattivazione del dominio enzimatico della PFK-2 e attivazione del FBPasi (diminuzione della concentrazione di **FBP2**). Il risultato è un aumento della via gluconeogenica.

Viceversa, se c'è molto glucosio nel sangue, interviene l'insulina, che a sua volta attiva la PP1 che defosforila, e quindi attiva, la PFK-2 e disattiva la FBPasi (aumento della concentrazione di F2,6P), risultando in un aumento della via glicolitica.

Il **FBP2** si forma da una piccola parte di fruttosio 6-fosfato prodotto nella glicolisi e sottratto grazie all'azione di un enzima bifunzionale.

Oltre ad essere un attivatore della fosfofruttochinasi (enzima glicolitico), è un potente inibitore della fruttosio bisfosfatasi (enzima gluconeogenico).

Nei tessuti non gluconeogenici, come il muscolo, **FBP2** è controllata in maniera differente per la presenza di isoenzimi differenti di PFK-2/FBPasi-2. Per esempio nel muscolo cardiaco la fosforilazione dell'enzima PFK-2 lo attiva invece di inibirlo. L'isoenzima del muscolo scheletrico, invece, non ha siti di fosforilazione, quindi è indipendente dalle variazioni di cAMP.

Finestra 2 : FBP2 e sarcomi

La notevole eterogeneità cellulare e genetica dei sarcomi dei tessuti molli (STS) limita il beneficio clinico delle terapie mirate.

L'espressione dell'isoenzima gluconeogenico **FBP2** è messa a tacere in un ampio spettro di sottotipi di sarcoma, rivelando un'apparente caratteristica metabolica comune condivisa da diversi STS. L'espressione forzata di **FBP2** inibisce la crescita delle cellule del sarcoma e del tumore attraverso due meccanismi distinti. In primo luogo, l'**FBP2** citosolico antagonizza l'elevata glicolisi associata all'"effetto Warburg", inibendo così la proliferazione delle cellule del sarcoma. In secondo luogo, l'**FBP2** localizzato nel nucleo limita la biogenesi e la respirazione mitocondriale in modo indipendente dall'attività catalitica inibendo l'espressione del fattore respiratorio nucleare e del fattore di trascrizione mitocondriale A (TFAM). Nello specifico, **FBP2** nucleare colocalizza con il fattore di trascrizione c-Myc nel locus TFAM e reprime l'espressione TFAM c-Myc-dipendente. Questa doppia funzione unica di **FBP2** fornisce una motivazione per la sua soppressione selettiva negli STS, identificando una potenziale vulnerabilità metabolica di questa neoplasia e un possibile bersaglio terapeutico.

Huangyang P et al. *Fructose-1,6-Bisphosphatase 2 Inhibits Sarcoma Progression by Restraining Mitochondrial Biogenesis*. *Cell Metab.* 2020 Jan 7;31(1):174-188.e7.

Un consiglio

Da non perdere Il Colibrì di **Francesca Archibugi** tratto dal romanzo di **Sandro Veronesi** vincitore del Premio Strega 2020, pubblicato in Italia nel 2019 dalla casa editrice La Nave di Teseo e tradotto in 25 lingue



Per osservare gli uccelli è necessario diventare parte del silenzio.

Robert Lynd

Un anno fa... Baedeker/Replay del 15 gennaio 2022

Il bollettino della sera

Marshall McLuhan, il grande teorico della comunicazione, riteneva che il mezzo tecnologico che utilizziamo per diffondere una informazione produce effetti pervasivi sull'immaginario indipendentemente dai contenuti veicolati. La sua celebre tesi secondo cui "il medium è il messaggio" ci fa capire l'efficacia dei quotidiani bollettini televisivi del Ministero della salute su quanti sono diventati rispetto a ieri i nuovi positivi omicron. McLuhan riteneva che il pubblico è un numero bombardato da numeri e che la televisione, in particolare, è uno strumento che conforta, consola, conferma capace di inchiodare gli spettatori in una stasi fisica (stare per del tempo seduti a guardarla) e mentale (poiché favorisce lo sviluppo

di una forma mentis non interattiva, diversamente da altri ambienti comunicativi a due o più sensi). Se negli anni passati il quotidiano bollettino diramato alle 18.00 era atteso con trepidazione e speranza nell'attuale pandemia la conta quotidiana dei positivi per qualcuno ormai non serve a nulla se non a creare ansia. E avanza il fronte dei "basta report", quelli che chiedono di evitare il bollettino quotidiano dei dati perché ritenuto "ansiogeno" e "inutile".

*Il virologo **Matteo Bassetti** ha dichiarato che "il report serale non dice nulla e non serve a nulla se non mettere l'ansia alle persone, siamo rimasti gli unici a farlo. Che senso ha - chiede - dire che abbiamo 250mila persone che hanno tampone positivo? Bisogna specificare se sono sintomatici, asintomatici, sono ricoverati, stanno a casa. Sono numeri che ci fanno fare brutta figura col resto del mondo, perché sembra che vada tutto male e invece non è così, nella realtà altri Paesi che hanno molti più contagi di noi cercano di gestirli in maniera diversa. Se continuiamo così finiremo con l'andare in lockdown di tipo psicologico e sociale".*

A parte il fatto che in tutti i palinsesti televisivi, con format e ritualità differenti, è previsto il rito della comunicazione giornaliera dei positivi attraverso il mezzo televisivo, il punto, a mio avviso di questa querelle è molto più complesso e serio. Nello studio di **Albert Einstein** all'Università di Princeton c'era un cartello scritto a mano che diceva: Non tutto ciò che può essere contato conta e non tutto ciò che conta può essere contato. **Marguerite Yourcenar** ci ricorda che l'uomo è un'istituzione che ha contro di sé il tempo, la necessità, la fortuna e l'imbecille e sempre crescente supremazia del numero. Non si possono adoperare gli uomini come numeri in operazioni d'aritmetica politica, perché essi si comportano come i simboli dello zero e dell'infinito, che sconvolgono tutte le operazioni matematiche. **Kant** ci ricorda che prima di valutare se una risposta è esatta si deve valutare se la domanda è corretta. e la domanda non interrogarsi sui numeri del bollettino, quanto piuttosto se questi numeri misurano effettivamente la presenza di Sars-Cov-2

La domanda che personalmente mi pongo da due anni ed a cui sto cercando ogni giorno attraverso questo crono-racconto di dare una risposta è se il paradigma interpretativo di questa pandemia sia corretto per quanto riguarda le origini del virus (che ignoriamo), le modalità di identificazione (importanti bias nelle procedure) la stessa storia naturale (imprevedibile) che mi ha portato come cronista ad usare l'espressione prudenziale di cosa pensiamo di aver capito. Ho il sospetto che misuriamo questa pandemia con strumenti inadeguati o starati ne loro funzionamento, come si misurassimo la febbre con un termometro starato e che i numeri che registriamo non sempre sono adeguata alla realtà biologica

Limite di rilevamento di una piattaforma diagnostica : LOD Per poter definire il LOD di un test è indispensabile conoscere due valori propedeutici ed indispensabili . La dose infettiva è il numero di particelle virali sufficienti per infettare il 50% di una data popolazione che nel caso del SARS-Cov-2 è attualmente sconosciuta La carica virale il numero di particelle virali in un individuo infetto, un valore attualmente stocastico All'inizio di questa pandemia, l'identificazione molecolare del CoV-2 si è basata sul processo di trascrizione inversa associato alla PCR (RT-PCR, Reverse Transcription-Polymerase Chain Reaction) da tamponi nasofaringei od orofaringei, che è stata eseguita usando il metodo SYBRGreen® o TaqMan®. All'inizio della pandemia, la Food and Drug Administration (FDA) ha approvato un test commerciale per la rilevazione del CoV-2 della ditta Roche (cobas® SARS-CoV-2). Questo test richiede 3,5 ore; più avanti è successivamente un altro test, più veloce, l'Xpert® Xpress SARS-CoV-2 della ditta Cepheid Inc (US), con cui si ottiene il risultato in 45 minuti. A proposito di questi test vi invito a rileggere nella sezione News il report "Cosa significa essere Covid- positivo" del mese di novembre 2020 sull'utilizzo della PCR a fini diagnostici: College of American Pathologist Consiglia cautela nella interpretazione dei risultati ottenuti attraverso i tamponi molecolari (Rhoads D 2020) **Center for Evidence-Based Medicine** Mentre uno stadio infettivo può durare circa una settimana, l'RNA inattivato intracellulare si degrada lentamente e può ancora essere amplificato molte settimane dopo che l'infezione si è risolta dando una pCR positiva Centers for Disease Control and Prevention (CDC) Nonostante la sensibilità analitica molto elevata dei tamponi-PCR utilizzati ai fini di una sorveglianza hanno, nella migliore delle ipotesi hanno evidenziato solo il 10% di sensibilità per rilevare le infezioni e pertanto non sono filtri anti Covid affidabili. Inoltre dopo lo stadio trasmissibile riporta che molte, se non la maggior parte, dei positivi, non lo sono più al momento del rilevamento, oltre il 50% dei PCR positivi presenta valori soglia del ciclo bassi riconducibili ad una infezione allo stadio iniziale o avanzato. La lunga durata della "coda positiva all'RNA" suggerisce che la maggior parte delle persone infette vengono identificate dopo che il periodo infettivo è passato. Migliaia di positivi

vengono posti in quarantena di 10 giorni dopo i test RNA positivi nonostante abbiano già superato la fase trasmissibile dell'infezione

Il team di Xiao L dell'Università di Pechino Segnala che alcuni pazienti "positivi" possono rimanere tali fino a cinque settimane dopo la comparsa della sintomatologia. **Bernard La Sola, infettivologo del Insitute Ihu-Méditerranée Infecction di Marsiglia** Su un pool di 3790 campioni positivi I "positivi" con un Ct elevato possono continuare a risultare "positivi" per settimane dopo la guarigione clinica, suggerendo che attraverso la PCR è stato individuato materiale genetico di "residui virali" non infettivi. **Monica Gandhi,Università di San Francisco** Ritiene che la carica virale non causa necessariamente sintomi: il 40% dei positivi è asintomatico pur avendo una quantità di virus simile a quella di chi si ammala (Gandhi M. 2020). Gandhi M, Yokoe DS, Havlir DV. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid19. N Engl J Med. 2020 May 28;382(22):2158-2160. Marta Gaglia virologa della Tufts University. Lo stesso campione può dare risultati diversi se si usano strumentazioni per la PCR di aziende differenti, e più tamponi della stessa persona se replicati possono dare risultati con differenze significative Gaglia MM, Glaunsinger BA. Viruses and the cellular RNA decay machinery. Wiley Interdiscip Rev RNA. 2010 JulAug;1(1):47-59. doi: 10.1002/wrna.3. Epub 2010 May 6. PMID: 21956906; PMCID: PMC7169783 **David Paltiel, epidemiologo di Haward** esperto nel campo della ricerca operativa e della modellazione della simulazione delle malattie ritiene che nelle condizioni su riportate "La PCR viene ingannata tutto il tempo, sta solo raccogliendo fili di spazzatura virale". Se l'obiettivo finale dei tamponi è avere informazioni per rallentare un'epidemia, osserva, questo tipo di dati non è effettivamente così utile Paltiel AD, Zheng A, Walensky RP. Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United States. JAMA Netw Open. 2020 Jul 1;3(7):e2016818.

Ad oggi nessuna di queste dichiarazioni è stata ritrattata e/o modificata. Successivamente una moltitudine di metodi sono stati applicati, dimostrando come differenti competenze in ambito clinico, biomolecolare e matematico-statistico possono convergere su un terreno comune. Ampli Amplificazione con polimerasi e ricombinasi (RPA) Amplificazione isoterma mediata da loop (LAMP) Microarray Metodi basati sul sequenziamento di nuova generazione (NGS) Metodi basati sulla tecnologia CRISPR/Cas Dai metodi di RT-PCR, alle ultime frontiere della tecnologia CRISPR/Cas e dell'intelligenza artificiale, si sta facendo un enorme sforzo per assicurare non solo l'identificazione veloce e sicura degli individui con il CoV-2, ma soprattutto si può migliorare la loro prognosi e si possono definire protocolli terapeutici sempre più appropriati. Una domanda:

Cosa misuriamo veramente nei bollettini delle 18.00 ?

(per continuare vai all'originale)