Cell Host & Microbe

Un'omologia di sequenza e un approccio bioinformatico possono prevedere target candidati per risposte immunitarie a SARS-CoV-2

Astratto grafico



Punti salienti

- d Dieci regioni definite sperimentalmente all'interno di SARS-CoV hanno alta omologia con SARS-CoV-2
- d La bioinformatica parallela ha predetto potenziali cellule B e T.

epitopi per SARS-CoV-2

d Approcci indipendenti identificati allo stesso modo

regioni immunodominanti

d Le regioni immunitarie conservate hanno implicazioni per il vaccino progettazione contro più CoV

autori

Alba Grifoni, John Sidney, Yun Zhang, Richard H. Scheuermann, Bjoern Peters, Alessandro Sette

Corrispondenza

alex@lji.org

In breve

Grifoni et al. identificare potenziali bersagli per le risposte immunitarie al romanzo coronavirus 2019 (SARS-CoV-2) per omologia di sequenza con SARS-CoV strettamente **correlato e da a priori previsione dell'epitopo usando** approcci bioinformatici. Questa analisi fornisce informazioni essenziali per comprendere le risposte immunitarie umane a questo virus e per valutare i candidati diagnostici e vaccinali.



Grifoni et al., 2020, Cell Host & Microbe 27, 671–680, 8 aprile 2020 ^a 2020 Elsevier Inc. https://doi.org/10.1016/j.chom.2020.03.002



Teoria

Un'omologia di sequenza e un approccio bioinformatico possono prevedere target candidati per risposte immunitarie a SARS-CoV-2

Alba Grifoni, 1 John Sidney, 1 Yun Zhang, 2 Richard H. Scheuermann, 1, 2, 3 Bjoern Peters, 1, 4 e Alessandro Sette 1, 4, 5, *

1 Division of Vaccine Discovery, La Jolla Institute for Immunology, La Jolla, CA 92037, USA

2 J. Craig Venter Institute, La Jolla, CA 92037, USA

3 Dipartimento di Patologia, Università della California, San Diego, San Diego, CA 92093, USA

4 Dipartimento di Medicina, Università della California, San Diego, San Diego, CA 92093, USA

5 Contatto principale

*Corrispondenza: alex@lji.org

https://doi.org/10.1016/i.chom.2020.03.002

SOMMARIO

Contromisure efficaci contro la recente comparsa e la rapida espansione del nuovo coronavirus del 2019 (SARS-CoV-2) richiedono lo sviluppo di dati e strumenti per comprendere e monitorare la sua diffusione e le risposte immunitarie ad esso. Tuttavia, sono disponibili poche informazioni sugli obiettivi delle risposte immunitarie a SARS-CoV-2. Abbiamo usato il Database Epitope Immune e le risorse di analisi (IEDB) per catalogare i dati disponibili relativi ad altri coronavirus. Ciò include SARS-CoV, che presenta un'elevata somiglianza di sequenza con SARS-CoV-2 ed è il coronavirus meglio caratterizzato in termini di risposte epitopiche. Abbiamo identificato più regioni specifiche in SARS-CoV-2 che hanno un'elevata omologia del virus SARS-CoV. Pronostici bioinformatici paralleli identificati a priori potenziali epitopi di cellule B e T per SARS-CoV-2. L'identificazione indipendente delle stesse regioni utilizzando due approcci riflette l'alta probabilità che queste regioni siano obiettivi promettenti per il riconoscimento immunitario di SARS-CoV-2. Queste previsioni possono facilitare un'efficace progettazione del vaccino contro questo virus di alta priorità.

nam, India, Stati Uniti, Canada, Germania, Francia, Italia e Emirati Arabi Uniti. Questi numeri stanno cambiando rapidamente. Per informazioni aggiornate su COVID-19, consultare il sito Web dell'OMS all'indirizzo https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

Il database e analisi delle risorse immunitarie degli epitopi (IEDB) è un deposito di informazioni relative agli epitopi curate dalla letteratura scientifica nel contesto di **malattie infettive, allergie e autoimmunità (Vita et al., 2019). L'IEB fornisce anche** strumenti e algoritmi bioinformatici che consentono l'analisi dei dati degli epitopi e la previsione di potenziali epitopi da nuove sequenze. Virus Pathogen Resource (ViPR) è un repository complementare di informazioni sui virus patogeni umani che integra informazioni sul genoma, sul gene e sulla sequenza proteica con dati su epitopi immunitari, strutture proteiche e risposte dell'ospite alle infezioni virali (Pickett et al., 2012). Sono attualmente disponibili informazioni limitate su quali parti della sequenza SARS-CoV-2 sono riconosciute dalle risposte immunitarie umane. Tale conoscenza è di immediata rilevanza e aiuterebbe la progettazione del vaccino e faciliterebbe la valutazione dell'immunogenicità del candidato vaccino, nonché il monitoraggio delle potenziali conseguenze degli eventi mutazionali e della fuga dell'epitopo mentre il virus viene trasmesso attraverso le popolazioni umane.

Sebbene non siano ancora disponibili dati sugli epitopi per SARS-CoV-2, esiste un corpus significativo di informazioni sugli epitopi per i coronavirus in generale, e in particolare per *Betacoronaviruses*

come SARS-CoV e MERS-CoV, che causano malattie respiratorie nell'uomo (de Wit et al., 2016; Song et al., 2019). Qui, abbiamo usato le risorse IEDB e ViPR per compilare siti epitopici noti da altri coronavirus, mappare le regioni corrispondenti nelle sequenze SARS-CoV-2 e prevedere probabili epitopi. Abbiamo anche usato strumenti bioinformatici validati per predire gli epitopi delle cellule B e T che probabilmente saranno riconosciuti nell'uomo e per valutare la conservazione di questi epitopi tra le diverse specie di coronavirus.

INTRODUZIONE

Il 31 dicembre 2019, il Centro cinese per il controllo delle malattie (ChinaCDC) ha s riportato un gruppo di gravi casi di polmonite di eziologia sconosciuta nella città di p Wuhan, nella provincia cinese di Hubei. Poco dopo, i professionisti della sanità q pubblica hanno identificato il probabile agente causale come un romanzo *betacoronavirus*

(SARS-CoV-2). L'attuale epidemia, COVID-19, ha 81.109 casi confermati in tutto il mondo con 2.718 morti, a febbraio

26, 2020, secondo l'Organizzazione mondiale della sanità (OMS) in collaborazione con il CDC cinese e i centri di salute pubblica di altri paesi. Sebbene la maggior parte dei casi si sia verificata in Cina, un piccolo numero è stato confermato in altri 24 paesi, tra cui Giappone, Tailandia, Corea del Sud, Singapore, Vietnam.

RISULTATI

Una ricchezza di dati relativi ai coronavirus è disponibile nell'IEB

I coronavirus appartengono alla famiglia Coronaviradae, ordine nidovirales, e può essere ulteriormente suddiviso in quattro generi principali



SARS-CoV-2 orf	f1ab	S	ORF3a	E	М	ORF6	ORF7a	ORF8	N	ORF10
Bat-SL-CoV 95	5%	80%	91%	100%	98%	93%	88%	94%	94%	-
SARS-CoV 86	6%	76%	72%	94%	90%	68%	85%	40%	90%	-
MERS-CoV 50	0%	35%	-	36%	42%	-		-	48%	-

Figura 1. Confronto della struttura del genoma SARS-CoV-2 (Wuhan-Hu-1) con il suo parente di pipistrello più vicino (bat-SL-CoVZXC21), Tor2 SARS-CoV e HCoV-EMC MERS-CoV

Sopra: le regioni della sequenza di codifica (CDS) corrispondenti alle proteine omologhe tra i quattro virus sono riempite con lo stesso colore nello schema del genoma per indicare l'omologia; le regioni senza omologia delle proteine SARS-CoV-2 previste sono di colore bianco. Sotto: tabella delle somiglianze proteiche a coppie (espresse come% identità) tra SARS-CoV-2 e gli altri tre virus.

(Alfa-, Beta-, Gamma-, e Deltacoronaviruses). Parecchi

Alfa- e *Betacoronaviruses* causa lievi infezioni respiratorie e sintomi di raffreddore comuni nell'uomo, mentre altri sono zoonotici e infettano uccelli, maiali, pipistrelli e altri animali. Oltre alla SARS-CoV-2, altri due coronavirus, la SARS-CoV e la MERS-CoV, hanno causato gravi epidemie con alti tassi di mortalità (10% -30%) e società diffusa

urto

in caso di emergenza (Figura 1) (deWit et al., 2016; Song et al., 2019). La risposta immunitaria alla SARS-CoV-2 nell'uomo attende la caratterizzazione, ma sono state studiate le risposte immunitarie umane contro altri coronavirus. A partire dal 27 gennaio

Nel 2020, lo IEDB ha curato 581 lineari e 81 come epitopi discontinui di cellule B che sono stati riportati nella letteratura peerreviewed. Inoltre, 320 peptidi sono stati segnalati come epitopi delle cellule T (Tabella 1). La stragrande maggioranza di questi epitopi deriva da *Betacoronavirues*, e più precisamente da SARS-CoV, che da solo rappresenta oltre il 60% di essi. In termini di ospite in cui sono stati riconosciuti i vari epitopi delle cellule B e T (Tavolo 2), la maggior parte degli epitopi (B o T) sono stati definiti nell'uomo o nei sistemi murini. In particolare, tutti tranne 2 di

gli 417 epitopi delle cellule B e T descritti nell'uomo provengono da *Betacoronaviruses*, di cui 398 provenienti dalla SARS-CoV.

SARS-CoV-2 Somiglianza con altri Betacoronavirus

Il confronto tra una sequenza di proteine SARS-CoV-2 consensuale e sequenze per SARS-CoV, MERS-CoV e bat-SL-CoVZXC21 ha rivelato un alto grado di somiglianza (espressa come identità%) tra SARS-CoV-2, bat-SL- CoVZXC21 e SARS-CoV, ma una somiglianza più limitata con MERS-CoV (Figura 1). Ciò è in accordo con un recente documento pubblicato il 7 febbraio 2020 che mostra la massima somiglianza tra SARS-CoV-2 e SARS o SARS simili a SARS (Wu et al., 2020). Inoltre, SARS-Cov è il virus correlato più vicino a SARS-CoV-2 per il quale è stato definito un numero significativo di epitopi nell'uomo (e in altre specie) e che causa anche malattie umane con esiti letali. Di conseguenza, nelle seguenti analisi, ci siamo concentrati sul confronto tra sequenze di epitopi SARS-CoV note e sequenza SARSCoV-2.

Abbiamo prima valutato la distribuzione degli epitopi derivati dalla SARS-CoV in funzione della proteina di origine (Tabella 3). Nel contesto

Tabella 1. Inventario IEDB di epitopi di coronavirus B e T								
Set di epitopi	genere	Coronavirus	Coronavirus					
		Alfa	Beta	Beta				
			SARS-CoV	MERS-CoV	Altro			
Cellula B.	conformazionale	18	27	23	2	11	81	
	Lineare	81	405	5	60	30	581	
Cellula T.		61	164	25	54	16	320	

delle risposte delle cellule B, la maggior parte dei 12 antigeni nel proteoma SARS-CoV sono associati agli epitopi, con il maggior numero derivato da glicoproteina di picco, nucleoproteina e proteina di membrana (Tabella 3). La scarsità di epitopi di cellule B associati ad altre proteine è probabilmente perché, in media, gli studi di screening degli epitopi di cellule B finora hanno regioni sondate che costituiscono meno del 20% di ciascuna rispettiva sequenza, incluso <1% della polipoteina di Orf 1ab. In confronto, la gamma completa delle sequenze di glicoproteina spike, nucleoproteina e proteine di membrana è stata analizzata almeno in una certa misura nei test sulle cellule B.

Una situazione simile è stata osservata nel caso degli epitopi delle cellule T. Qui, abbiamo considerato solo epitopi il cui riconoscimento è limitato dal maggiore complesso di istocompatibilità (MHC) dell'antigene leucocitario umano (HLA), poiché il polimorfismo dell'MHC si traduce in genere nel riconoscimento di diversi epitopi nell'uomo e nei topi.

Definire le regioni immunodefinenti all'interno del genoma SARSCoV

Gli epitopi delle cellule B derivati da SARS-CoV sono stati mappati indietro a una sequenza di riferimento SARS-CoV utilizzando lo strumento Immunobrowser dello IEDB (Dhadatla membrana sono stati maggiormente conservati (8/10 e 2/3, rispettivamente, et al., 2018). Questo strumento combina tutti i record disponibili lungo una sequenza di riferimento e produce un punteggio del fattore di risposta (RF) che tiene conto del tasso di positività (frequenza con cui è stato trovato un residuo in un epitopo positivo) e del numero di record (come vengono riportati molti saggi indipendenti). Le regioni dominanti sono state identificate considerando i tratti in cui si trovava il punteggio RF R 0.3.

Sono mostrate le analisi della glicoproteina a spillo, delle proteine di membrana e delle nucleoproteine figura 2. Nel caso della glicoproteina a spillo (figura 2 A), identifichiamo cinque regioni di potenziale interesse (residui 274–306, 510–586, 587–628, 784–803 e 870–893), tutti rappresentanti delle regioni associate ad alti tassi di risposta immunitaria. Tre di queste regioni immunodominanti si trovano nella subunità S1 nel CTD2 e CTD3 (dominio C-terminale), mentre le altre due si trovano nel dominio HR1 della subunità S2.

Successivamente, abbiamo allineato le sequenze della regione dell'epitopo delle cellule B SARS-CoV alla sequenza SARS-CoV-2 per calcolare l'identità percentuale tra ciascuna delle regioni dominanti SARS-CoV e SARS-CoV-2 (Tabella 4). Delle 10 regioni identificate, 6 avevano il 90% o più di identità con SARS-CoV-2, 2 erano identiche tra l'80% e l'89% e 2 avevano un'omologia inferiore ma ancora apprezzabile (69% e 78%).

In un'analisi simile, è stato anche scoperto che gli epitopi delle cellule T sono prevalentemente associati alla glicoproteina a spike e alla nucleoproteina (Tabella 3). Tabella 5 mostra un elenco degli epitopi individuali SARS-CoV più dominanti finora identificati negli esseri umani. Abbiamo anche allineato le sequenze di epitopi delle cellule T SARS-CoV e calcolato per ciascun epitopo l'identità percentuale a SARSCoV-2. Per ogni epitopo di cellule T, Tabella 5 mostra l'antigene di origine, la sequenza epitopica, la sequenza omologa SARS-CoV-2 e la percentuale corrispondente dell'identità della sequenza. Complessivamente, la fosfoproteina nucleocapside e gli epitopi derivati idalla membrana sono stati maggiormente conservati (8/10 e 2/3, rispettivamente, avevano R 85% di identità con SARS-CoV-2). Gli epitopi di Orf1ab e della glicoproteina di superficie sono stati moderatamente conservati (rispettivamente 3/7 e 10/23 R L'identità dell'85% con SARSCoV-2) e gli epitopi di Orf 3a erano i meno conservati.

Predizione degli epitopi delle cellule B SARS-CoV-2

Per definire i potenziali epitopi delle cellule B con un metodo alternativo, abbiamo utilizzato gli strumenti predittivi forniti con l'IEDB. Le previsioni sugli epitopi delle cellule B sono state eseguite utilizzando la superficie SARS-CoV-2

Tabella 2. Inventario IEDB di epitopi di coronavirus B e T								
Set di epitopi	Ospite	Coronavirus B						
	Alfa Beta					Gamma		
			SARS-CoV	MERS-CoV	Altro			
Cellula B. 🖛	Gli esseri umani	0	306	16	0	0	322	
	Торі	62	154	9	58	20	303	
	Altro	42	142	5	6	23	218	
	Торі Тд	0	0	0	0	0	0	
Cellula T.	Gli esseri umani	2	92	0	1	0	95	
	Торі	16	99	25	53	1	194	
	Altro	46	1	0	0	15	62	
	Торі Тд	0	29	0	0	0	29	

un La cellula B include sia epitipi conformazionali che lineari.

B Totali tra Tabelle 1 e 2 potrebbe non essere uguale poiché diversi epitopi sono riconosciuti in più specie.

Tabella 3. Inventario IEDB di epitopi di coronavirus B e T							
Proteine SARS-CoV	B Cell	Cellula T.					
Glicoproteina Spike	279	48					
nucleoprotein	113	33					
Proteina di membrana	20	4					
Replicase polyprotein 1ab	8	9					
Proteina 3a	2	7					
Busta piccola proteina di membrana	2	0					
Proteina non strutturale 3b	2	0					
Proteina 7a	2	0					
Proteine 9b	2	0					
Proteine non strutturali 6	1	0					
Proteine non strutturali 8a	1	0					
U totala da ali azitazi dalla azili da Tarazzanda azitazi							

Il totale degli epitopi delle cellule T comprende epitopi riconosciuti nell'uomo e / o nei topi transgenici.

glicoproteina, fosfoproteina nucleocapside e seguenze di glicoproteina di membrana. che, come descritto sopra, sono state trovate essere i principali target proteici per le risposte delle cellule B ad altri coronavirus. Parallelamente, abbiamo eseguito previsioni per epitopi di cellule B lineari con Bepipred 2.0 (Jespersen et al., 2017) e per epitopi conformazionali con Discotope 2.0 (Kringelum et al., 2012). Entrambi gli algoritmi di predizione sono disponibili nella pagina dello strumento di predizione delle cellule IEDB B (http://tools.iedb.org/main/bcell/). Viene fornito un elenco completo dei risultati di predizione dell'epitopo di cellule B per posizione di aminoacidi per proteina Tabella S1 . Utilizzando Bepipred 2.0 e un limite di R 0,55 (corrispondente a

una soglia specifica dell'80%) (Jespersen et al., 2017), la glicoproteina di superficie presentava il numero più elevato di epitopi previsti per le cellule B, seguita dalla glicoproteina di membrana e dalla fosfoproteina nucleocapsidica (Tabella S2). Per prevedere e mappare gli epitopi delle cellule B conformazionali, abbiamo usato la struttura della glicoproteina con picco SARS-CoV-2 recentemente presentata (PDB: 6VSB). Un elenco di posizioni di aminoacidi glicoproteici di superficie con un'alta probabilità di essere inclusi negli epitopi delle cellule B previsti, basato sull'analisi con l'algoritmo Discotope 2.0, è mostrato in Tabella S1 (cutoff di R 2.5, corrispondente all'80% della specifica). Abbiamo quindi localizzato le posizioni rilevanti degli aminoacidi sulla struttura del modello, che ha permesso l'identificazione di sette residui / regioni di epitopi previsti (491-505, 558-562, 703-704, 793-794, 810, 914 e 1140-1146) nella glicoproteina di superficie (Figura 3).

Predizione degli epitopi delle cellule T SARS-CoV-2

Per predire gli epitopi delle cellule T CD4, abbiamo usato il metodo descritto da Paul e co-autori (Paul et al., 2015a), come implementato nella risorsa Tepitool in IEDB (Paulquelle previste da BebiPred 2.0. Come detto sopra, nessuna previsione Discotope et al., 2016). Questo approccio è stato progettato e validato per prevedere gli epitopi dominanti indipendentemente dall'etnia e dal polimorfismo HLA, sfruttando l'ampia reattività crociata e il repertorio che si sovrappongono tra i diversi loci HLA di classe Il e le varianti alleliche. Qui, abbiamo selezionato peptidi con un percentuale mediana di consenso% 20, una soglia associata ai pannelli epitopici responsabili di circa il 50% delle risposte specifiche del target. Usando questa soglia, abbiamo identificato 241 candidati nella sequenza SARS-CoV-2 (vedi

In esperimenti precedenti, abbiamo dimostrato che i pool basati su numeri peptidici simili possono essere generati dalla liofilizzazione sequenziale (Carrasco Pro et al., 2015). Questi pool di peptidi (o megapool) incorporano epitopi previsti o validati sperimentalmente e consentono la misurazione dell'entità e della caratterizzazione del fenotipo delle risposte delle cellule T umane nelle indicazioni di malattie infettive come Bordetella pertussis. Mycobacteria tuberculosis. Dengue e virus Zika (Carrasco Pro et al., 2015; da Silva Antunes et al., 2018; Grifoni et al., 2017, 2018). La megapool CD4 SARS-CoV-2 copre tutte e 10 le proteine previste, con il numero di potenziali epitopi proporzionale alla dimensione di ciascuna proteina (Tabella S4).

Parallelamente, abbiamo anche cercato di definire probabili epitopi CD8. In questo caso, è stato richiesto un approccio diverso poiché la sovrapposizione tra diverse varianti alleliche e loci HLA di classe I è più limitata ai gruppi specifici di alleli o supertipi (Sidney et al., 2008). Seguendo un approccio precedentemente validato (Weiskopf et al., 2013), abbiamo assemblato una serie dei 12 più importanti alleli HLA di classe I che hanno dimostrato di consentire un'ampia copertura della popolazione generale, come descritto nella Metodi STAR (Guarda anche Tabella S5). Abbiarno quindi eseguito previsioni di associazione HLA di classe I utilizzando l'algoritmo Net MHC pan 4.0 EL (Jurtz et al., 2017) disponibile presso lo IEDB. Per ogni allele, abbiamo selezionato i peptidi con il punteggio superiore dell'1% nella sequenza SARS-CoV-2, in base alla previsione. Dopo aver eliminato ridondanze e peptidi nidificati, abbiamo ottenuto un risultato finale " in silico " megapool di 628 epitopi unici previsti. Tabella S6 elenca quegli epitopi previsti unici per proteina, indicando per ciascuno le rispettive restrizioni HI A

Corrispondenza tra gli epitopi identificati dai due diversi approcci

Gli epitopi identificati dall'omologia agli epitopi SARS-CoV definiti sperimentalmente mostrati in Tabelle 4 e 5 sono stati successivamente confrontati con gli epitopi identificati dalle previsioni degli epitopi mostrate in Tabelle S2 , S3 , e S6 . Si presume che gli epitopi identificati in modo indipendente in entrambi gli approcci siano i contatti più preziosi.

Abbiamo prima confrontato le regioni immunodominanti delle cellule B identificate in SARS-CoV e mappate alle omologhe proteine SARSCoV-2 (Tabella 4), con il lineare previsto (Tabella S2) e conformazionale (Tabella S1) Epitopi delle cellule B. Delle cinque regioni immunodominanti delle cinque cellule B dalla glicoproteina con picco SARS che sono state mappate su SARS-CoV-2, tre regioni si sono sovrapposte a quelle identificate da BebiPred 2.0 e due si sono sovrapposte alle regioni previste da Discotope 2.0 (Figura 3;

Tabella S1). Non sono state osservate sovrapposizioni per le cinque regioni della proteina di membrana SARS-CoV e la nucleoproteina mappate su SARS-CoV-2 e 2.0 era disponibile per quelle due proteine.

L'analisi di predizione eseguita con Discotope 2.0 basata sulla struttura PDB della glicoproteina con picco SARS-CoV-2 conferma in modo indipendente due delle probabili regioni epitopiche definite sulla base dei dati SARS-CoV. In particolare, un epitopo dominante corrisponde all'epitopo 524-599 di Tabella 5, che si sovrappone all'epitopo previsto 558-562, e anche la regione 802-819 è prevista (cfr. il residuo 810 previsto si trova nel terreno di questa regione). Infine, la regione 888-909 è mancata da poco,

Proteina SARS-CoV Spike (ID: P59594)



Figura 2. Regioni immunodominanti delle cellule B basate sulla mappatura epitopica specifica SARS È stato calcolato il punteggio RF per ciascuna posizione dell'amminoacido (vedere Metodi STAR) e tracciato sulla sequenza di consenso SARS-CoV di glicoproteina spike (A), proteina di membrana (B) e nucleoproteina (C).

perché il residuo 914, che è previsto, è proprio fuori dall'epitopo.

Quando abbiamo confrontato gli epitopi delle cellule T SARS-CoV mappati a SARS-CoV-2 (Tabella 5) con gli epitopi previsti delle cellule T CD4 e CD8 (Tabelle S3 e S6, rispettivamente), abbiamo scoperto che 12 di 17 epitopi di cellule T SARS-CoV-2 con identità ad alta sequenza (R 90%) al SARS-CoV sono stati identificati indipendentemente dai due metodi. Altri 7 epitopi su 16 con identità di sequenza moderata (70% -89%) e 6 su 12 epitopi con identità di sequenza bassa (<70%) sono stati identificati con entrambi i metodi. La mancanza di corrispondenza assoluta non è sorprendente, dato che i dati sperimentali sono derivati da un insieme distorto di restrizioni HLA (in gran parte HLA A * 02: 01) e che la nostra previsione di classe I HLA la strategia mirava a una serie più limitata di alleli selezionati per rappresentare le varianti mondiali più frequenti; allo stesso tempo, le previsioni di classe II dovrebbero coprire il 50% delle risposte di classe II (Paul et al., 2015b).

DISCUSSIONE

Il presente studio identifica i probabili target della risposta immunitaria umana alla SARS-CoV-2, includendo sia i bracci delle cellule B che T della risposta immunitaria adattativa. Ciò è rilevante di fronte alla sempre maggiore urgenza medica e sociale che circonda COVID-19, soprattutto data l'attuale scarsità di dati sperimentali riguardanti qualsiasi immunitario corrispondente

Host cellulare e microbo 27, 671-680, 8 aprile 2020 675

Tabella 4. Regioni di epitopi cellulari dominanti SARS-CoV B.						
SARS-CoV		SARS-CoV-2				
Sequenza	Max RF	Sequenza	Proteina un	Inizio-Fine mappato	Identità (%)	
DAVDCSQNPLAELKCSVKSFEIDK GIYQTSNF	0,504	DAVDCALDPLSETKCTLKS FTVEKGIYQTSN	S	287-317	69	
VCGPKLSTDLIKNQCVNFNFNGL TGTGVLTPSSKRFQPFQQFGRD VSDFTDSVRDPKTSEILDISPCSF GGVSVIT	0,745	VCGPKKSTNLVKNKCVNFNFN GLTGTGVLTESNKKFLPFQQF GRDIADTTDAVRDPQTLEILDI TPCSFGGVSVI	S	524-598	80	
GTNASSEVAVLYQDVNCTDVSTA IHADQLTPAWRIYSTGNN	0,709	GTNTSNQVAVLYQDVNCTEVPVA IHADQLTPTWRVYSTGS	S	601-640	78	
FSQILPDPLKPTKRSFIED	0,365	FSQILPDPSKPSKRSFIE	S	802-819	89	
FGAGAALQIPFAMQMAYRFNGIG	0.367	FGAGAALQIPFAMQMAYRFNGI	S	888-909	100	
MADNGTITVEELKQLLEQWNLVIG	0.460	MADSNGTITVEELKKLLEQWNLVI	М	1-24	92	
PLMESELVIGAVIIRGHLRMA	0,457	PLLESELVIGAVILRGHLRI	М	132-151	90	
PQGLPNNTASWFTALTQHGKEE	0,537	RPQGLPNNTASWFTALTQHGK	N	42-62	95	
NNAATVLQLPQGTTLPKGFYA	0,543	NNNAATVLQLPQGTTLPKGF	N	153-172	95	
KHIDAYKTFPPTEPKKDKKKKTDEAQ PLPQRQKKQPTVTLLPAADMDD	0.82	NKHIDAYKTFPPTEPKKDKKKKTD EAQPLPQRQKKQPTVTLLPAADM	Ν	355-401	90	
S, glicoproteina di superficie; M, proteina di membra	ana; N, fosfoproteir	na nucleocapside				

risposta. L'approccio che abbiamo seguito si basa sulla determinazione di diverse linee di evidenza che individuano chiaramente SARS-CoV come un modello rilevante per estrapolare probabili target di risposte a SARS-CoV-2, il virus associato a COVID-19

La prima linea di evidenza riguarda il fatto che dei coronavirus noti per infettare l'uomo, SARS-CoV è il più simile in termini filogenetici a SARS-CoV-2. La seconda linea di prova è che SARS-CoV-2 è il più (e altamente) simile a SARS-CoV a livello di identità di sequenza. In terzo luogo, quando abbiamo rivisto criticamente le conoscenze relative agli epitopi precisi riconosciuti dalle risposte adattative nel contesto dei coronavirus in forma aggregata, era evidente che tutti tranne 2 degli epitopi 417 B e T descritti finora nell'uomo provengono da

Betacoronaviruses, di cui 398 provenienti dalla SARS-CoV.

La nostra analisi ha mostrato che alcune regioni SARS-CoV erano dominanti per le risposte delle cellule B e che quelle regioni erano ben conservate in termini di sequenza con SARS-CoV-2. Cinque regioni contengono epitopi riconosciuti dagli anticorpi neutralizzanti nei sieri convalescenti della SARS (Guo et al., 2004; Shichijo et al., 2004). Tra questi, di particolare interesse è la regione 587-628 che nidifica la 604 625 rispettive proteine, il che ha reso più difficile l'identificazione di regioni epitopiche peptidi, che sono stati identificati in un paziente convalescente SARS e hanno scoperto di avere la capacità di suscitare anticorpi che prevengono efficacemente l'infezione nei primati non umani (Hu et al., 2005; Wang et al., 2016).

A causa dell'elevato livello complessivo di somiglianza di seguenza di SARS-CoV e SARS-CoV-2. deduciamo che le regioni dominanti in SARS-CoV hanno un'alta probabilità di essere anche dominanti in SARS-CoV-2, anche se le seguenze effettive sono diversi. Questa ipotesi è in accordo con la recente struttura della microscopia crioelettronica (cryoEM) della glicoproteina a spillo di SARSCoV-2, che mostra un'alta somiglianza nella struttura generale con la proteina spike SARS-CoV (Wrapp et al., 2020). Nello stesso studio, tuttavia, gli autori non osservano il riconoscimento incrociato di anticorpi monoclonali SARS-CoV con SARSCoV-2. In effetti, non hanno osservato alcuna reattività con gli anticorpi SARS-CoV che riconoscono il dominio di legame del recettore spike SARS-CoV-2 (RBD), nonostante il fatto che SARS-CoV-2 mantenga la stessa capacità di legare il recettore ACE2 di SARS-CoV (Wrapp et al., 2020). Ciò suggerisce che la previsione delle cellule B eseguita sul dominio RBD richiederà ulteriori studi.

Abbiamo anche analizzato gli epitopi delle cellule T SARS-CoV. In guesti casi, le regioni epitopiche e i singoli epitopi sono stati più ampiamente dispersi tra le discrete e dominanti. Questo risultato non è inatteso dato che le cellule T riconoscono i peptidi corti generati dall'elaborazione cellulare di antigeni virali che possono essere derivati da qualsiasi segmento della proteina.

2 B) e tre regioni sono state identificate per la nucleoproteina (43-65, 154-175 e 356-404) (figura 2 C). È stato dimostrato che le due regioni della proteina di membrana suscitano risposte IgM e IgG marcate e un ampio spettro di riconoscimento, evidenziandole come potenziali candidati diagnostici Chow et al., 2006; Wang et al., 2003). Delle tre regioni identificate nella nucleoproteina, 156-175 ha mostrato una forte reattività nei confronti dei sieri dei pazienti con SARS e l'immunogenicità in più specie, tra cui topi, scimmie e umani (Liu et al., 2006).

Due regioni sono state identificate dalla proteina di membrana (1-25 e 131-152) (figura Si prevede generalmente che gli epitopi delle cellule T CD8 saranno derivati da entrambe le proteine strutturali e non strutturali (Tian et al., 2019), poiché entrambi i tipi di proteine vengono elaborati endogenamente da cellule infette. Nel caso degli epitopi di classe II, le proteine strutturali sarebbero di particolare interesse, poiché molto probabilmente forniranno aiuto attraverso l'interazione cognitiva (Sette et al., 2008). Quando si esaminano le regioni omologhe di SARS-CoV, è stato riscontrato che i probabili epitopi delle cellule T sono positivi in test come ELISPOT, colorazione intracellulare (ICS) e colorazione multimer / tetramero (vedere, ad esempio, Cheung et al., 2007, 2008; Kohyama et al., 2009; Tsao et al., 2006; Yang et al., 2009).

Tabella 5. Epitopi dominanti di cellule T SARS-CoV							
SARS			SARS-CoV-2				
Sequenza	Limitazione HLA de	el punteggio RF ur	Sequenza Inizio-Fine m		opato alle proteine	Identità (%)	
VRGWVFGSTMNNKSQSVI	0.15	DRB1 * 04: 01	IRGWIFGTTLDSKTQSLL	S	101-118	50	
CTFEYISDAFSLD	0,21	DRB1 * 04: 01	CTFEYVSQPFLMD	S	166-178	62	
DAFSLDVSEKSGN	0.62	DRB1 * 04: 01	QPFLMDLEGKQGN	S	173-185	38	
TNFRAILTAFSPAQDIW	0,32	DRB1 * 04: 01	TRFQTLLALHRSYLTPGD SSSGW	S	236-258	17	
KSFEIDKGIYQTSNFRVV	0.40	DRB1 * 04: 01, DRB1 * 07: 01	KSFTVEKGIYQTSNFRVQ	S	304-321	78	
STFFSTFKCYGVSATKL	0.50	DRB1 * 07: 01, DR8	SASFSTFKCYGVSPTKL	S	371-387	82	
KLPDDFMGCV	0.55	A * 02: 01	KLPDDFTGCV	S	424-433	90	
NIDATSTGNYNYKYRYLR	0,29	Classe II	NLDSKVGGNYNYLYRLFR S		440-457	56	
YLRHGKLRPFERDISNVP	0.16	DRB1 * 04: 01	YLYRLFRKSNLKPFERDI	S	451-468	58	
RPFERDISNVPFS	0,36	DRB1 * 04: 01	KPFERDISTEIYQ	S	462-474	54	
KSIVAYTMSLGADSSIAY	0.15	DRB1 * 04: 01, DRB1 * 07: 01	QSIIAYTMSLGAENSVAY	S	690-707	72	
SIVAYTMSL	0,29	A * 02: 01	SIIAYTMSL	S	691-699	89	
TECANLLLQYGSFCTQL	0.50	DR8	TECSNLLLQYGSFCTQL	S	747-763	94	
VKQMYKTPTLKYFGGFNF	0.20	DRB1 * 04: 01	VKQIYKTPPIKDFGGFNF	S	785-802	78	
ESLTTTSTALGKLQDVV	0.42	DRB1 * 04: 01	DSLSSTASALGKLQDVV	S	936-952	71	
ALNTLVKQL	0,29	A * 02: 01	ALNTLVKQL	S	958-966	100	
VLNDILSRL	0,29	A * 02: 01	VLNDILSRL	S	976-984	100	
LITGRLQSL	0.42	A * 02: 01	LITGRLQSL	s	996-1004	100	
QLIRAAEIRASANLAATK	0.20	DRB1 * 04: 01	QLIRAAEIRASANLAATK	S	1011-1028	100	
SWFITQRNFFSPQII	0.60	DRB1 * 04: 01	HWFVTQRNFYEPQII	S	1101-1115	73	
RLNEVAKNL	0.42	A * 02: 01	RLNEVAKNL	S	1185-1193	100	
NLNESLIDL	0,29	A * 02: 01	NLNESLIDL	S	1192-1200	100	
FIAGLIAIV	0,80	A * 02: 01	FIAGLIAIV	S	1220-1228	100	
RFFTLGSITAQPVKI	0,18	B * 58: 01	RIFTIGTVTLKQGEI	Orf 3a	6-20	40	
SITAQPVKI	0,29	B * 58: 01	TVTLKQGEI	Orf 3a	12-20	22	
TLACFVLAAV	0.59	A * 02: 01	TLACFVLAAV	М	61-70	100	
GLMWLSYFV	0.59	A * 02: 01	GLMWLSYFI	M	89-97	89	
HLRMAGHSL	0.40	Classe I.	HLRIAGHHL	М	148-156	78	
ALNTPKDHI	0,29	A * 02: 01	ALNTPKDHI	N	138-146	100	
LQLPQGTTL	0,29	A * 02: 01	LQLPQGTTL	N	159-167	100	
GETALALLLL	0.38	B * 40: 01	GDAALALLLL	N	215-224	80	
LALLLLDRL	0,29	A * 02: 01	LALLLDRL	N	219-227	100	
LLLDRLNQL	0.42	A * 02: 01	LLLDRLNQL	N	222-230	100	
RLNQLESKV	0.42	A * 02: 01	RLNQLESKM	N	226-234	89	
TKQYNVTQAF	0,29	Classe I.	TKAYNVTQAF	N	265-274	90	
GMSRIGMEV	0.42	A * 02: 01	GMSRIGMEV	N	316-324	100	
MEVTPSGTWL	0.42	B * 40: 01	MEVTPSGTWL	N	322-331	100	
	0.50	A * 24: 02	NEKDQVILI	N	345-353	78	
CLDAGINYV	0.42	A * 02: 01	CLEASFNYL	Orf 1ab	2139-2147	56	
WLMWFIISI	0.42	A * 02: 01	WLMWLIINL	Orf 1ab	2292-2300	67	
ILLLDQVLV	0.42	A * 02: 01	ILLLDQALV	Orf 1ab	2498-2506	89	
	0.42	A * 02: 01	SACVLAAEC	Orf 1ab	2840-2848	56	
ALSGVECGV	0.42	A * 02· 01	SLPGVECGV	Orf 1ab	2942-2950	78	
	0.42	A * 02: 01		Orf 1ab	3639-3647	89	
SMWALVISV	0.42	A * 02: 01	SMWALIISV	Orf 1ab	3661-3669	89	

S, glicoproteina di superficie; M, proteina di membrana; N, fosfoproteina nucleocapside. wr Le restrizioni definite solo nei topi transgenici HLA sono indicate dal carattere corsivo.



Figura 3. Glicoproteina con punte SARS-CoV-2 (PDB: 6VSB)

La superficie calcolata dei primi 13 residui di aminoacidi previsti come epitopi di cellule B in base alla classifica eseguita con Discotope 2.0 è mostrata in rosso. Il monomero è mostrato in alto a sinistra. In alto a destra e in basso al centro presentano il rifinitore in due diversi orientamenti. Il rendering 3D è stato eseguito utilizzando YASARA (Krieger e Vriend, 2014).

Abbiamo anche cercato di affrontare i potenziali epitopi SARS-CoV-2 con un metodo completamente diverso, vale a dire utilizzando le previsioni degli epitopi ospitate dallo IEDB (Dhanda et al., 2019; Vita et al., 2019). Per gli epitopi delle cellule B, abbiamo usato metodi che prevedono epitopi lineari (Jespersen et al., 2017) e nel caso della glicoproteina a spillo in cui recentemente è diventata disponibile una struttura affidabile (Wrapp et al., 2020), Discotope 2.0 (Kringelum et al., 2012) metodo che prevede anche epitopi in base alla conformazione proteica e all'esposizione ai residui. La previsione Discotope ha confermato in modo indipendente due delle probabili regioni epitopiche definite sulla base dei dati SARS-CoV.

Nel caso degli epitopi delle cellule T, abbiamo utilizzato algoritmi predittivi (Jurtz et al., 2017; Paul et al., 2016) per mappare centinaia di potenziali epitopi umani per tenere conto del polimorfismo dell'HLA e del fatto che gli epitopi delle cellule T sono tipicamente derivati da proteine strutturali e non strutturali e non si limitano alle regioni esposte. Qui, come convalida indipendente delle previsioni, abbiamo chiesto se le previsioni identificassero effettivamente i relativamente pochi epitopi identificati sperimentalmente in SARS-CoV, limitato dall'HLA umana e conservato in SARS-CoV-2. In effetti, abbiamo scoperto che 12 di 17 epitopi di cellule T SARS-CoV-2 con identità ad alta sequenza (R 90%) al SARS-CoV sono stati identificati in modo indipendente dalle previsioni degli epitopi basate sulle sequenze SARS-CoV-2.

In conclusione, l'uso delle informazioni disponibili relative agli epitopi SARS-CoV in combinazione con previsioni bioinformatiche indica regioni specifiche di SARS-CoV-2 che hanno un'alta probabilità di essere riconosciute dalle risposte immunitarie umane. L'osservazione che molti epitopi di cellule B e T sono altamente conservati tra SARS-CoV-2 e SARS-CoV è importante. Potrebbero essere le strategie di vaccinazione progettate per indirizzare la risposta immunitaria verso queste regioni di epitopi conservate generare un'immunità che non è solo cross-protettiva Betacoronaviruses ma anche relativamente resistente all'evoluzione in corso del virus.

STAR + METODI

I metodi dettagliati sono forniti nella versione online di questo documento e includono quanto seque:

a TABELLA DELLE RISORSE CHIAVE
d DISPONIBILITÀ DEI CONTATTI E DEI MATERIALI
d DETTAGLI DEL METODO
B Analisi IEDB degli epitopi di Coronavirus T e B.
B Confronto tra sequenze di coronavirus e SARS-CoV-2
B Determinazione del consumo di sequenza SARS-CoV-2vazione
B Previsione dell'epitopo delle cellule B SARS-CoV-2
B Previsione dell'epitopo delle cellule T SARS-CoV-2
d QUANTIFICAZIONE E ANALISI STATISTICA
d DISPONIBILITÀ DI DATI E CODICI

Ulteriori informazioni sono disponibili online all'indirizzo https://doi.org/10.1016/j. chom.2020.03.002 .

RINGRAZIAMENTI

Ringraziamo Erica Ollmann Saphire, Sharon Schendel e Mitchell Kronenberg per la lettura critica del manoscritto e numerosi suggerimenti utili. Ringraziamo anche Jason McLellan e Barney Graham per l'accesso anticipato alla struttura PDB della glicoproteina con spike SARS-2. Supporto per il lavoro incluso finanziamento fornito tramite contratti NIH-NIAID 75N9301900065 (AS) e 75N93019C00001 (AS e BP). Ulteriore supporto è stato fornito dal contratto NIHNIAID 75N93019C00076 (YZ e RHS).

CONTRIBUTI D'AUTORE

AG, JS e YZ hanno eseguito analisi e scritto il documento; RHS ha scritto il documento e fornito indicazioni per le analisi virali; BP ha ideato il progetto, scritto il documento e fornito esperienza ai componenti bioinformatici; e

AS ha ideato il progetto, scritto il documento e fornito la direzione generale per lo studio.

DICHIARAZIONE DI INTERESSI

L'Istituto di Immunologia della Jolla (LJI) ha presentato una domanda di brevetto relativa a questo manoscritto.

Ricevuto: 13 febbraio 2020 Revisionato: 26 febbraio 2020 Accettato: 5 marzo 2020

26 febbraio 2020 Accettato: 5 marzo 2020 Pubblicato: 16 marzo 2020

RIFERIMENTI

Carrasco Pro, S., Sidney, J., Paul, S., Lindestam Arlehamn, C., Weiskopf, D., Peters, B. e Sette, A. (2015). Generazione automatica di set di epitopi specifici convalidati. J. Immunol. Res. 2015, 763.461.

Cheung, YK, Cheng, SC, Sin, FW, Chan, KT e Xie, Y. (2007). Induzione della risposta delle cellule T da parte di un vaccino a DNA codificante per un epitopo coronavirus della sindrome respiratoria acuta grave HLA-A * 0201 nuovo. Vaccino 25, 6070-6077.

Cheung, YK, Cheng, SC, Sin, FW, Chan, KT e Xie, Y. (2008). Studio degli epitopi immunogenici delle cellule T nella proteina nucleocapsidica del virus SARS e del loro ruolo nella prevenzione e nel trattamento dell'infezione da SARS. Hong Kong Med. J. 14 (Suppl 4), 27-30.

Chow, SC, Ho, CY, Tam, TT, Wu, C., Cheung, T., Chan, PK, Ng, MH, Hui, PK, Ng, HK, Au, DM e Lo, AW (2006). Epitopi specifici delle proteine strutturali e ipotetiche suscitano risposte umorali variabili nei pazienti con SARS. J. Clin. Pathol. *59*, 468-476.

da Silva Antunes, R., Paul, S., Sidney, J., Weiskopf, D., Dan, JM, Phillips, E., Mallal, S., Crotty, S., Sette, A. e Lindestam Arlehamn, CS (2018). Correzione: definizione di epitopi umani riconosciuti nel tossoide tetano e sviluppo di una strategia di analisi per rilevare le risposte delle cellule T CD4 + Ex Vivo Tetanus. PLoS One *13*, e0193382 area della cellule T CD4 + Ex Vivo Tetanus. PLoS One *13*, e0193382

de Wit, E., van Doremalen, N., Falzarano, D., and Munster, VJ (2016). SARS e MERS: recenti approfondimenti sui coronavirus emergenti. Nat. Rev. Microbiol. 14. 523-534.

Dhanda, SK, Mahajan, S., Paul, S., Yan, Z., Kim, H., Jespersen, MC, Jurtz, V., Andreatta, M., Greenbaum, JA, Marcatili, P., Sette, A., Nielsen, M. e Peters, B. (2019). IEDB-AR: risorsa di analisi del database di epitopi immuni in

2019, Ris, Di acidi nucleici, 47, W502-W506 .

Dhanda, SK, Vita, R., Ha, B., Grifoni, A., Peters, B. e Sette, A. (2018). ImmunomeBrowser: uno strumento per aggregare e visualizzare epitopi complessi ed eterogenei nelle proteine di riferimento. Bioinformatica *34*, 3931-3933

Grifoni, A., Angelo, MA, Lopez, B., O'Rourke, PH, Sidney, J., Cerpas, C., Balmaseda, A., Silveira, CGT, Maestri, A., Costa, PR, et al. (2017). Valutazione globale di Dengue Virus-Speci fi c CD4 + Risposte alle cellule T nelle aree endemiche della dengue. Davanti. Immunol. *8*, 1309.

Grifoni, A., Costa-Ramos, P., Pham, J., Tian, Y., Rosales, SL, Seumois, G., Sidney, J., de Silva, AD, Premkumar, L., Collins, MH, et al. (2018). All'avanguardia: il processo di trascrizione rivela le risposte antivirali multifunzionali e citotossiche di Zika Virus-Speci fi c CD8 + Cellule T. J. Immunol. 201,

3487-3491.

Guo, JP, Petric, M., Campbell, W. e McGeer, PL (2004). Peptidi del virus corona SARS riconosciuti dagli anticorpi nei sieri dei casi convalescenti. Virologia 324, 251-256. Hu, H., Li, L., Kao, RY, Kou, B., Wang, Z., Zhang, L., Zhang, H., Hao, Z., Tsui,

WH, Ni, A., et al. (2005). Lo screening e l'identificazione di epitopi lineari a cellule B e peptidi con blocco dell'entrata della sindrome respiratoria acuta grave (SARS) associata a coronavirus associato a una libreria di peptidi sintetici sovrapposti. J. Comb. Chem. 7, 648-656.

Jespersen, MC, Peters, B., Nielsen, M. e Marcatili, P. (2017). BepiPred-2.0: migliorare la previsione degli epitopi di cellule B basata su sequenza usando epitopi conformazionali. Acidi Nucleici Res. 45 (W1), W24 – W29.

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B. e Nielsen, M. (2017). NetMHCpan-4.0: Migliorato Peptide-MHC Classe I Interazione

Pronostici che integrano i dati sull'infinità di legame tra ligando eletto e peptidi. J. Immunol. 199, 3360-3368.

Kohyama, S., Ohno, S., Suda, T., Taneichi, M., Yokoyama, S., Mori, M., Kobayashi, A., Hayashi, H., Uchida, T. e Matsui, M. (2009). Efficace induzione di linfociti T citolossici specifici per coronavirus associato alla sindrome respiratoria acuta grave (SARS) associata all'immunizzazione con peptidi liposomiali collegati alla superficie derivati da una polipoteina non strutturale 1a. Res. Antivirale *84*,

168-177.

Krieger, E. e Vriend, G. (2014). YASARA View - grafica molecolare per tutti i dispositivi - dagli smartphone alle workstation. Bionformatics *30*, 2981-2982.

Kringelum, JV, Lundegaard, C., Lund, O. e Nielsen, M. (2012). Previsioni affidabili degli epitopi delle cellule B: impatto dello sviluppo del metodo e miglioramento del benchmarking. PLoS Comput. Biol. 8, e1002829.

Liu, SJ, Leng, CH, Lien, SP, Chi, HY, Huang, CY, Lin, CL, Lian, WC, Chen, CJ, Hsieh, SL e Chong, P. (2006). Caratterizzazioni immunologiche dei candidati al vaccino SARS a base di proteine nucleocapsidiche. Vaccino

24, 3100-3108 .

Middleton, D., Menchaca, L., Rood, H. e Komerofsky, R. (2003). Nuovo database delle frequenze degli alleli: http://www.allelefrequencies.net. Antigeni tissutali *61*,

403-407 .

Paul, S., Weiskopf, D., Angelo, MA, Sidney, J., Peters, B. e Sette, A. (2013). Gli alleli HLA di classe I sono associati a repertori peptidici di dimensioni diverse, af fi nità e immunogenicità. J. Immunol. 191, 5831-5839.

Paul, S., Dillon, MBC, Arlehamn, CSL, Huang, H., Davis, MM, McKinney, DM, Scriba, TJ, Sidney, J., Peters, B. e Sette, A. (2015a). Un approccio di analisi della risposta della popolazione per assegnare le restrizioni HLA-epitopi di classe II.

J. Immunol. 194, 6164-6176 .

Paul, S., LindestamArlehamn, CS, Scriba, TJ, Dillon, MB, Oseroff, C., Hinz, D., McKinney, DM, Carrasco Pro, S., Sidney, J., Peters, B. e Sette, A. (2015b). Sviluppo e validazione di un 382 ampio schema per la previsione di epitopi di cellule T limitati HLA di classe II. J. Immunol. metodi 422, 28-34.

Paul, S., Sidney, J., Sette, A. e Peters, B. (2016). TepiTool: una pipeline per la previsione computazionale dei candidati epitopi delle cellule T. Curr. ProtoC. Immunol. *114*, 18.19.1-18.19.24.

Pickett, BE, Sadat, EL, Zhang, Y., Noronha, JM, Squires, RB, Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., et al. (2012). ViPR: un database bioinformatico aperto e risorsa di analisi per la ricerca virologica. Acidi Nucleici Res. 40, D593-D598.

Sette, A., Moutaftsi, M., Moyron-Quiroz, J., McCausland, MM, Davies, DH, Johnston, RJ, Peters, B., Ra fi i-El-Idrissi Benhnia, M., Hoffmann, J., Su , HP, et al. (2008). Aiuto selettivo per le cellule T CD4 + per le risposte anticorpali a un grande patogeno virale: legame deterministico di speci fi che. Immunità 28, 847-858.

Shichijo, S., Keicho, N., Long, HT, Quy, T., Phi, NC, Ha, LD, Ban, VV, Itoyama, S., Hu, CJ, Komatsu, N., et al. (2004). Valutazione dei peptidi sintetici della sindrome respiratoria acuta grave coronavirus riconosciuta da immunità duratura. Antigeni tissutali *64*, 600-607.

Sidney, J., Peters, B., Frahm, N., Brander, C. e Sette, A. (2008). Supertipi di classe I HLA: una classificazione aggiornata e aggiornata. BMC Immunol. *9*, 1 . Song, Z., Xu, Y., Bao, L., Zhang, L., Yu, P., Qu, Y., Zhu, H., Zhao, W., Han, Y. e Qin, C. (2019). Dalla SARS alla MERS, spingendo i coronavirus sotto i riflettori. I virus *11*, https://doi.org/10.3390/v11010059

Tian, Y., Grifoni, A., Sette, A. e Weiskopf, D. (2019). Risposta TCell umana all'infezione da virus della dengue Davanti. Immunol. 10. 2125. Tsao, YP, Lin, JY, Jan, JT, Leng, CH, Chu, CC, Yang, YC e Chen, SL (2006). Epitopi di cellule T HLA-A * 0201 in coronavirus nucleocapsid e proteine del picco coronavirus della sindrome respiratoria acuta grave (SARS). Biochem. Biophys. Res. Commun. 344, 63-71.

Vita, R., Mahajan, S., Overton, JA, Dhanda, SK, Martini, S., Cantrell, JR, Wheeler, DK, Sette, A. e Peters, B. (2019). Immune Epitope Database (IEDB): aggiornamento 2018. Acidi Nucleici Res. 47 (D1), D339 – D343 .

Wang, J., Wen, J., Li, J., Yin, J., Zhu, Q., Wang, H., Yang, Y., Qin, E., You, B., Li,
W., et al. (2003). Valutazione dei peptidi sintetici immunoreattivi dalle proteine strutturali della sindrome respiratoria acuta grave coronavirus. Clin. Chem. 49, 1989-1996.

Wang, Q., Zhang, L., Kuwahara, K., Li, L., Liu, Z., Li, T., Zhu, H., Liu, J., Xu, Y., Xie, J., et al. (2016). Gli epitopi di coronavirus SARS immunodominanti nell'uomo hanno provocato effetti sia di potenziamento sia di neutralizzazione sull'infezione in primati non umani. ACS Infect. Dis. *2*, 361-376.

Weiskopf, D., Angelo, MA, Azeredo, EL, Sidney, J., Greenbaum, JA, Fernando, AN, Broadwater, A., Kolla, RV, De Silva, AD, de Silva, AM, et al. (2013). Analisi completa delle risposte specifiche del virus dengue porta un ruolo protettivo collegato ad HLA per le cellule T CD8 +. Proc. Natl. Acad. Sci. Stati Uniti d'America 110. E2046-E2053 .

Wrapp, D., Wang, N., Corbett, KS, Goldsmith, JA, Hsieh, CL, Abiona, O., Graham, BS e McLellan, JS (2020). Struttura Cryo-EM del picco 2019nCoV nella conformazione di prefusione. Scienza. https://doi.org/10.1126/ science.abb2507.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., et al. (2020). Composizione del genoma e divergenza del romanzo Coronavirus (2019-nCoV) originaria della Cina. Microbo ospite cellulare.

https://doi.org/10.1016/j.chom.2020.02.001 .

Yang, J., James, E., Roti, M., Huston, L., Gebe, JA e Kwok, WW (2009). Ricerca di epitopi immunodominanti prima dell'epidemia: epitopi proteici con picco SARS-CoV di classe II HLA in soggetti non esposti. Int. Immunol. 21, 63-71.

Zhang, Q., Wang, P., Kim, Y., Haste-Andersen, P., Beaver, J., Bourne, PE, Bui, HH, Buus, S., Frankild, S., Greenbaum, J., et al. (2008). Risorsa di analisi del database di epitopi immunitari (IEDB-AR). Acidi Nucleici Res. *36*,

W513-8.

STAR + METODI

TABELLA DELLE RISORSE CHIAVE

	FONTE	IDENTIFIER
Dati depositati		
Struttura 3D della glicoproteina con punte SARS-CoV-2	Wrapp et al., 2020	ID PDB: 6VSB
Isolato di RNA di Wuhan-Hu-1	Database nuccore NCBI	GenBank: MN908947
Proteina ORF10	Database di proteine NCBI	NCBI: YP_009725255.1
Fosfoproteina nucleocapside	Database di proteine NCBI	NCBI: YP_009724397.2
Proteina ORF8	Database di proteine NCBI	NCBI: YP_009724396.1
Proteina ORF7a	Database di proteine NCBI	NCBI: YP_009724395.1
Proteina ORF6	Database di proteine NCBI	NCBI: YP_009724394.1
glicoproteina di membrana	Database di proteine NCBI	NCBI: YP_009724393.1
proteine dell'involucro	Database di proteine NCBI	NCBI: YP_009724392.1
PRURDER Pigepitopi. REAGENTE o RISORSE	Database di proteine NCBI	NCBI: YP_009724391.1
glicoproteina di superficie	Database di proteine NCBI	NCBI: YP_009724390.1
polipoteina di orf1ab	Database di proteine NCBI	NCBI: YP_009724389.1
Software e algoritmi YASARA		
	Krieger e Vriend, 2014	http://www.yasara.org
IEDB in ViPR. Le sequenze proteiche di isolati di virus naturali con sequenze identic BebiPred 2.0	Vita et al., 2019 the al consenso SARS-CoV e MERS-CoV sono state sele Jespersen et al., 2017	https://www.iedb.org zionate per l'uso nell'analisi delle http://tools.iedb.org/bcell/
Discotope 2.0	Kringelum et al., 2012	http://tools.iedb.org/bcell/
NetMHCpan EL 4.0	Jurtz et al., 2017	http://tools.iedb.org/mhci/
Tepitool	Paul et al., 2016	http://tools.iedb.org/tepitool/

finale. Le sequenze di proteine di consenso di ciascun gruppo di virus sono state determinate dagli allineamenti finali usando lo strumento Analisi delle variazioni di sequenza DISPONIBILITÀ DEI CONTATTI E DEI MATERIALI

Si prega di contattare AS (alex@lji.org) per aliquote di insiemi di peptidi sintetizzati identificati in questo studio. Esistono restrizioni alla disponibilità dei reagenti peptidici a causa di costi e quantità limitate.

DETTAGLI DEL METODO

state allineate usando l'algoritmo MUSCLE in ViPR. Le sequenze che causano scarsi allineamenti in un'analisi preliminare sono state rimosse prima di calcolare l'allineamento Analisi IEDB degli epitopi di Coronavirus T e B.

Gli epitopi delle cellule T e B per i coronavirus sono stati identificati mediante la ricerca dell'IEDB alla fine di gennaio 2020. Le query sono state condotte in senso ampio per i coronavirus (tassonomia IDno. 11118), selezionando saggi positivi nei contesti di cellule T, Bcell e / o ligando. Sono state tabulate le caratteristiche di ciascun epitopo unico (ovvero specie, proteine di provenienza, tipi di test positivi, restrizione MHC), nonché il numero totale di donatori testati e il corrispondente numero totale di donatori con risposte positive nelle cellule B o T saggi e in funzione dell'host. Infine, i punteggi della frequenza di risposta specifica (RF) del saggio T o B sono stati calcolati in modo ampio (cioè, qualsiasi host) o per contesti specifici (ad es. Saggi su cellule T nell'uomo). In particolare, RF = [(r - sqrt (r)] / t, dove r è il numero totale di donatori che rispondono e t è il numero totale di donatori gennaio 2020. Al fine di escludere sequenze di ceppi sperimentali, sono state escluse dall'analisi sequenze di "topi sconosciuti" e host di scimmie. Le sequenze rimanenti sono testati (Carrasco Pro et al., 2015)).

La densità degli epitopi di sequenza SARS-CoV (codice fiscale 694009) è stata visualizzata con lo strumento IEDB Immunobrowser (Dhanda et al., 2018).

Per identificare le regioni dominanti contigue, i punteggi RF per ciascun residuo sono stati ricalcolati per rappresentare una finestra scorrevole di 10 residui.

Confronto tra sequenze di coronavirus e SARS-CoV-2

Tutte le sequenze di proteine a lunghezza intera di SARS-CoV e MERS-CoV sono state recuperate da ViPR (https://www.viprbrc.org/brc/home.spg? decoratore = corona) il 31

Determinazione della conservazione della sequenza SARS-CoV-2

Ogni sequenza di proteine Wuhan-Hu-1 (GeneBank: MN908947) è stata confrontata con le sequenze di proteine di consenso di SARSCoV e MERS-CoV e le sequenze di proteine del parente di pipistrello più vicino (bat-SL-CoVZXC21) usando l'algoritmo BLAST (ViPR;

https://www.viprbrc.org/brc/blast.spg?method=ShowCleanInputPage&decorator=corona) per calcolare l'identità a coppie tra le proteine Wuhan-Hu-1 e il loro obiettivo di confronto.

Previsione dell'epitopo delle cellule B SARS-CoV-2

Sono state eseguite predizioni epitopiche a cellule B lineari su tre diverse proteine coronavirus: glicoproteina di superficie (S), fosfoproteina nucleocapside (N) e glicoproteina di membrana (M) (NCBI: YP_009724390.1, YP_009724393.1, rispettivamente) versioni omologhe di queste proteine sono gli obiettivi primari delle risposte immunitarie delle cellule B per SARS-CoV. Abbiamo usato BebiPred 2.0 (Jespersen et al., 2017) algoritmo incorporato nello strumento di analisi della previsione delle cellule B disponibile in IEDB (Zhang et al., 2008). Per ogni proteina, è stato recuperato il punteggio di probabilità dell'epitopo per ciascun amminoacido e la probabilità di esposizione. I potenziali epitopi delle cellule B sono stati previsti usando un valore soglia di 0,55 (corrispondente a una specifica maggiore di 0,81 e una sensibilità inferiore a 0,3) e considerando sequenze con più di 7 residui di amminoacidi. La previsione anticorpale basata sulla struttura è stata eseguita utilizzando Discotope 2.0 (Kringelum et al., 2012), disponibile in IEDB (Zhang et al., 2008) e un cutoff di positività maggiore di 2.5 è stato applicato (corrispondente a una specifica maggiore o uguale a 0,80 e sensibilità inferiore a 0,39), utilizzando la struttura della glicoproteina con spike SARS-CoV-2 (ID PDB: 6VSB).

Previsione dell'epitopo delle cellule T SARS-CoV-2

La predizione dell'epitopo è stata effettuata usando le dieci proteine previste per l'isolato di riferimento SARS-CoV-2, Wuhan-Hu-1. I corrispondenti numeri di identificazione dell'adesione proteica sono: NCBI: YP_00972555.1 (Orf 10), NCBI: YP_009724397.2 (N), NCBI: YP_009724396.1 (Orf 8), NCBI: YP_009724395.1 (Orf 7a), NCBI: YP_009724395.1 (Orf 7a), NCBI: YP_009724393.1 (Orf 6), NCBI: YP_009724393.1 (M), NCBI: YP_009724392.1 (Envelope protein, E), NCBI: YP_009724391.1 (Orf 3a), NCBI: YP_009724390.1 (S) e NCBI: YP_009724398.1 (Orf 1ab).

Per la previsione degli epitopi delle cellule T CD4, abbiamo applicato un algoritmo precedentemente descritto che è stato sviluppato per predire gli epitopi HLA di classe II dominante, usando il percentile di consenso degli americani del cutoff di predizione% 20 come raccomandato (Paul et al., 2015). Per la previsione degli epitopi delle cellule T CD8, abbiamo selezionato i 12 alleli HLA di classe I più frequenti nella popolazione mondiale (Middleton et al., 2003; Paul et al., 2013), utilizzando un limite di frequenza fenotipica R 6%. Gli alleli specifici inclusi erano: HLA-A * 01: 01, HLA-A * 02: 01, HLA-A * 03: 01, HLA-A * 11: 01, HLA-A * 23: 01, HLA-A * 24:02, HLA-B * 07: 02, HLA-B * 08: 01, HLA-B * 35: 01, HLA-B * 40: 01, HLA-B * 44: 02, HLA-B * 44: 03. Le sequenze di proteine SARS-CoV-2 sono state eseguite su questo insieme di alleli usando l'algoritmo NetMHCpan EL 4.0 e un intervallo di dimensioni di 8-14 metri (Jurtz et al., 2017). Per ogni allele HLA di classe che ho analizzato, abbiamo selezionato i primi 1% di epitopi classificati in base al punteggio di predizione. Per generare un set finale per la sintesi, i peptidi duplicati (cioè quelli selezionati per più alleli) sono stati ridotti a una singola occorrenza e i peptidi nidificati sono stati accorpati in sequenze più lunghe, fino a 14 residui di lunghezza, prima di assegnare le restrizioni HLA corrispondenti multiple per ogni regione.

QUANTIFICAZIONE E ANALISI STATISTICA

Nel presente studio teorico non sono state utilizzate analisi statistiche basate sui dati della letteratura pubblicata e sui database disponibili al pubblico. I calcoli della percentuale di identità e dei punteggi dei fattori di risposta sono stati eseguiti come descritto nei dettagli del metodo sopra.

DISPONIBILITÀ DI DATI E CODICI

Tutti i dati presentati e analizzati nel presente studio sono stati recuperati da IEDB e PDB, come sopra descritto. L'articolo pubblicato include tutti i dati generati o analizzati durante questo studio e riassunti nelle tabelle, nelle figure e nelle tabelle di accompagnamento Materiali supplementari . I file di testo dei dati scaricati dall'IEB sono disponibili su richiesta dell'autore corrispondente.