

Sottigliezze tecniche:

Come dividere i campioni inizialmente? Facile, si usa la regola delle potenze di due, ovvero sia $1/N$ la probabilità che un campione sia positivo, e poniamo per esempio $N=100$. Allora esiste una unica k che è la più grande potenza di due che è più piccola di N (Nell'esempio $k=64$). Il numero di campioni da accorpare insieme nella prima iterazione dell'esperimento è allora $k/2$ (nell'esempio 32 campioni). Questo metodo garantisce due cose: 1) il numero di campioni analizzati per unità di tempo è il massimo possibile 2) in nessun passaggio della nostra analisi noi ci troveremo mai ad avere più provette da analizzare di quante ne possa accomodare il termociclatore* (vedi nota1) (nel senso che solo in un'analisi su 10 capiterà che effettivamente il termociclatore sia pieno e si debba o usare un altro termociclatore oppure finire l'analisi di 2 pazienti su 1000 con 3 ore di ritardo. Se si hanno a disposizione 3 termociclatori il numero di analisi che ritarda di 3 ore anche un solo paziente scende ad $1/50$). **Cosa non fare quando si stima la probabilità a priori p che un campione sia infetto?** La risposta è che non si deve stimarla al ribasso, altrimenti il termociclatore sarà pieno e molte analisi (quelle dei positivi per giunta) saranno ritardate di 3 ore. Quindi se siete indecisi tra $p=1/100$ e $p=1/130$ dovete scegliere $p=1/100$ e dividere i campioni in gruppi da 32.

Come si relaziona la nostra procedura per la PCR con tutto il preprocessing dei campioni? Una sola è la cosa da ricordare, che è quella che ci ha portati fino a qui, ed è che accorpare le analisi vince sul farle singole, quindi ad esempio la fase di estrazione del genomico NON può avvenire paziente per paziente, altrimenti sarebbe questa fase il collo di bottiglia, ma deve avvenire in parallelo ai processi di PCR e con i campioni che serviranno nel prossimo step della PCR.

Quindi i tamponi di tutti i pazienti (ad esempio 1024 pazienti) arrivano nel laboratorio, i tamponi vengono immersi in una soluzione in cui quindi saranno presenti particelle virali, vengono prelevate delle aliquote* (vedi nota2) da ogni campione che si mischiano insieme in gruppi da 32 ad esempio (ovviamente 32 dipende dalla probabilità che un campione sia positivo come spiegato nella sezione apposita). Ognuno di questi gruppi verrà sottoposto all'estrazione del genomico, dando origine a 32 provette ($32 \times 32 = 1024$) in triplice copia=96, il giusto numero per effettuare la prima PCR. Nelle 3 ore in cui la PCR è occupata, si riprendono aliquote da tutti i campioni e li si divide in gruppi da 16 in modo che ogni gruppo da 32 sia esattamente diviso in 2, e da ognuno di questi si estrae il genomico, per un totale di 64 provette in triplice copia=192 provette. Per nostra fortuna il throughput degli estrattori è circa il doppio del throughput della PCR* (vedi nota3), quindi la PCR e l'estrazione dei campioni dovrebbero circa finire insieme. Ora che la PCR è finita alcuni gruppi da 32 sono risultati negativi al test (più del 60%), e tutti gli estratti da 16 tamponi che corrispondevano a provette da 32 che sono risultati negativi possono essere buttati all'istante. Ora è il momento di fare la PCR con i gruppi da 16 che venivano da gruppi da 32 positivi. Questi non occuperanno quasi mai totalmente il termociclatore* (ma se dovesse capitare vedere nota4). Ora quindi si fa in parallelo la PCR ai gruppi da sedici e l'estrazione ai gruppi da 8 che stavano nei gruppi da 32 che sappiamo essere con almeno un positivo. Per lo stesso motivo di prima estrazione e PCR finiscono circa insieme, e si procede in questo modo fino a quando la PCR non sta facendo test su una sola persona, e alla fine di quello si sa esattamente quali pazienti sono positivi e quali no.

Logistica, convenienza e criticità del metodo.

Lati positivi: Non c'è che dire, quando la probabilità p di avere un tampone positivo è bassa il nostro metodo funziona meglio di quello standard di un fattore 4 se $p=1/32$, fino ad arrivare ad un fattore 10 se $p=1/128$ per esempio, ed inoltre dopo solo la prima iterazione il numero di negativi individuati sono la metà delle persone totali analizzate, ovvero per $p=1/32$ si ha un guadagno nel numero di esiti dopo 4h 30m pari

ad un fattore 8, e per $p=1/128$ addirittura un fattore 32.

Criticità: il nostro metodo è conveniente rispetto al metodo attuale (nel senso che aumenta il throughput a scapito di aumentare di almeno 3 ore il tempo di analisi di alcuni pazienti) solo se la probabilità p che un tampone sia infetto è minore del 30% *(vedi nota 5), per cui non è utile per gli screening di pazienti in gravi condizioni o con crisi respiratorie, quando servono analisi veloci e siamo già quasi certi che il paziente sia positivo, ma piuttosto per screening “di massa”, oppure del personale sanitario, dove la percentuale di positivi è ben al di sotto del 30%. Anche se è vero che il nostro metodo avrebbe un throughput molto maggiore del metodo attualmente utilizzato, solo la metà dei pazienti analizzati potrebbero essere avvertiti dopo le normali 4h 30m di attesa, (comunque 600 pazienti rispetto ai 32 del metodo standard nel nostro esempio con $p=1/64$) mentre un quarto dei pazienti verrebbe avvertito dopo ulteriori 3 ore di attesa, un ottavo dopo altre 3 ore di attesa e così via fino ad arrivare ai positivi, gli ultimi ad essere individuati, che nel nostro esempio (1024 pazienti, prob infetto= $1/64$, gruppi iniziali da 32) verrebbero avvertiti solo dopo circa 19h 30m dall’inizio dell’analisi. Altre due criticità tecniche sono che bisogna conservare i tamponi dei singoli pazienti fino a quando non viene svolta l’ultima estrazione (e potrebbe volerci una ventina di ore dall’inizio della procedura), ed infine che la concentrazione di particelle virali potrebbe essere troppo bassa nelle prime fasi del procedimento (quando si formano i gruppi da 32) e potrebbero verificarsi falsi negativi. Questo si può risolvere per esempio con una eventuale concentrazione per centrifugazione da effettuare insieme all’ estrazione del genoma.

Note

1) Qui abbiamo usato la parola termociclatore come sinonimo di macchinario usato per fare la real time RT_PCR per abbreviare.

2) La frazione liquido del tampone che deve essere prelevato non deve eccedere $1/(N+1)$, dove N è l’esponente di due quando all’inizio si processano 2^N tamponi insieme (nel nostro 32 campioni, $N=5$)

3) Abitualmente negli ospedali la fase di inattivazione ed estrazione richiede circa 1h 30min e la PCR 3h.

4) se ad un certo punto la macchina della PCR si riempie e non tutte le provette potessero trovarvi posto basterebbe che nessun paziente venga costretto a due turni in cui il suo genomico non viene processato, questo è il modo di rendere il ritardo il minore possibile per tutti, quindi diciamo ad un certo punto una provetta col genomico di 4 persone rimane fuori, al turno dopo viene processato insieme ai raggruppamenti a due a due lasciando fuori un gruppo da due, al turno dopo se è risultato positivo si processano i 3 gruppi da 2 con tutti i gruppi da 1, e in questo modo si è tardato di al più sei pazienti per un tempo di una corsa di PCR, nessun paziente ha aspettato due corse di PCR in più. Questa eventualità è molto rara comunque.

5) il 30% è la probabilità di positività al di sotto della quale il nostro metodo ha un throughput di pazienti superiore al metodo standard, anche se la metà dei pazienti ci mette 2 corse di PCR a sapere il risultato, quindi in realtà sarebbe meglio usare il nostro metodo per positività sotto il 20% e mai quando il tempismo nella cura dei pazienti è vitale.