La comprensione dei meccanismi molecolari che regolano le interazioni tra macromolecole e il loro significato biologico richiede la caratterizzazione della struttura primaria (sequenza). La sequenza primaria può essere consultata attraverso il sequenziamento diretto, che rappresenta l'unica metodica che garantisce il 99% di affidabilità dei test clinici. Questo metodo consente di rilevare le variazioni nella sequenza nucleotidica e determinare se si tratta di una mutazione o di un polimorfismo.

La disponibilità della sequenza primaria di un gene ci consente anche di comprendere se la proteina codificata (in seguito alla variazione) è ancora funzionale o meno. Attualmente, il sequenziamento è una delle tecniche più ampiamente utilizzate nei laboratori di genetica ed è basato sul "reverse dot plot" (un grafico specifico per le mutazioni). Questa tecnica consente di caricare uno spot di nucleotidi contenente la mutazione, anziché l'intero genoma con enzimi di restrizione, e l'analisi viene eseguita utilizzando un DNA di controllo.Immagine che contiene testo, Carattere, schermata, design

Descrizione generata automaticamente

In un laboratorio di genetica medica, oltre al sequenziatore, si trovano anche altre tecniche, come\_

* la Real-time PCR, che rappresenta un'alternativa alla classica PCR e consente di quantificare il prodotto.
* La tecnica MLPA (Multiplex Ligation-dependent Probe Amplification) consente di identificare grandi delezioni e duplicazioni che spesso sfuggono al sequenziamento di Sanger e alle tecnologie di sequenziamento di nuova generazione (NGS). Nonostante l'NGS, il sequenziamento di Sanger non viene sostituito perché i dati derivanti da un alto livello di processamento richiedono una conferma tramite il sequenziamento di Sanger.

Il metodo di sequenziamento di Sanger è stato sviluppato da Frederick Sanger e colleghi presso l'Università di Cambridge nel 1977, e per questa scoperta Sanger ha ricevuto il premio Nobel. Questa tecnica consente di sequenziare 500.000 paia di basi al giorno e la lunghezza massima del frammento sequenziabile è di 600 basi. Ciò significa che se è necessario sequenziare un gene di grandi dimensioni, è necessario frammentarlo in regioni più piccole.

Il sequenziamento avviene a partire da un singolo frammento amplificato mediante PCR, successivamente la sequenza viene allineata a un riferimento e analizzata per identificare eventuali varianti nucleotidiche. Attraverso un elettroferogramma, è possibile determinare se la sequenza è normale o presenta una mutazione utilizzando un genome browser.

**SINTESI DNA IN VITRO**

Il sequenziamento si basa sul principio e sulla biochimica della replicazione del DNA. Per effettuare il sequenziamento, è necessario avere un template (la sequenza di DNA da sequenziare) e un innesco (primer), che si lega al template per avviare la replicazione.Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Il processo di sequenziamento inizia immergendo il template e l'innesco in una miscela di reagenti che contiene i nucleotidi trifosfato (dNTPs), che sono le unità costituenti del DNA (adenina, timina, citosina e guanina). Questi nucleotidi vengono aggiunti alla sequenza in crescita dall'enzima polimerasi.

Durante il sequenziamento, una Taq polimerasi termostabile viene utilizzata per sintetizzare la nuova elica di DNA. La Taq polimerasi è un enzima che può resistere alle temperature elevate richieste nelle reazioni di amplificazione del DNA.

Man mano che la polimerasi si muove lungo il template riconosce il nucleotide complementare presente nel template e lo lega al filamento in crescita. Questo processo si ripete per ogni nucleotide nella sequenza, consentendo la formazione della nuova elica di DNA.

Durante la sintesi del DNA, vengono utilizzati dNTPs contrassegnati con fluorescenza di diversi colori per identificare i diversi tipi di nucleotidi incorporati nella sequenza. La fluorescenza emessa dai nucleotidi viene rilevata e registrata da un'apparecchiatura di sequenziamento, che produce un tracciato di picchi che corrispondono ai diversi nucleotidi.

Attraverso l'analisi dei picchi generati e l'interpretazione dei dati, è possibile determinare l'ordine preciso dei nucleotidi nella sequenza di DNA originale, permettendo così di ottenere la sequenza completa.

**SEQUENZIAMENTO DI SANGER**

Nella reazione di Sanger, l'allungamento del frammento di DNA avviene grazie ai legami fosfodiesterici in direzione 5'-3', in cui i nucleotidi vengono aggiunti al filamento in crescita tramite l'enzima polimerasi. È importante che i dNTPs utilizzati abbiano un gruppo -OH in posizione 3', che consente la successiva estensione della catena. Tuttavia, nella reazione di Sanger, una parte dei dNTPs viene sostituita con i ddNTPs (dideossinucleotidi), che hanno un gruppo H in posizione 3' invece del gruppo -OH. Questa sostituzione blocca l'estensione ulteriore della catena di DNA.Immagine che contiene testo, schermata, diagramma, numero

Descrizione generata automaticamente

Per questo motivo, il sequenziamento di Sanger è anche noto come metodo di terminazione di catena o metodo dei dideossinucleotidi. Nel contesto del sequenziamento di Sanger, se consideriamo quattro provette separate, i componenti che vengono aggiunti per la replicazione in vitro includono:

Immagine che contiene testo, schermata, diagramma, design

Descrizione generata automaticamente

* Molte copie del DNA template;
* Un primer (a differenza della PCR, ne viene utilizzato solo uno poiché il sequenziamento avviene in avanti o all'indietro);
* Taq polimerasi;
* Abbondanti dNTPs;
* Una piccola quantità di ddNTPs (uno per ogni nucleotide marcato). In passato, venivano utilizzati marcatori radioattivi, mentre oggi si utilizzano fluorocromi come marcatori.

Una volta ottenuta questa miscela, la polimerasi inizia il processo di polimerizzazione e si ferma quando incontra il nucleotide marcato corrispondente. Ciò porta alla formazione di frammenti sequenziati di diverse lunghezze, in modo simile alla generazione dei frammenti di restrizione discusso nelle lezioni precedenti.

Poiché l'obiettivo è analizzare singoli nucleotidi, è necessario utilizzare gel di elettroforesi più sofisticati come il gel di poliacrilamide. I frammenti di DNA vengono caricati sul gel e iniziano la loro "corsa" dal polo negativo al polo positivo, con i frammenti più lunghi che migrano meno rispetto a quelli più corti (a causa del loro peso molecolare superiore). Il risultato della elettroforesi viene letto dall'estremità più pesante a quella più leggera (o viceversa). Ordinando i frammenti dal più leggero al più pesante, si ottiene la sequenza: A, T, G, C, A, C, T, T, G, A, T, C.

**SEQUENZIAMENTO DI SANGER AUTOMATIZZATO**

Il sequenziamento, al giorno d’oggi, è automatizzato. Non vi sono 4 provettine separate, ma una sola dove si inseriscono i ddNTPs marcati con fluorocromi diversi. L’elettroforesi che ne deriva ci mostra come le sequenze siano discriminate da diversi colori e anche i diversi nucleotidi sono discriminati.Immagine che contiene schermata, testo, Rettangolo, design

Descrizione generata automaticamente

Il risultato dell’elettroforesi è consultabile dal computer.

Nello strumento è presente un laser che colpisce la “corsa” elettroforetica e riconosce il fluorocromo ed emetterà un segnale di una determinata lunghezza d’onda che corrisponde alla rappresentazione di un picco specifico nell’elettroferogramma (output).

Nell’immagine a dx è rappresentato il dato grezzo che deriva dalla corsa elettroforetica.

Ogni picco dell’elettroferogramma corrisponde ad un determinato nucleotide che ha un suo colore specifico (Timina: rosso; Adenina: verde; Citosina: blu; Guanina: nero).

L’altezza dei picchi ci dà l’idea della qualità della base che è stata rilevata, ma non ci fornisce dati su una sua eventuale mutazione. I numeri rappresentano le paia di basi. Immagine che contiene testo, diagramma, schermata, Piano

Descrizione generata automaticamente

La mutazione viene segnalata nella linea della sequenza nucleotidica grazie ad un quadratino rosso.

Nel caso in cui ci fosse una mutazione in omozigosi, essa non si leggerebbe sull’elettroferogramma e pertanto dobbiamo confrontare la nostra sequenza con una WT (perché la sequenza non porta errori).

Se ci fosse una mutazione in eterozigosi, all’interno di un solo nucleotide ci sarebbero due picchi sovrapposti.

La sequenza nucleotidica è la sovrapposizione della sequenza del gene dell’esone degli alleli dell’individuo.

**FATTORI CHE INFLUENZANO LA QUALITÀ DEL DATO DI SEQUENZA**

* Qualità dello stampo di DNA;
* Quantità dello stampo di DNA;
* Sequence composition (è più difficile da sequenziare e denaturare se contiene molte GC che formano 3 legami a idrogeno a differenza di AT che ne formano 2);
* Elettroforesi(capillare, polimero);
* Primer design;
* Purificazione della reazione di sequenza (togliere l’eccesso dei reagenti che non hanno più una funzione e potrebbero contaminare il processo).

**QUALITY SCORE (usato soprattutto nell’NGS)**

Il Phred Quality Score (Q score) è utilizzato per valutare l'accuratezza della chiamata delle basi durante il sequenziamento. Il Q score è un punteggio numerico che indica la probabilità che una base sequenziata sia chiamata erroneamente dalla macchina sequenziatrice.Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

Nel sequenziamento Sanger, è possibile valutare immediatamente se ci sono stati problemi durante la lettura delle basi. Nell'immagine a destra, è evidente un groviglio che indica una lettura inaccurata delle basi.Immagine che contiene testo, schermata, software, linea

Descrizione generata automaticamente

Nel sequenziamento di nuova generazione (NGS), la macchina di sequenziamento calcola il Quality Score. Il software Phred è in grado di leggere i cromatogrammi nel formato ABI, che è un formato proprietario molto diffuso della Applied Biosystems. In output, produce file con formati adatti alla visualizzazione, sintesi o elaborazione successiva da parte di altri programmi. Questi file possono essere letti e analizzati utilizzando diversi software. In particolare, Phred può produrre output nel formato FASTA, che viene spesso utilizzato nei sistemi di allineamento.

Phred applica metodi statistici per valutare la qualità di ogni base sequenziata. Analizza la posizione di ogni base e l'area di ogni picco nell'elettroferogramma, confrontandoli con le posizioni teoriche attese. Sulla base di questo confronto, viene calcolato un valore di affidabilità per ogni base. I Quality Score attribuiti da Phred sono associati in modo logaritmico alle probabilità di errore [Q = -10 log10(Pe)].

In conclusione, il Phred Quality Score fornisce un'indicazione della qualità delle basi sequenziate, permettendo di valutare la probabilità di errori nelle chiamate delle basi e garantendo una stima della affidabilità dei dati di sequenziamento.

Immagine che contiene testo, schermata, Carattere, design

Descrizione generata automaticamente

**PROBLEMI POSSIBILI**

N= non sa che nucleotide mettere in quel punto.

Immagine che contiene testo, schermata, design

Descrizione generata automaticamente

Se ci sono delle frameshift, i picchi si sovrappongono.

Slargamenti o eccessiva altezza dei picchi possono essere dovuti a contaminazioni o di troppo DNA.

Immagine che contiene testo, schermata, diagramma, Carattere

Descrizione generata automaticamente

**ESEMPIO DI REFERTO**

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

La professoressa ha sequenziato un campione con il codice ALS762, focalizzandosi sull'esone EX4F.

Per ogni picco nel sequenziamento, è possibile ottenere i dati di affidabilità dal software. I picchi blu indicano basi affidabili, e maggiore è la loro profondità, maggiore è l'affidabilità. I picchi gialli rappresentano situazioni incerte, mentre i picchi rossi indicano anomalie.

L'inizio della sequenza è spesso difficile da interpretare, ed è per questo che si osservano quadrati rossi in quella regione. Questa parte rappresenta il punto di inizio della reazione di sequenziamento. Per evitare problemi di lettura, i primer vengono posizionati in regioni più distanti all'interno dell'esone, di solito nell'introne adiacente. Questo accorgimento permette di superare le difficoltà iniziali e facilita la diagnosi. I "rumori di fondo" visibili nell'immagine possono essere ignorati.

L'avviso rosso (M) può anche apparire quando due picchi sono molto vicini tra loro. Questo evento viene segnalato poiché è stato stabilito un intervallo minimo tra le basi per rilevare eventuali sovrapposizioni. Ad esempio, nella posizione 40, è possibile verificare se si tratta di una sovrapposizione o se i picchi sono semplicemente molto vicini confrontando la sequenza nucleotidica WT (wild type) nei database. Tuttavia, è probabile che si tratti di un errore del server, poiché i picchi hanno la stessa altezza. Se non sono presenti nella sequenza WT, allora indica una mutazione sovrapposta.

La professoressa consulta Ensembl per verificare la sequenza nucleotidica. Per trovare l'esone 4, si cerca il cDNA (complemento di DNA) dove è possibile visualizzare anche le varianti (come spiegato nelle lezioni precedenti). È importante prestare attenzione al genoma di riferimento utilizzato. Dalla ricerca emerge che la sequenza è normale e che il programma non ha correttamente distanziato i picchi.

La prima cosa da fare con la sequenza è individuare l'inizio dell'esone (in questo caso, CAT GTT alla POSIZIONE 87) e inserire uno slash (/) ogni 3 nucleotidi per suddividere la sequenza in codoni.

La mutazione identificata è un cambiamento di ASP91A. Il tempo richiesto per la relazione finale può aumentare se la mutazione è nuova e non viene trovata nei database.