

SAGE: Serial Analysis of Gene Expression

L'insieme di tutti gli mRNA presenti in una cellula si definisce *trascrittoma*. Ogni trascrittoma ha una composizione complessa, con migliaia di mRNA diversi, ciascuno presente ad una data concentrazione. Idealmente, per caratterizzare un trascrittoma è necessario identificare tutti questi mRNA. Il modo più diretto di caratterizzare un trascrittoma sarebbe quello di convertire tutti gli mRNA in cDNA e poi di determinare la sequenza di tutti i cloni della libreria di cDNA così ottenuta. Questo approccio, seppur realizzabile, è estremamente dispendioso sia in termini di tempo che di risorse, perché si analizza un gene alla volta. Diverse tecniche consentono invece l'analisi in serie dei geni: la *visualizzazione differenziale degli mRNA* è utile per capire come si modifica l'espressione genica nelle cellule in diverse fasi fisiologiche e di sviluppo (analisi comparativa); questo metodo però identifica solo un gruppo parziale di geni e non è utilizzabile per determinare l'abbondanza dei diversi trascritti. *L'analisi seriale dell'espressione genica (SAGE)* permette di determinare contemporaneamente la struttura, la funzione e l'espressione di un intero trascrittoma. Questa tecnica è stata ideata e messa a punto nel 1995 da Velculescu *et al.* (Science 270: 484), affrontando lo studio di mRNA espressi nel pancreas.

Il SAGE si basa su due principi:

- Una corta etichetta di sequenza (9-10 pb) contiene sufficienti informazioni per identificare un qualunque mRNA in modo univoco;
- La concatenazione di molte etichette (*tag*) in un unico clone di cDNA consente l'analisi in serie degli mRNA in modo più efficiente.

Opportuni programmi al computer saranno successivamente in grado di convertire le informazioni inizialmente presenti come un'unica stringa di dati in una tabella che riporta per ogni mRNA identificato descrizione e livello di espressione genica (numero di tag).

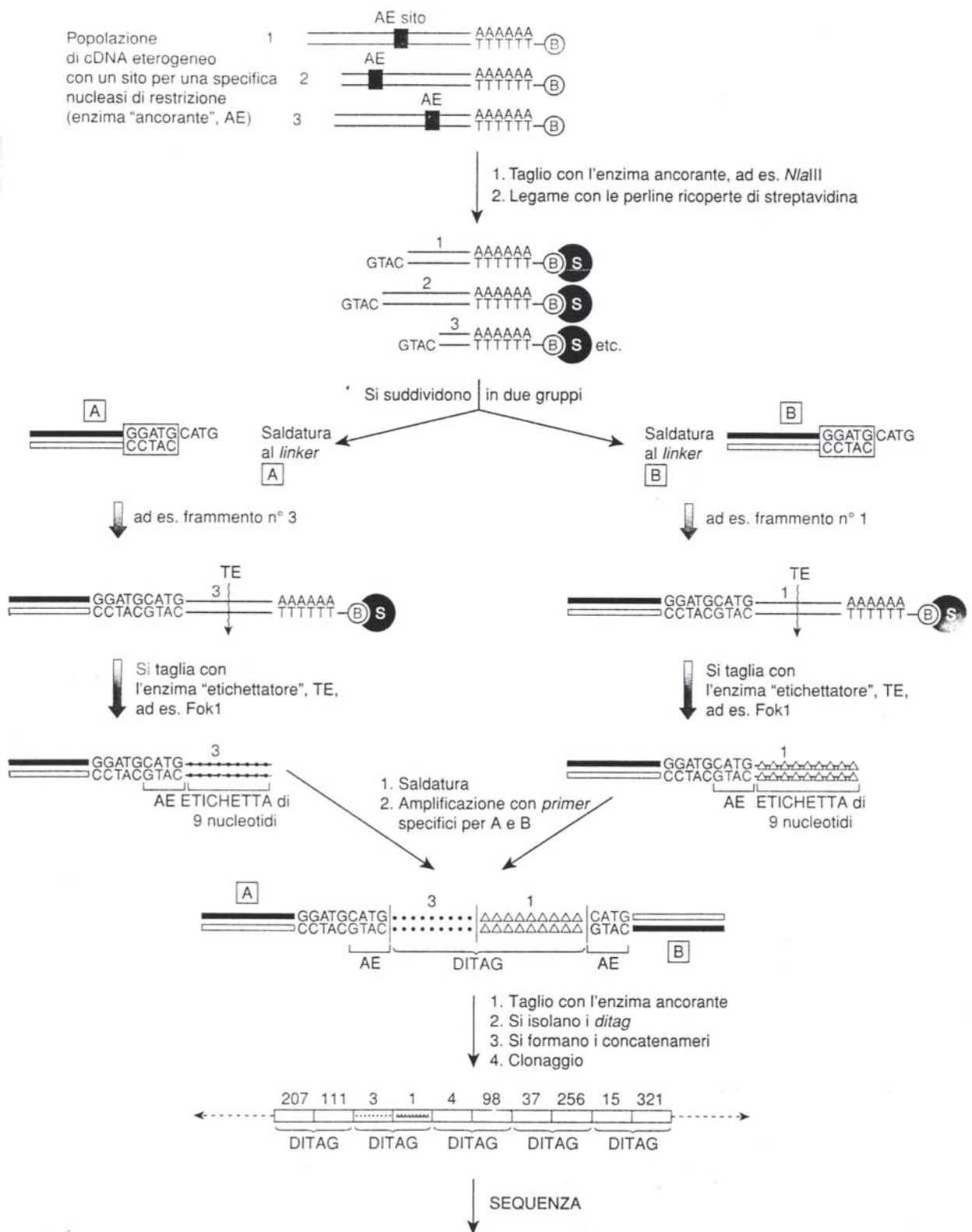
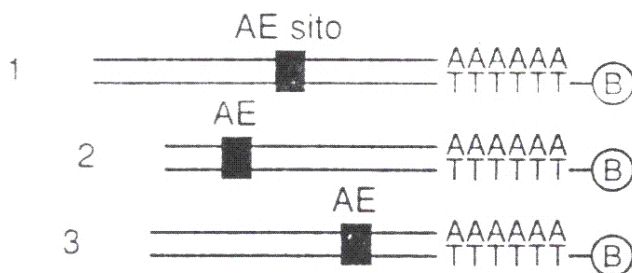


Figura 1 Schema illustrativo del SAGE. L'enzima ancorante è *Nla III* e l'enzima etichettatore è *Fok I*.

Per il SAGE si utilizzano due tipi diversi di enzimi di restrizione:

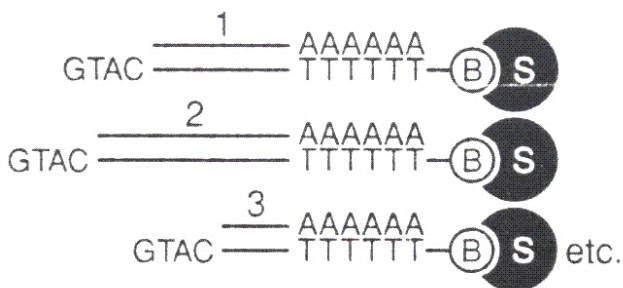
- **Enzima ancorante (AE)**: riconosce una sequenza di 4pb e taglia in media ogni 256 pb (es. Nla III, Alu I, etc.).
- **Enzima "etichettatore" (TE)**: taglia ad un numero definito (fino a 20) di basi a valle del sito di riconoscimento. Questi enzimi (es. Fok I, Bsm FI) appartengono alla sottoclasse IIS delle endonucleasi di restrizione.

1) Sintesi del cDNA



Per la sintesi del cDNA si usa un primer oligo-dT biotinilato (B).

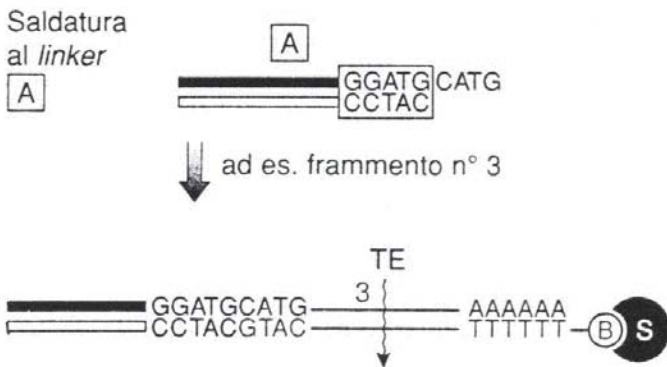
2) Taglio con l'enzima ancorante (AE)



I cDNA in doppia elica sono tagliati con l'enzima AE (in questo caso Nla III, riconosce le sequenze GTAC), creando una popolazione di frammenti lunghi in media 256 pb. Per isolare solo i frammenti 3' terminali si utilizzano delle sferette rivestite di streptavidina (S), una proteina che ha un'altissima affinità per la biotina. Notare che in questo caso non si possono studiare messaggeri privi della sequenza GTAC.

3) Saldatura ai linker

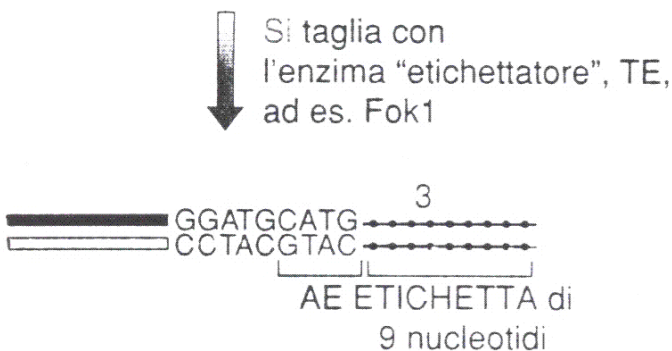
I frammenti di cDNA 3' terminali sono suddivisi in due frazioni, ciascuna delle quali è saldata ad un adattatore (*linker*) diverso.



Ogni adattatore è disegnato in modo tale da presentare:

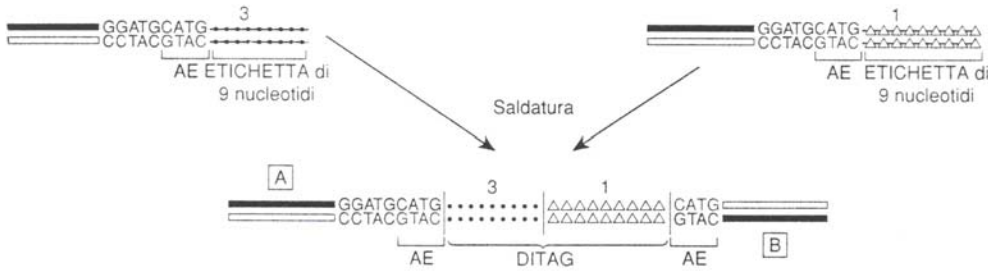
- un'estremità sporgente CATG compatibile con quella dei frammenti
- un sito di riconoscimento per l'enzima etichettatore
- sequenze aggiuntive per i primers specifici utilizzati per la PCR.

4) Taglio con l'enzima etichettatore



In seguito al taglio con l'enzima TE il frammento più lungo di cDNA 3' terminale rimane attaccato alle sferette, mentre l'adattatore con una breve etichetta si può recuperare in soluzione. Notare che dopo il taglio con l'enzima TE rimangono 4 basi sporgenti in 5', per cui prima del passaggio successivo è importante creare estremità piatte (reazione di *fill-in* con la T4 DNA polimerasi).

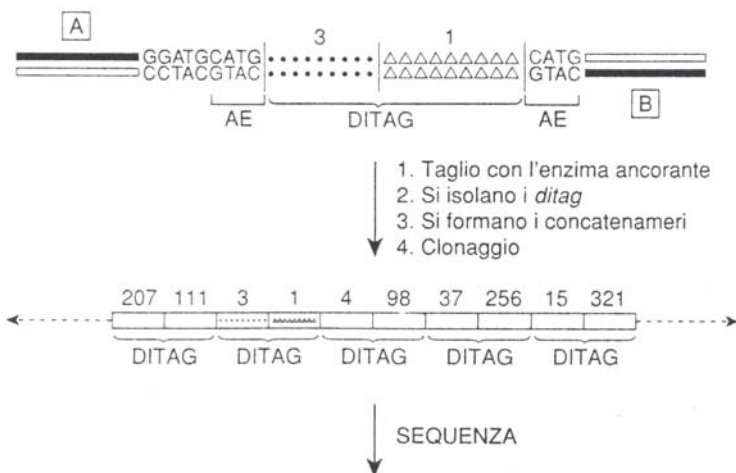
5) Formazione di etichette doppie



Anche se saldando insieme le molecole delle due frazioni linkerA-etichetta e linkerB-etichetta si ottengono diversi prodotti, per le fasi successive si considerano solo i **di-tag**, cioè un linkerA-etichetta legato a un linkerB-etichetta. I di-tag sono amplificati mediante PCR utilizzando primer specifici per gli adattatori **A** e **B**. In questo modo si ottengono i seguenti vantaggi:

- amplificazione del materiale per il clonaggio
- orientazione e puntuazione delle etichette in modo regolare
- l'analisi dei di-tag formati prima dell'amplificazione permette di correggerne eventuali distorsioni introdotte in modo selettivo durante la PCR

6 Concatenazione e clonazione dei di-tag



Dopo il taglio con l'enzima ancorante iniziale, i di-tag sono separati dai linker mediante estrazione da gel. Dopo ligazione dei di-tag fra di loro, si selezionano i concatenameri più lunghi di 200 pb, che sono quindi clonati in vettori plasmidici. Ogni sequenza prodotta dai cloni ottenuti produce da 10 a 50 etichette diverse.

L'articolo originale di Velculescu *et al.* riporta l'analisi di 1000 etichette prodotte da cDNA di pancreas:

- 30 tag ricorrono più di 4 volte (mRNA abbondanti, 45.2% dei tag totali)
- 15 tag ricorrono 3 volte
- 32 tag ricorrono 2 volte
- 351 tag ricorrono solo 1 volta (mRNA rari, 41.8% dei tag totali)

Notare che i tag dei messaggeri rari sono anche quelli meno affidabili, perché un singolo errore di assegnamento di base durante la lettura della sequenza è sufficiente ad invalidare i dati.

Per isolare geni sconosciuti identificati tramite SAGE si può vagliare una libreria di cDNA con oligonucleotidi contenenti una sequenza di 13 nt (9 nt tag + 4 nt sito riconosciuto da AE). In generale, si osserva una buona correlazione dei dati prodotti dal SAGE con quelli ottenuti dall'analisi di librerie di cDNA.

Per determinare l'espressione di un intero trascrittoma il numero di tag prodotti deve essere molto grande, intorno a 50.000 nel caso di cellule di mammifero.

ARTICOLO ORIGINALE

Velculescu VE, Zhang L, Vogelstein B and Kinzler W (1995) Serial Analysis of Gene Expression. *Science* Vol 270 (20 october): 484-487.

REVIEW

Velculescu VE, Vogelstein B and Kinzler W (2000) Analysing uncharted transcriptomes with SAGE. *Trends. Genet.* 16: 423-425.