

DIGIT~Bio~TECH



LO1 BIOLOGIA DEI SISTEMI E TECNOLOGIA OMICA

Livello basale

AUTORE:

ANNA TOMOVA



Sommario

Biologia dei sistemi e tecnologie Omics: il quadro generale	3
Biologia dei sistemi della cellula: dai dati degli esperimenti a omica singola a quelli multi-omici	3
Diversi rami dell'omica: sfide per combinare le informazioni biologiche	6
1. Genomica	8
2. Epigenomica.....	11
3. Trascritomica.....	13
4. Proteomica.....	16
5. Metabolomica.....	19
Vantaggi e svantaggi delle tecnologie omiche	23
Referenze:	26



Biologia dei sistemi e tecnologie Omics: il quadro generale

Negli ultimi decenni il rapido progresso della biologia molecolare e i risultati in questo settore scientifico sono direttamente correlati ai progressi delle scienze informatiche e allo sviluppo di nuovi strumenti software. L'applicazione di diverse tecnologie legate all'omica (genomica, epigenomica, trascrittomica, proteomica, metabolomica, ecc.) porta all'accumulo di dati biologici versatili e la loro interpretazione combinata apre nuove possibilità per gli scienziati di passare dallo studio di molecole biologiche isolate a un'ampia analisi di grandi serie di molecole biologiche. Le scienze biologiche sono diventate le scienze dei big-data che sono sempre più importanti per risolvere i problemi riguardanti la salute umana e l'ambiente. Le sfide dei big-data non sono solo le loro dimensioni, ma anche la loro crescente complessità.

BIOLOGIA DEI SISTEMI DELLA CELLULA: DAI DATI DEGLI ESPERIMENTI A OMICA SINGOLA A QUELLI MULTI-OMICI

La biologia dei sistemi è emersa come un nuovo campo di studio interdisciplinare che ha attirato un'enorme attenzione negli ultimi anni. Sebbene il termine "Biologia dei sistemi" sia stato usato in vari modi, è generalmente inteso per descrivere la ricerca che combina la biologia con discipline apparentemente disparate come fisica, biochimica, ingegneria, biostatistica, matematica, informatica, bioinformatica e altre. Lo scopo principale di questo campo interdisciplinare è quello di ottenere una visione altamente parallela dei sistemi biologici a livello molecolare e capire come funzionano nel loro insieme invece come somma di parti. Chiarendo i meccanismi e i processi molecolari a diversi livelli di sistema nell'organismo o in una cellula, si potrebbe prevedere come questi sistemi cambino nel tempo e in condizioni variabili. La grande quantità di informazioni ottenute sulle basi molecolari della fisiologia e dell'organizzazione cellulare potrebbe fornire soluzioni riguardanti malattie, tossicità, terapie, scoperta di farmaci, ecc.

La biologia dei sistemi è stata lanciata come disciplina distinta negli anni '50 con il lavoro dei teorici dei sistemi Mihajlo Mesarovic e Ludwig von Bertalanffy sulla teoria generale dei sistemi. Il concetto principale di questa teoria è che la dinamica di ogni sistema è il risultato delle relazioni tra le sue unità separate che ne determinano la funzione. A quel tempo, studiando a fondo enzimi e cinetica delle reazioni enzimatiche, biochimici, seguaci della teoria dei sistemi, cercarono di esaminare il comportamento delle vie biochimiche come una rete invece che come somma delle sue parti. Allontanandosi dall'area biochimica, Reinhart Heinrich sviluppò approcci teorici per la descrizione e



2019-1-BG01-KA203-062371

l'indagine quantitativa delle vie di segnalazione e sviluppò una teoria del controllo metabolico. È diventata evidente la necessità di un approccio integrato per uno studio e una comprensione più dettagliati dei complessi processi biologici.

La biologia dei sistemi dipende fortemente dalle informazioni biologiche ottenute dalla biologia molecolare e / o dai singoli approcci omici, ma spesso sono in gran parte basati su ipotesi o riduzionisti. Nell'approccio riduzionista vengono studiate le proprietà funzionali dei singoli componenti molecolari nei sistemi biologici complessi. Dopo il completamento del Progetto Genoma Umano, la scienza moderna si è sviluppata al di là della visione incentrata sul gene della precedente era genomica. Questo periodo designato come era postgenomica è caratterizzato da importanti cambiamenti nella conduzione della ricerca scientifica e nell'interpretazione dei risultati e secondo la Bloom questa è la fine del "riduzionismo ingenuo". Tuttavia, questi metodi classici da soli non possono fornire una comprensione completa degli organismi viventi, ma continueranno ad essere un elemento essenziale di tutte le ricerche biologiche. In questo contesto, la biologia dei sistemi ha causato il cambiamento fondamentale negli approcci tradizionali. Utilizza un nuovo approccio olistico che genera ipotesi, si concentra sullo studio dei sottosistemi, che consente una comprensione più profonda dell'intero processo. La via di utilizzo del galattosio (GAL) del lievito *Saccharomyces cerevisiae* è un esempio dell'applicazione dell'approccio di biologia dei sistemi per la rilettura e l'interpretazione dei risultati ottenuti dall'approccio riduzionista (un gene/una proteina alla volta). Sulla base dei dati sperimentali ottenuti dalle analisi dei livelli di proteine e RNA, nonché dalle interazioni proteina-proteina e proteina-DNA e dalla loro integrazione in un unico modello, è stata fornita una nuova ipotesi sulla regolazione della via GAL, che è stata sperimentalmente verificata in seguito.

Attraverso l'uso di strumenti computazionali e matematici una grande quantità di dati sperimentali viene raccolta e integrata dalla biologia dei sistemi per rivelare modelli sconosciuti e generazione di ipotesi. Questa è una strategia importante per ottenere nuove intuizioni sui sistemi biologici e anche aiutare nella progettazione sperimentale.

L'approccio sistemico dei sistemi biologici è volto a chiarire le seguenti tre questioni nel contesto della rete molecolare: i) quali sono le singole componenti del sistema; ii) come funzionano separatamente? e iii) come lavorano insieme questi componenti per svolgere un compito? (3). Nel contesto delle reti molecolari, lo scopo fondamentale della biologia dei sistemi può essere riassunto come segue: (i) una comprensione della struttura di tutti i componenti di una cellula / organismo fino a livello molecolare, (ii) la capacità di prevedere lo stato futuro della cellula / organismo in un ambiente normale, iii) la capacità di prevedere le risposte di output per un dato stimolo di input, e iv) la capacità di stimare i cambiamenti nel comportamento del sistema al momento della perturbazione dei componenti o dell'ambiente.

Il termine "sistemi" in biologia dei Sistemi definisce diversi intervalli di complessità che vanno da due macromolecole che interagiscono per svolgere un particolare compito a organismi interi (Fig.1).



2019-1-BG01-KA203-062371

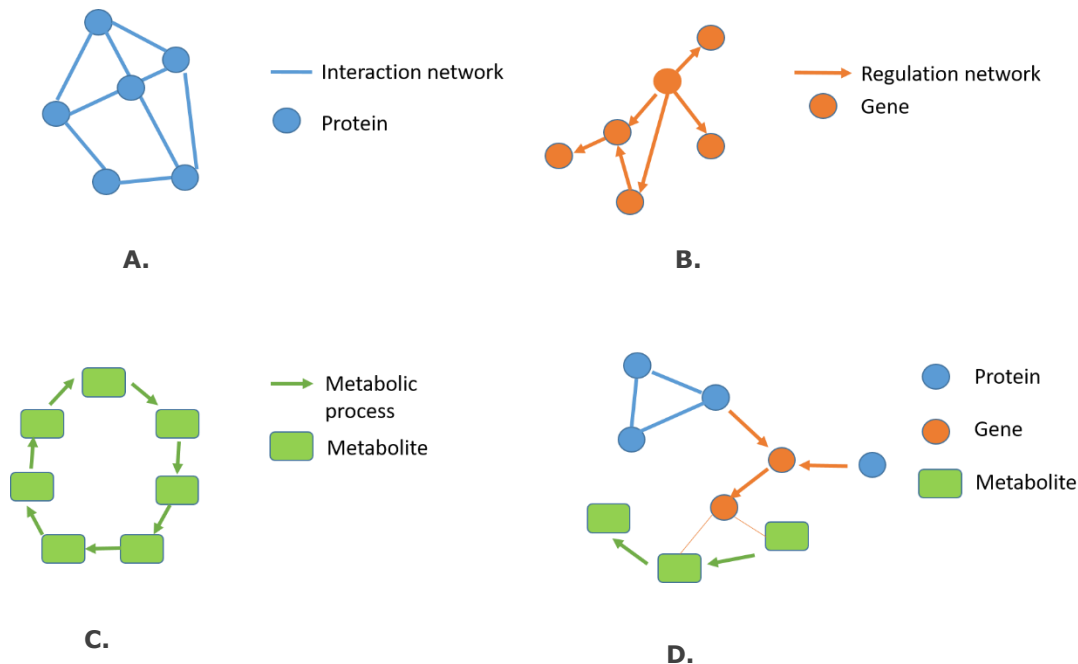


Fig. 1. Principali tipi di reti biologiche. A. Interazioni proteina-proteina; B. Rete di regolazione genica; C. Reti metaboliche; D. Diversi tipi di rete nella cellula (9)

Pertanto, la biologia dei sistemi condivide un obiettivo scientifico comune con la disciplina della fisiologia che è dedicata allo studio della funzione integrata dell'intero complesso sistema biologico. Gli organismi sono molto più della somma delle loro parti e la complessità dei processi fisiologici non può essere compresa semplicemente studiando come funzionano i componenti in isolamento. Ad esempio, i geni codificano la struttura primaria delle proteine cellulari che a loro volta svolgono funzioni specifiche a supporto del metabolismo cellulare e della fisiologia e dello sviluppo dell'organismo. Allo stesso tempo, le proteine in ogni cellula non agiscono in isolamento ma in una rete complessa che è importante per specificare i fenotipi. La maggior parte dei processi biologici sono altamente dinamici e spesso coinvolgono più di un tipo di molecole. Allo stesso tempo un fenotipo può essere condizionato da diversi meccanismi molecolari ed epigenetici e un tipo di molecola può essere coinvolto in diversi fenotipi. In realtà, anche nello stesso organismo, una proteina può svolgere ruoli diversi in cellule diverse; o effetti del percorso del segnale possono indurre vari programmi di differenziamento in diverse linee cellulari. Inoltre, negli organismi multicellulari, le singole cellule non hanno un'esistenza indipendente dall'intero organismo, sono legate ontogeneticamente.

Indipendentemente dal loro obiettivo comune, la biologia e la fisiologia dei sistemi utilizzano diversi strumenti e approcci sperimentali, il che porta ad ottenere un diverso insieme di dati sperimentali.



2019-1-BG01-KA203-062371

La biologia dei Sistemi interdisciplinari è una disciplina moderna in rapida evoluzione a causa del fatto che utilizza una varietà di metodi e strumenti tra cui la genomica funzionale su larga scala e altre tecnologie omiche, la bioinformatica e la modellazione al computer, che non sono sfruttate dai fisiologi. Il ruolo della biologia computazionale nella biologia dei Sistemi è quello di elaborare e analizzare enormi quantità di dati empirici prodotti da diversi livelli di omica che a loro volta portano alla scoperta di conoscenze biologiche e alla generazione di ipotesi di ricerca. L'applicazione di analisi *in silico* e basate sulla simulazione offre l'opportunità di fare previsioni che sono ulteriormente confermate da saggi sperimentali. Il rapido progresso delle tecnologie dell'informazione, compreso il miglioramento delle basi di dati pubbliche web dell'informazione biologica, sostiene anche lo sviluppo di studi omici e portano rispettivamente al progresso della biologia dei sistemi.

Al fine di studiare in modo completo i processi biologici, è fondamentale capire come gli strati biologici separati (genoma, epigenoma, trascritoma, proteoma, metaboloma e ionoma) si interconnettano tra loro in un sistema cellulare e come avviene il flusso di informazioni biologiche. La combinazione di diverse analisi omiche che utilizzano un approccio multi-omico è necessaria per progettare un quadro preciso degli organismi viventi. I dati di trascritomica, proteomica e metabolomica possono rispondere alle principali domande biologiche riguardanti l'espressione di trascritti, proteine e metaboliti, indipendentemente, ma un'integrazione sistematica multi-omica può assimilare, annotare e modellare in modo completo questi grandi set di dati.

DIVERSI RAMI DELL'OMICA: SFIDE PER COMBINARE LE INFORMAZIONI BIOLOGICHE

Il trasferimento dell'informazione genetica nei sistemi biologici si realizza dal DNA all'mRNA alla proteina e questo è considerato come il dogma Centrale della biologia molecolare. I tre processi principali in ogni cellula sono la replicazione, la trascrizione e la traduzione. Il loro flusso costante garantisce il mantenimento e la conversione delle informazioni genetiche, codificate nel DNA in prodotti genetici, che sono RNA o proteine, a seconda del gene. La replicazione è un processo di duplicazione del DNA di una cellula ed è la base per l'ereditarietà biologica. Viene effettuata dall'enzima DNA polimerasi che copia una singola molecola di DNA a doppio filamento parentale in due molecole di DNA a doppio filamento figlie. L'enzima RNA polimerasi crea una molecola di RNA dal DNA e tale processo è noto come trascrizione. La molecola di RNA appena sintetizzata è complementare a un tratto gene-codificante del DNA. La traduzione produce proteine dall'mRNA. Il ribosoma genera una catena polipeptidica di amminoacidi usando l'mRNA come modello. La catena polipeptidica si piega fino a diventare una proteina. Nelle cellule eucariotiche, o in quelle cellule che hanno un nucleo, la replicazione e la trascrizione avviene all'interno del nucleo mentre la traduzione avviene al di fuori del nucleo nel citoplasma. Nelle cellule procariotiche, o in quelle cellule che non hanno un nucleo, tutti e tre i processi si verificano nel citoplasma. Il fenotipo degli organismi è determinato da quel paradigma di trasferimento



2019-1-BG01-KA203-062371

delle informazioni. I biologi hanno studiato queste "omi" per anni sotto forma di genomica, trascrittomica e proteomica. I dati di questi approcci sperimentali sono integrati da epigenomica e metabolomica che sono stati recentemente utilizzati per risolvere problemi specifici riguardanti molte funzioni di un organismo. Il rapido sviluppo e il progresso delle tecnologie "omiche" determinano la progressiva espansione del volume di informazioni che possono essere raccolte nei singoli studi. Inoltre, l'attuale elevata produttività di queste tecniche ha aumentato l'accessibilità a queste informazioni in termini di tempo e costi. Molti ricercatori si trovano in una situazione in cui possono raccogliere diversi set di dati omici sugli stessi campioni sperimentali. Al fine di ottenere conclusioni più complete sui processi biologici, questi insiemi di dati devono essere integrati con un approccio multi-omico e analizzati come sistema olistico (Fig. 2).

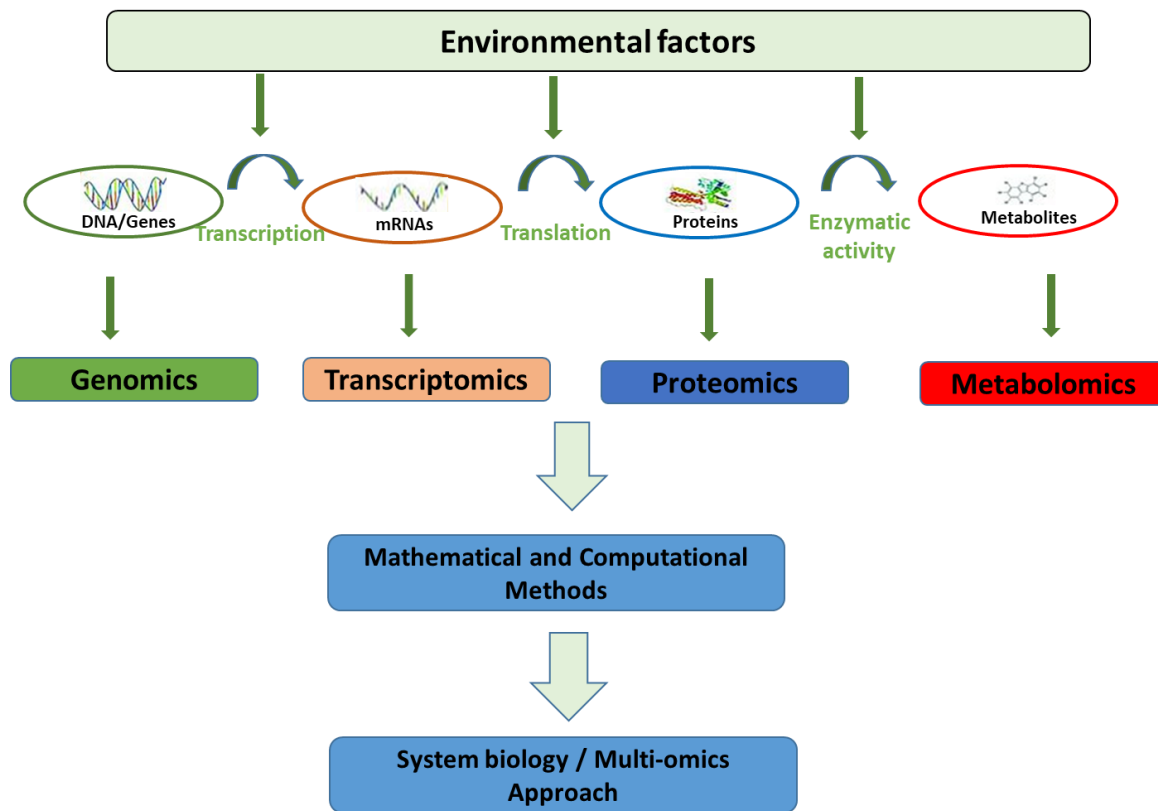


Fig. 2. Approcci omici integrati della biologia dei Sistemi

Il termine "Omica" derivato da una parola Greca e l'aggiunta del suffisso -oma alle molecole cellulari, come gene, trascritto, proteina, metabolita, dà significato di "intero", "tutto" o "completo". I



2019-1-BG01-KA203-062371

diversi strati della cellula, costituiti da DNA e modifiche (Genoma, Epigenoma), RNA e contenuto proteico (Transcriptome, Proteoma), piccole molecole (Metaboloma, Lipidoma) e composizione elementare (misurata come 'Ionome'), possono essere analizzati dalle tecnologie omiche. La combinazione di strati omici in un set di dati multi-oma è integrata attraverso una solida biologia dei sistemi che è in grado di rivelare meccanismi e interazioni inter-strato, nonché la funzione dei tessuti delle popolazioni cellulari, degli organi e dell'intero organismo. Gli approcci Omici comprendono un numero maggiore di misurazioni per endpoint e sebbene il numero di parametri misurati per analisi sia aumentato, il numero di repliche viene ridotto. Ciò è dovuto, da un lato, alla super valutazione dei metodi, poiché si ritiene che un maggior numero di misurazioni compenserebbe un piccolo numero di campioni e, dall'altro lato, è dovuto al costo e al tempo degli esperimenti omici.

Le discipline mono-omiche miravano a studiare specifiche questioni biologiche senza richiedere una comprensione preliminare delle basi biologiche coinvolte. A seconda del tipo di biomolecola su cui si concentrano principalmente in uno specifico campione biologico, le tecnologie omiche sono suddivise in: genomica (genoma/gene), metagenomica (genomi recuperati direttamente da campioni ambientali), epigenomica (struttura di supporto del genoma, inclusi leganti proteici e RNA, strutture alternative del DNA e modifiche chimiche sul DNA), trascrittomica (mRNA), proteomica (peptidi / proteine), metabolomica (metaboliti), lipidomica (lipidi) glicomica (carboidrati e zuccheri), ionomica (ioni).

Sebbene nessuna delle attuali tecnologie omiche sia perfetta, alcune di esse riescono a fornire un quadro più completo dello strato biologico che mirano a studiare rispetto ad altre. Questo fatto è dovuto non solo alle differenze nello stato degli sviluppi tecnologici, ma piuttosto alle differenze nella complessità chimica e fisica di ogni livello biologico.

1. Genomica

La genomica è lo studio sistematico di tutti i geni di un organismo (il genoma), comprese le interazioni di tali geni tra loro e con l'ambiente dell'organismo. Il genoma è lo strato biologico basale nella cellula e rappresenta il DNA totale di una cellula o di un organismo. L'acido deossiribonucleico (DNA) è il composto chimico che contiene tutte le informazioni genetiche necessarie per sviluppare e dirigere le attività di quasi tutti gli organismi viventi. Le molecole di DNA sono fatte di due filamenti intrecciati, spesso indicati come una doppia elica. Ogni filamento di DNA è costituito da quattro basi nucleotidiche - adenina (A), timina (T), guanina (G) e citosina (C) che si accoppiano specificamente su filamenti opposti: A si accoppia sempre con una T; una C si abbina sempre a una G. Una sequenza di tre nucleotidi adiacenti (codone) codifica per uno specifico amminoacido durante la sintesi o la traduzione delle proteine. L'ordine dei nucleotidi lungo le molecole di DNA determina il codice genetico. Il codice genetico è universale perché è lo stesso tra tutti gli organismi ed è anche degenerato perché 64 codoni codificano solo 22 amminoacidi. Con il suo linguaggio di quattro lettere, il DNA contiene le informazioni necessarie per costruire un intero organismo. Un gene rappresenta l'unità di DNA che porta



2019-1-BG01-KA203-062371

codici per la creazione di una proteina specifica o di un insieme di proteine. Sulla base della complementarità intrinseca delle basi nucleotidiche nel DNA, è diventato possibile sequenziare rapidamente un numero enorme di genomi a un costo relativamente basso. Dai genomi efficientemente sequenziati si possono fare previsioni sulle sequenze di RNA e proteine, che sono combinate negli approcci multi-omici. La forma digitale delle sequenze di DNA come una sorta di informazione biologica omica potrebbe essere facilmente memorizzata in database biologici e condivisa tra scienziati di tutto il mondo. Il sequenziamento del primo intero genoma del batterio *Haemophilus influenza* nel 1995 ha fatto una rivoluzione nella biologia molecolare. È stato prodotto il grande volume di dati delle sequenze che non era possibile interpretare completamente. Decifrare le informazioni genetiche rilevanti dal materiale genetico di base era quasi un compito impossibile. Per superare questo problema, erano necessarie informazioni biologiche più dettagliate. Erano necessarie informazioni sulla trascrizione del materiale genetico e sulla successiva produzione di proteine.

Una direzione di ricerca essenziale nel campo della biologia dei sistemi è la genomica funzionale. Questa disciplina sviluppa e sfrutta metodologie su larga scala e ad alta produttività con l'obiettivo di definire e analizzare la funzione genica a livello globale. È molto importante sviluppare una comprensione a livello di sistema di un processo biologico per identificare i geni e le proteine che codificano che lavorano insieme per dare origine a quel processo. La genomica funzionale è un campo scientifico integrativo che combina più set di dati su larga scala nel tentativo di generare intuizioni sulla funzione genica. All'inizio, i geni sono stati analizzati individualmente, ma il progresso delle tecnologie negli ultimi anni rende possibile l'analisi contemporanea dell'espressione di migliaia di geni. Questa analisi su larga scala della funzione genica è chiamata tecnologia dei microarray di DNA (Fig. 3).



2019-1-BG01-KA203-062371

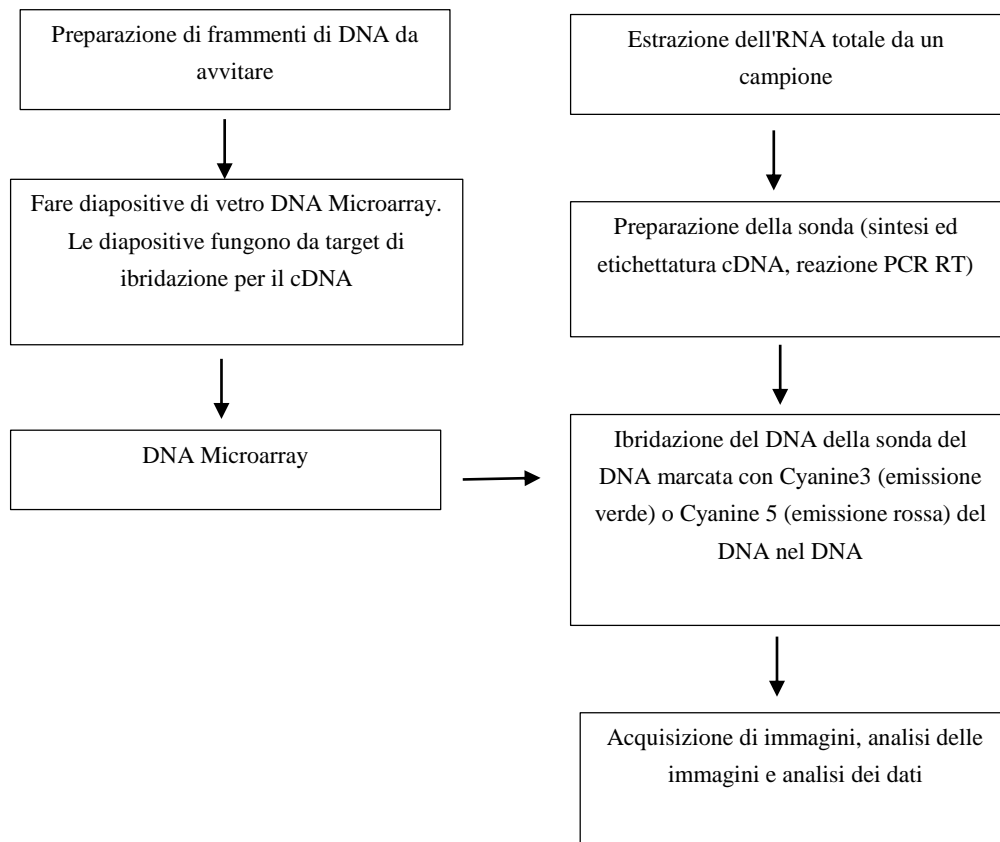


Fig. 3. Schema di analisi del Microarray del DNA

I microarray di DNA misurano le differenze nella sequenza del DNA tra gli individui e questa tecnologia può fornire informazioni per la funzione di geni non caratterizzati e può anche rivelare gruppi di geni interagenti che danno origine a un processo biologico di interesse. Le analisi microarray possono anche fornire approfondimenti sui meccanismi di regolazione genica, evoluzione ed eziologia della malattia. Ad esempio, l'analisi dei dati dei microarray potrebbe rivelare anomalie come inserzioni e delezioni cromosomiche o numeri cromosomici anomali in un processo chiamato ibridazione genomica comparativa. Le variazioni più comuni nelle sequenze di DNA tra le persone sono i polimorfismi a singolo nucleotide (SNP), in cui un nucleotide viene sostituito da un altro; questo può avere un significato funzionale se il cambiamento si traduce in un codone per un amminoacido diverso. Sono di particolare interesse se legati a malattie con una determinazione genetica. Il profilo del polimorfismo mono nucleotidico ha anche un ruolo nella farmacogenomica nell'esplorare le risposte dei singoli pazienti ai farmaci.



2019-1-BG01-KA203-062371

2. Epigenomica

Le modifiche epigenetiche come la metilazione del DNA, le modifiche istoniche, l'analisi 2D e 3D della struttura della cromatina e l'RNA non codificante sono studiate con i metodi dell'epigenomica. Questa direzione -omica si concentra sull'analisi dei cambiamenti epigenetici complessivi che fornisce informazioni importanti sui meccanismi e la funzione della regolazione genica in molti geni in una cellula o organismo. Gli scienziati hanno capito che il fenotipo dell'individuo non è controllato solo dal genoma, ma anche dai cambiamenti nella regolazione delle attività geniche. Esperimenti genetici nell'uomo e negli animali hanno dimostrato che, oltre alla sequenza del DNA, i segni epigenetici possono essere trasmessi da genitore a prole attraverso i gameti e influenzare il fenotipo della prole.

L'epigenomica definisce le modifiche nella regolazione delle attività geniche che agiscono senza, o indipendentemente da, cambiamenti nelle sequenze geniche. Alcune definizioni limitano l'epigenetica/epigenomica alle modifiche del fenotipo senza cambiamenti della sequenza del DNA che vengono trasmessi alle generazioni successive. Le modifiche epigenetiche sono modifiche chimiche, che non sono codificate dal genoma e coordinano come e quando i geni sono espressi. L'epigenomica esplora modifiche ereditabili e reversibili del DNA e della cromatina che non influenzano le sequenze nucleotidiche primarie. Mentre il termine epigenomica descriverebbe l'analisi dei cambiamenti epigenetici in molti geni in una cellula o in un intero organismo, l'epigenetica si concentra su processi che regolano come e quando specifici geni vengono accesi e spenti. Diversi fattori sono noti per influenzare la regolazione epigenetica: 1) Nutrizione (fattori dietetici); 2) Fattori ambientali; 3) Esposizione alle radiazioni; 4) Agenti infettivi; 5) Fattori immunologici; 6) Fattori genetici; 7) Agenti tossici; 8) Mutageni.

Metodi di rilevamento versatili vengono applicati per l'analisi delle modifiche epigenetiche nella cellula. La metilazione del DNA viene analizzata mediante saggi di digestione e sequenziamento bisolfito del DNA. Nei saggi di digestione, il DNA genomico è frammentato con endonucleasi sensibili alla metilazione e insensibili alla metilazione. Gli enzimi di restrizione sensibili alla metilazione tagliano solo il DNA non metilato e lasciano i frammenti di DNA metilati non digeriti. Il DNA frammentato può essere analizzato tramite sequenziamento o microarray, e quindi i siti di metilazione vengono mappati. Lo svantaggio di questo metodo è che studia solo le sequenze di DNA vicino agli obiettivi degli enzimi di restrizione scelti, e di solito viene utilizzato per caratterizzare solo i livelli globali di metilazione del DNA invece di identificare il DNA metilato a livello del singolo residuo. La risoluzione viene migliorata utilizzando il metodo del sequenziamento con bisolfito. Questo metodo specifico per il filamento viene utilizzato per convertire la citosina non metilata in uracile, mentre i residui metilati di citosina rimangono inalterati. Il DNA risultante viene amplificato in PCR e può essere analizzato sequenziando le regioni di interesse. L'analisi può essere eseguita anche utilizzando spettrometria di massa MALDI-TOF o microarray. La metilazione del DNA può anche essere studiata attraverso metodi basati sul principio della cromatografia di affinità. Il campione contenente il DNA frammentato viene caricato su una colonna con dominio di legame legante il metile (MBD) di MeCP2, specifico per il DNA metilato. Le frazioni di DNA metilate vengono eluite e analizzate con tecniche a livello genomico come MBDCap-seq/MethylCap-seq. Un altro approccio è l'immunoprecipitazione del DNA metilato (MeDIP), che si



2019-1-BG01-KA203-062371

basa sul legame specifico degli anticorpi alla citosina metilata (5mC) nel DNA. Questo metodo è utilizzato anche per l'analisi della metilcitosina idrossilata (5hmC). I frammenti purificati vengono analizzati con PCR, sequenziamento o microarray.

Una tecnica ampiamente utilizzata per il rilevamento delle modifiche istoniche è l'immunoprecipitazione della cromatina (ChIP). È usato per identificare le modifiche post-traduzionali locali delle code istoniche e per monitorare i cambiamenti nelle modifiche in risposta a diversi stimoli. Gli anticorpi contro specifiche modifiche istoniche (come la trimetilazione dell'istone H3 lisina K27) sono usati per l'immunoprecipitazione di quella regione di cromatina che hanno queste modifiche. Il ChIP può anche essere applicato per studiare il legame di fattori di trascrizione ed enzimi alla cromatina, utilizzando un anticorpo specifico contro il fattore di interesse. Le regioni di cromatina associate possono essere analizzate con PCR per rilevare loci specifici o sequenziamento estensivo per una visione più globale, così come utilizzando microarray (ChIP-chip), ma nell'ultimo caso la risoluzione dei dati è più debole rispetto al sequenziamento. Nell'ultimo decennio è stato sviluppato il metodo chip di base che porta all'emergere di altri diversi metodi ChIP come μ ChIP-seq per l'analisi della quantità di campioni a bassa, micro-scala e il moderno sequenziamento in tempo reale a singola molecola (SMRT; "sequenziamento di terza generazione") di campioni ChIP.

Le modifiche epigenetiche possono anche essere studiate con metodi che identificano le regioni attive della cromatina. Nel metodo DNase-seq, l'enzima DNasi I viene utilizzato per digerire il DNA che non è protetto dalla struttura del nucleosoma, quindi le regioni sensibili alla DNasi sono associate ai geni attivi. Le tecniche di sequenziamento, microarray o Southern Blot possono essere utilizzate per l'analisi e l'interpretazione dei risultati. La cromatina aperta può essere rilevata anche dall'isolamento assistito da formaldeide degli elementi regolatori (FAIRE), che si traducono nell'isolamento di regioni cromatiche aperte e prive di nucleosomi. Il metodo FAIRE si basa sul cross-linking tra formaldeide e DNA, istoni e altre proteine ad esso associate. Il DNA viene sonicato per essere frammentato e dopo ciò viene isolato con estrazione fenolo-cloroformio. Solo il DNA che non è legato dai nucleosomi e dalle proteine associate rimane nella fase acquosa dell'estrazione, con conseguente isolamento delle regioni aperte e attive del genoma. I frammenti isolati possono essere nuovamente analizzati con diversi metodi, come PCR, microarray e sequenziamento. La cromatina può anche essere studiata con la tecnica di cattura della conformazione della cromatina (3C) che identifica le regioni di cromatina che sono fisicamente associate insieme, come i promotori con gli enhancer. Il 3C è spesso analizzato tramite PCR, ma al giorno d'oggi anche tramite sequenziamento profondo delle interazioni a livello globale (Hi-C).

Gli RNA non codificanti (ncRNA) agiscono come modificatori epigenetici per regolare fortemente l'espressione genica. La loro espressione aberrante principalmente sotto forma di microRNA (miRNA) e di RNA lunghi non codificanti, può modificare l'espressione genica e innescare complicati disturbi immunitari. L'analisi della componente RNA dell'epigenetica spesso viene eseguita su scala genomica utilizzando metodi di sequenziamento di nuova generazione. I cambiamenti in ncRNA e mRNA possono essere caratterizzati con il sequenziamento estensivo. L'espressione degli RNA può essere analizzata anche mediante reazione quantitativa a catena della polimerasi (qPCR). La PCR quantitativa nota anche come PCR in tempo reale è una tecnica di laboratorio in biologia molecolare,



2019-1-BG01-KA203-062371

con la quale la quantità del prodotto PCR può essere determinata in tempo reale, ed è molto utile per studiare l'espressione genica.

3. Trascritomica

Quando i geni sono espressi, le informazioni genetiche memorizzate nel DNA vengono trasferite all'RNA (acido ribonucleico). Gli RNA sono importanti macromolecole, composte da catene lineari di nucleotidi (Fig. 4), che sono prodotte dal processo cellulare di trascrizione. Gli RNA svolgono diverse funzioni cellulari e biologiche in quanto servono da template per la sintesi proteica o svolgono ruoli critici di catalisi e regolazione. Nella trascrizione le informazioni genetiche vengono trasferite dal DNA all'mRNA. Questo processo viene eseguito dall'enzima RNA polimerasi. Nelle cellule esistono diverse classi di RNA (RNA messaggero (mRNA), RNA di trasferimento (tRNA), RNA ribosomiale (rRNA), piccolo RNA nucleare (snRNA), piccolo RNA nucleolare (snoRNA), corti RNA interferenti (siRNA), micro RNA (miRNA), RNA lungo non codificante (lncRNA) e pseudogeni, ma quelli che prendono parte alla sintesi proteica sono l'RNA messaggero (mRNA), l'RNA di trasferimento (tRNA) e l'RNA ribosomiale. L'RNA messaggero (mRNA) è una molecola a singolo filamento che media il trasferimento di informazioni genetiche ai ribosomi in cui le proteine sono sintetizzate. Negli eucarioti, ogni gene è trascritto per produrre un singolo mRNA, mentre nei procarioti, una singola molecola di mRNA può trasportare le informazioni genetiche di diversi geni; cioè diverse regioni codificanti proteine. Esiste una corrispondenza lineare tra la sequenza di base di un gene e la sequenza amminoacidica di un polipeptide. Ogni gruppo di tre nucleotidi consecutivi codifica la posizione di un particolare amminoacido in una molecola proteica e ognuna di queste triplette di basi è chiamata codone. I codoni sono tradotti in sequenze di amminoacidi da ribosomi (che a loro volta consistono in proteine e rRNA), tRNA e proteine di aiuto chiamate fattori di traduzione.

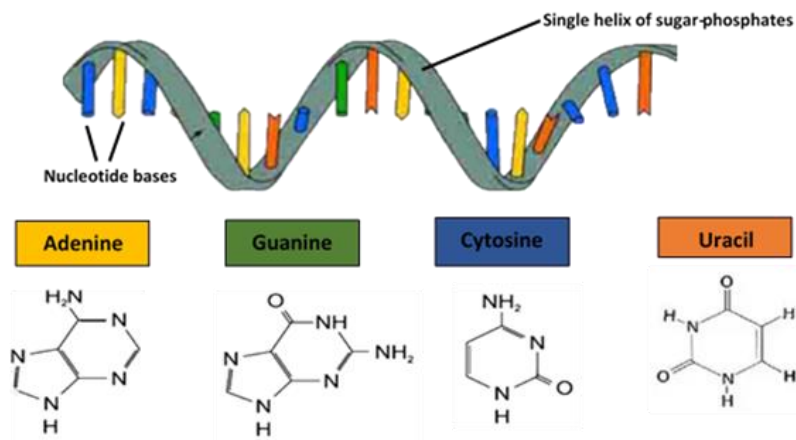


Fig. 4. Struttura a singolo filamento di RNA con le quattro basi nucleotidiche: Adenina, Uracile, Guanina, Citosina



2019-1-BG01-KA203-062371

Il termine "trascrittoma" è ampiamente usato per designare l'insieme completo di tutte le molecole di acido ribonucleico (RNA) in una cellula, tessuto o organismo. Il trascrittoma riflette l'attività molecolare nelle cellule, i geni che si esprimono attivamente in un dato momento. Comprende tutte le forme di molecole di RNA, comprese le molecole che codificano per proteine, che non codificano per proteine, derivate da splicing alternativo, alternativamente poliadenilate, alternativamente iniziate, senso, antisenso e trascrizioni modificate con RNA. Rispettivamente, la trascrittomica studia tutti i tipi di trascrizioni all'interno di una cellula o di un organismo, inclusi mRNA, miRNA e diversi tipi di long non-coding RNA (lncRNA). La trascrittomica copre tutto ciò che riguarda gli RNA come i loro livelli di trascrizione ed espressione, funzioni, luoghi, traffico e degradazione. Include anche le strutture dei trascritti e dei loro geni genitori per quanto riguarda i siti di inizio, le sequenze finali 5' e 3', i modelli di splicing e le modifiche posttrascrizionali.

I principali studi di trascrittomica sono finalizzati a:

- caratterizzazione dei diversi stati delle cellule (cioè stadi di sviluppo), tessuti o fasi del ciclo cellulare attraverso i modelli di espressione;
- studio dei meccanismi molecolari alla base di un fenotipo;
- identificazione di biomarcatori diversamente espressi tra lo stato mato e lo stato sano;
- differenziazione delle fasi o dei sottotipi della malattia (ad esempio stadi tumorali);
- stabilire la relazione causale tra varianti genetiche e modelli di espressione genica per evidenziare l'eziologia delle malattie.

Negli ultimi trent'anni il progresso tecnologico ha rivoluzionato il profilo del trascrittoma e ridefinito ciò che è possibile indagare. L'integrazione dei dati trascrittomici con altre omiche sta fornendo una visione sempre più integrata delle complessità cellulari facilitando approcci olistici alla ricerca biomedica.

Le principali tecniche applicate per lo studio del trascrittoma sono:

- Metodi basati su Expressed sequence tag (EST)
- SAGE
- Microarray basato sull'ibridazione
- Real-time PCR
- Metodi di sequenziamento dell'RNA basati su NGS (RNA-seq),
- RNA interference
- Strumenti bioinformatici per l'analisi dei trascrittomi.

La selezione della tecnica dipende dall'efficacia dei costi, dalla sensibilità, dall'elevata produttività e dalla concentrazione minima dell'RNA iniziale. La metodologia prevede l'isolamento dell'RNA, la purificazione, la quantificazione, la costruzione della libreria cDNA e il sequenziamento ad alta produttività.



2019-1-BG01-KA203-062371

- I metodi basati su Expressed sequence tag (EST) - I tag di sequenza espressi (EST) sono sequenze di DNA relativamente brevi (di solito 200-300 nucleotidi) solitamente generate dalle estremità 3' dei cloni di cDNA da cui i primer PCR possono essere derivati e utilizzati per rilevare la presenza della specifica sequenza codificante nel DNA genomico. Il sequenziamento degli EST fornisce una panoramica sul livello di espressione del gene. Poiché sempre più dati EST sono diventati disponibili al pubblico, l'uso degli EST si è esteso ad altre aree, come la scoperta di marcatori genetici *in silico*, la scoperta di geni *in silico*, la costruzione di modelli genetici, la previsione dello splicing alternativo, l'annotazione del genoma, il profilo di espressione e la genomica comparativa. Rispetto al sequenziamento dell'intero genoma, la tecnologia EST è più semplice e meno costosa, soprattutto nel caso di genomi di grandi dimensioni. Esistono database EST come dbEST (NCBI EST) (<https://www.ncbi.nlm.nih.gov/genbank/dbest/>) che contengono dati di sequenza e altre informazioni sulle sequenze cDNA "single-pass", o " Expressed Sequence Tags ", da un certo numero di organismi e quindi fungono da riferimento per il profilo di espressione di un organismo.

- L'analisi seriale dell'espressione genica (SAGE) è una tecnica trascrittomica utilizzata dagli scienziati per produrre un'istantanea della popolazione di RNA messaggero in un campione di interesse sotto forma di piccoli tag che corrispondono a frammenti di tali trascrizioni. Questa tecnica è vantaggiosa rispetto a EST perché vengono sequenziati solo brevi "tag" di circa 15 basi. I brevi frammenti generati vengono quindi uniti e sequenziati. Un pool di cDNA può essere sottoposto a NGS ad alta produttività noto come RNA-seq per la quantificazione, la scoperta di nuovi EST e il profilo degli RNA.

- Chip di DNA, chip genetico, biochip o microarray è una raccolta di spot di DNA, cDNA, oligonucleotidi attaccati a un supporto solido come vetro o chip di silicio. Attraverso questo metodo basato sull'ibridazione i livelli di espressione di migliaia di geni vengono monitorati contemporaneamente. La principale restrizione di questa tecnologia è che l'informazione sulla sequenza genomica è un prerequisito e anche un background più elevato inerente alla tecnica di ibridazione.

- La PCR quantitativa in tempo reale (qRT-PCR) è un tipo di PCR per una quantificazione affidabile di mRNA poco abbondanti o di trascrizioni a basso numero di copie. I principali vantaggi di questa tecnica sono la sua elevata sensibilità, una migliore riproducibilità e un ampio intervallo dinamico di quantificazione. Facilita le espressioni geniche e gli studi di regolazione anche in una singola cellula in base alla sua capacità di amplificazione esponenziale. La disponibilità di diversi tipi di sistema di monitoraggio della fluorescenza collegato alla PCR ha portato alla sua popolarità per gli studi sull'espressione genica. I problemi dell'amplificazione aspecifica, della formazione di dimeri di primer sono alcuni dei limiti del qRT-PCR.

- Il sequenziamento dell'RNA è una delle tecnologie avanzate ad alta produttività per la trascrittomica. RNA-seq, noto anche come sequenziamento shotgun dell'intero trascrittoma del, utilizza strumenti NGS. I vantaggi dell'RNA-seq sono che non si basa sulla disponibilità della sequenza genomica, non ha limiti di quantificazione superiori, mostra un'elevata riproducibilità e possiede un ampio intervallo dinamico di rilevamento.



2019-1-BG01-KA203-062371

Il progresso della trascrittomica ha portato ad una serie di scoperte biologiche. Questa tecnologia omica ha posto le basi dei primi veri studi multiomici, basati sui confronti tra sequenza di DNA ed espressione di mRNA. Al giorno d'oggi, l'analisi trascrizionale rimane più frequentemente utilizzata dalla maggior parte dei biologi poiché i dati ottenuti sono ancora più facilmente analizzabili e condivisi rispetto agli "omici a valle" come la proteomica e la metabolomica. Più recentemente, la trascrittomica sta godendo di una seconda rinascita, in quanto in molti casi è applicabile alle singole cellule.

4. Proteomica

Le proteine controllano la struttura e l'attività cellulare, forniscono i meccanismi per la segnalazione tra cellule e tessuti e catalizzano reazioni chimiche che supportano il metabolismo. La struttura proteica detta la funzione (o disfunzione). Le proteine possono essere la causa principale di malattie (come l'Alzheimer o il morbo di Huntington) e possono essere utilizzate per curarlo (ad esempio, gli anticorpi sono usati come terapie contro le infezioni virali e batteriche). L'insieme di tutte le proteine espresse in un sistema biologico in condizioni specifiche e definite è noto come proteoma. La parola "proteoma" è una combinazione di proteine e genoma ed è stata coniata da Mark Wilkins nel 1994.

L'obiettivo principale della proteomica è un'analisi sperimentale su larga scala della struttura e della funzione di questo intero insieme di proteine prodotte da un organismo vivente. Il termine "proteomica" è apparso per la prima volta nel 1997 e si riferisce a una tecnologia di base negli approcci di biologia dei sistemi che studia il proteoma. La funzione delle cellule dipende dalle proteine presenti nello spazio intra- e intercellulare e dalla loro abbondanza. La sintesi delle proteine nella cellula si basa sui precursori dell'mRNA, ma è impossibile prevedere l'abbondanza di proteine specifiche solo sulla base dell'analisi dell'espressione genica. La ragione di ciò è che le proteine native subiscono modifiche post-traduzionali (PTM) o alterazioni in risposta ai cambiamenti ambientali. Il proteoma è un riflesso dinamico sia dei geni che dell'ambiente ed è una fonte preziosa per la scoperta di biomarcatori perché è più probabile che le proteine siano ubiquitariamente colpite nella malattia e nella risposta alle malattie. La caratterizzazione completa di tutte le proteine è stato l'obiettivo della proteomica sin dal suo inizio quasi 25 anni fa. Tende a fare di più che identificare semplicemente le proteine potenzialmente presenti in un campione, ma anche a valutare l'abbondanza proteica, la localizzazione, le modifiche post-traduzionali, le isoforme e le interazioni molecolari. La proteomica mira a studiare il flusso di informazioni biologiche attraverso percorsi e reti proteiche, con l'obiettivo finale di comprendere la rilevanza funzionale delle proteine. Ciò richiede lo sviluppo di tecnologie in grado di rilevare una vasta gamma di proteine in campioni di origini diverse. Varie tecnologie sono utilizzate nella proteomica e il più delle volte vengono applicate in combinazione, ad esempio elettroforesi in gel mono o bidimensionale con spettrometria di massa (MS) o cromatografia liquida e SM.

Le tecniche convenzionali per l'isolamento e la purificazione delle proteine sono basate sulla cromatografia come la cromatografia a scambio ionico (IEC), l'idrossiapatite, la cromatografia ad



2019-1-BG01-KA203-062371

esclusione dimensionale (SEC) e la cromatografia per affinità. Per l'analisi di proteine selettive, è possibile utilizzare il saggio immunoassorbimento legato all'enzima (ELISA) e western blotting. Queste tecniche possono essere applicate per l'analisi delle singole proteine, ma non possono definire il livello di espressione proteica. Per la separazione di campioni proteici complessi vengono utilizzate tecniche di elettroforesi del gel di sodio dodecil solfato -poliacrilammide (SDS-PAGE), elettroforesi su gel bidimensionale (2-DE) ed elettroforesi su gel differenziale bidimensionale (2D-DIGE).

- La spettrometria di massa (MS) consente l'analisi dei proteomi e di solito è il metodo preferito per identificare le proteine presenti nei sistemi biologici. Le tre principali applicazioni della MS alla proteomica sono: catalogazione dell'espressione proteica, definizione delle interazioni proteiche e identificazione dei siti di modifica delle proteine. La spettrometria di massa misura il rapporto massa-carica (m/z) degli ioni in fase gassosa. Gli spettrometri di massa sono costituiti da una sorgente di ionica che converte le molecole di analita in ioni in fase gassosa, un analizzatore di massa che separa gli analiti ionizzati in base al rapporto m/z , e un rivelatore che registra il numero di ioni ad ogni valore m/z . Lo sviluppo della ionizzazione elettrospray (ESI) e del desorbimento/ionizzazione laser assistito da matrice (MALDI), le due tecniche di ionizzazione morbida in grado di ionizzare peptidi o proteine, ha rivoluzionato l'analisi proteica utilizzando la MS. L'analizzatore di massa è fondamentale per la tecnologia MS. Per la ricerca proteomica, quattro tipi di analizzatori di massa sono comunemente usati: quadrupolo (Q), trappola ionica (trappola ionica quadrupolo, QIT; trappola ionica lineare, LIT o LTQ), analizzatore di massa del tempo di volo (TOF) e analizzatore di massa di risonanza ciclotronica ionica a trasformata di Fourier (FTICR). Variano nei loro principi fisici e nelle loro prestazioni analitiche. Gli strumenti "ibridi" sono stati progettati per combinare le capacità di diversi analizzatori di massa e includono Q-Q-Q, Q-Q-LIT, Q-TOF, TOF-TOF e LTQ-FTICR.

- La spettrometria di massa tandem è una procedura sperimentale aggiuntiva nota anche come MS/MS o MS2. Questa è una tecnica chiave per il sequenziamento di proteine o peptidi e l'analisi PTM in cui due o più analizzatori di massa sono accoppiati. Le molecole di un dato campione sono ionizzate e il primo spettrometro (designato MS1) separa questi ioni in base al loro rapporto massa-carica (spesso dato come m/z o m/Q). Gli ioni di un particolare rapporto m/z proveniente da MS1 vengono selezionati e quindi fatti dividere in ioni di frammento più piccoli, ad esempio con dissociazione indotta dalla collisione (CID), dissociazione di cattura elettronica (ECD), dissociazione a trasferimento di elettroni (ETD), reazione ione-molecola o dissociazione fotografica. Questi frammenti vengono quindi introdotti nel secondo spettrometro di massa (MS2), che a sua volta separa i frammenti per il loro rapporto m/z e li rileva. La fase di frammentazione consente di identificare e separare gli ioni che hanno rapporti m/z molto simili negli spettrometri di massa regolari.

La complessità dei sistemi biologici richiede che il proteoma sia separato prima dell'analisi. Sia la cromatografia su gel che le separazioni a base di cromatografia liquida si sono dimostrate utili a questo proposito. Tipicamente, dopo queste estese separazioni, le proteine sono caratterizzate dall'analisi MS di proteine intatte (top-down) o peptidi proteici digeriti enzimaticamente (bottom-up). Le identificazioni proteiche vengono effettuate confrontando masse misurate di proteine intatte (top-down) o peptidi proteici digeriti (bottom-up) con le masse calcolate ottenute dai dati del genoma.



2019-1-BG01-KA203-062371

- Isobaric Tag Labelling (ITL) è un metodo comunemente usato per la quantificazione delle proteine. In proteomica la quantificazione dell'abbondanza proteica è un obiettivo importante. I livelli di espressione proteica rappresentano l'equilibrio tra la traduzione e la degradazione delle proteine nelle cellule. Si presume quindi che l'abbondanza di una proteina specifica sia correlata al suo ruolo nella funzione cellulare. ITL offre l'opportunità di identificare e quantificare simultaneamente le proteine da più campioni in un'unica analisi. Per misurare la quantità di proteine in un dato campione, i peptidi sono etichettati con tag chimici che hanno la stessa struttura e massa nominale, ma variano nella distribuzione degli isotopi pesanti nella loro struttura. Questi tag, comunemente indicati come tag di massa tandem, sono progettati in modo che l'etichetta di massa sia scissa in una specifica regione del linker su dissociazione indotta da collisione ad alta energia (HCD) durante la spettrometria di massa tandem che produce ioni reporter di diverse masse. La quantificazione delle proteine si ottiene confrontando le intensità degli ioni reporter negli spettri MS/MS. Altre tecniche quantitative sono l'etichettatura ICAT e l'Etichettatura Isotopica Stabile con Aminoacidi nella Coltura Cellulare (SILAC). L'ICAT ha anche ampliato la gamma di proteine che possono essere analizzate e consente l'accurata quantificazione e identificazione della sequenza delle proteine da miscele complesse. SILAC è un approccio basato sulla MS per la proteomica quantitativa che dipende dall'etichettatura metabolica dell'intero proteoma cellulare. I proteomi di diverse cellule coltivate in coltura cellulare sono etichettati con forma "leggera" o "pesante" di amminoacidi e differenziati attraverso la MS.

- La cristallografia a raggi X e la spettroscopia a risonanza magnetica nucleare (NMR) sono due importanti tecniche ad alta produttività che forniscono una struttura tridimensionale (3D) delle proteine che potrebbe essere utile per comprenderne la funzione biologica. Con il supporto di tecnologie ad alta produttività, viene raccolto un enorme volume di dati proteomici. Vengono istituite banche dati bioinformatiche per gestire un'enorme quantità di dati e la loro conservazione. Vari strumenti bioinformatici sono sviluppati per la previsione della struttura 3D, l'analisi del dominio proteico e del motivo, la rapida analisi dell'interazione proteina-proteina e l'analisi dei dati della SM. Gli strumenti di allineamento sono utili per l'allineamento della sequenza e della struttura per scoprire la relazione evolutiva. L'analisi del proteoma fornisce la rappresentazione completa delle informazioni strutturali e funzionali delle cellule, nonché il meccanismo di risposta della cellula contro vari tipi di stress e farmaci utilizzando tecniche di proteomica singola o multipla. Le principali tecniche utilizzate nella ricerca sulla proteomica sono fornite nella Fig. 5.



2019-1-BG01-KA203-062371

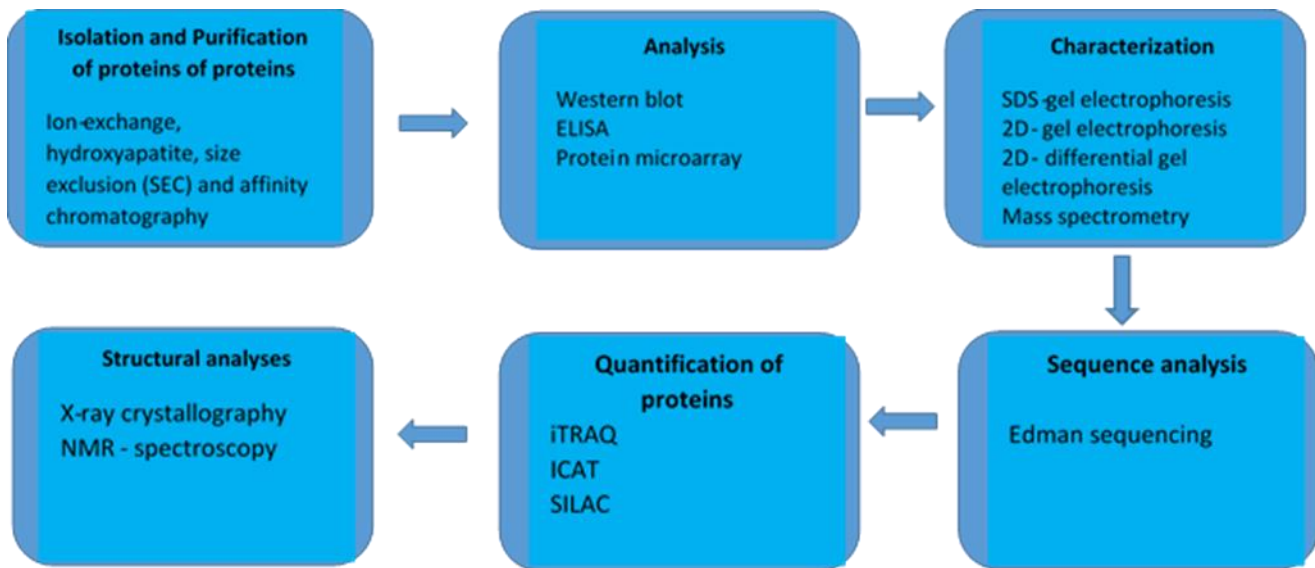


Fig. 5. Metodi applicati nel campo della proteomica per indagini su larga scala sulle proteine

5. Metabolomica

Il metaboloma è stato descritto per la prima volta da Oliver e colleghi nel 1998, durante il loro lavoro pionieristico sul metabolismo del lievito, come il complemento completo di piccole molecole in un sistema biologico o fluido. Comprende tutte le piccole molecole come lipidi, amminoacidi, acidi grassi, carboidrati e vitamine che sono noti come metaboliti e sono prodotti a seguito del metabolismo cellulare. Partecipano ai processi metabolici nelle cellule interagendo con altre molecole biologiche seguendo vie metaboliche. Lo stato dei metaboliti nei sistemi biologici è altamente variabile e dipende dal tempo e si osservano cambiamenti nei livelli dei metaboliti chiave a causa dei fattori genetici, ambientali, nutrizionali e di altro tipo.

La metabolomica (analisi del metaboloma) è la strategia introdotta più di recentemente tra le Omiche che identifica e quantifica sistematicamente i metaboliti presenti in una cellula, tessuto, organo, biofluidi o organismo in un determinato momento. I metodi utilizzati nella metabolomica mirano a misurare composti a basso peso molecolare (metaboliti) con diverse caratteristiche fisiche come polarità del composto, gruppi funzionali e somiglianza strutturale. Sulla base di queste proprietà il metaboloma è diviso in sottoinsiemi di vari metaboliti e per la loro indagine le procedure analitiche ottimizzate per ogni tipo di molecola vengono applicate nella metabolomica.



2019-1-BG01-KA203-062371

I metaboliti sono prodotti di reazioni biochimiche che vengono effettuate con la partecipazione di proteine del proteoma (Fig. 6). Le loro concentrazioni dipendono dalle interazioni di altri processi (trascrizione, traduzione, segnalazione cellulare, ecc.), sono stati studiati come reporter per il metabolismo nella PD. Questo a sua volta determina la struttura biologica e la funzione del fenotipo finale dell'organismo. Pertanto, i cambiamenti nell'espressione genica, la funzione delle proteine e l'ambiente influenzano direttamente la concentrazione dei metaboliti in un sistema biologico. Il metaboloma è composto da un numero relativamente piccolo (circa 5000), ma diversi tipi di metaboliti che rendono il metaboloma più complesso fisicamente e chimicamente rispetto al genoma, al trascrittoma e al proteoma. Inoltre, il genoma, il trascrittoma e il proteoma sono costituiti da composti che fanno parte del metaboloma. Inoltre, i componenti del metaboloma sono altamente conservati tra gli organismi rispetto al genoma, al trascrittoma e al proteoma, motivo per cui si ritiene che il metaboloma sia evolutivamente la parte più antica della cellula.

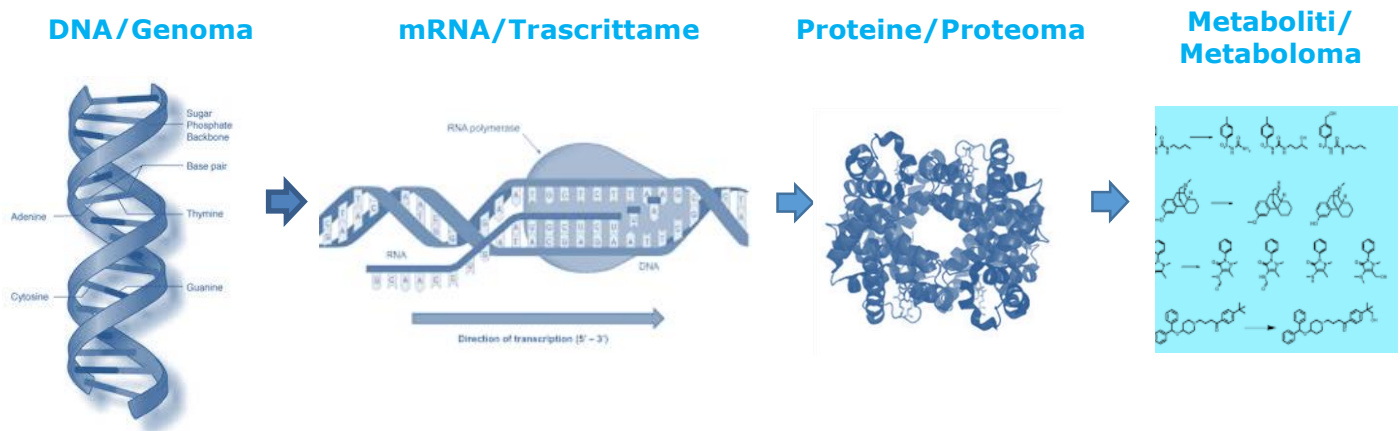


Fig.6. Approcci di integrazione dei dati in Biologia dei Sistemi

La metabolomica è la strategia più appropriata per la determinazione dei biomarcatori associati alla malattia. Lo studio completo dei metaboliti è uno strumento auspicabile per diagnosticare le malattie, identificare nuovi bersagli terapeutici e consentire trattamenti appropriati. Le analisi di un piccolo numero di metaboliti sono utilizzate nella diagnostica delle malattie per decenni, ad esempio lo sviluppo di strisce per il test del glucosio nel sangue negli anni '50 per il test del diabete o la quantificazione della fenilalanina nei neonati per lo screening della fenilchetonuria.

Le analisi della metabolomica possono essere raggruppate in due approcci principali: metabolomica mirata e non mirata. Nell'analisi mirata i metaboliti che vengono determinati sono noti e rappresentano percorsi o classi specifiche di molecole. Al contrario, la metabolomica non mirata mira a quantificare e identificare il maggior numero possibile di metaboliti. La metabolomica mirata si riferisce



2019-1-BG01-KA203-062371

alla quantificazione assoluta (nM o mg/ml) e utilizza uno standard interno e un'analisi semi quantitativa o quantitativa per rilevare composti noti relativi a percorsi specifici. L'approccio non mirato misura tutti i metaboliti presenti in un campione e applica la quantificazione relativa (fold change) e il confronto tra campioni. Il ruolo chiave che la metabolomica svolge nell'integrazione multi-omica e nella modellazione dei sistemi è dovuto al fatto che può essere quantitativa. La modellazione dei sistemi non può essere eseguita senza valori accurati o concentrazioni accurate come input e, allo stesso modo, i modelli di sistema non possono essere facilmente verificati senza concentrazioni accurate e quantitative come output. La metabolomica può fornire sia (dati quantitativi di input che di output), rendendola estremamente preziosa per i modellisti di sistemi.

A causa dell'enorme complessità chimica del metaboloma, non può essere studiato in modo completo da una singola tecnologia. La prima applicazione della metabolomica risale agli anni '70, quando la gascromatografia-spettrometria di massa fu utilizzata per il profilazione dei metaboliti di campioni clinici di urina. Nicholson e altri negli anni '80 hanno fatto un profilo di campioni clinici mediante l'applicazione della spettroscopia di risonanza magnetica nucleare. La spettrometria di massa accoppiata sia con la cromatografia liquida che con l'elettroforesi capillare viene applicata anche in metabolomica per completare le tecniche analitiche disponibili. Alcuni degli strumenti analitici utilizzati ai fini della ricerca sulla metabolomica vengono applicati frequentemente e regolarmente, mentre altri hanno ruoli specifici e sono applicati meno frequentemente.

- Gascromatografia - Spettrometria di massa (GC-MS) - Tra i metodi comunemente usati nella metabolomica c'è GC-MS. Come suggerisce il nome, GC-MS unifica due tecniche per formare un unico metodo per analizzare miscele di sostanze chimiche. La gascromatografia separa i componenti di una miscela e la spettroscopia di massa caratterizza ciascuno dei componenti singolarmente. Combinando le due tecniche, ogni metabolita in un campione può essere valutato sia qualitativamente che quantitativamente. Questa tecnica cromatografica viene utilizzata per studiare metaboliti che hanno un basso punto di ebollizione e che saranno presenti nella fase gassosa all'intervallo di temperatura 50-350°C. Questi metaboliti possono avere un basso punto di ebollizione nella loro forma biologicamente nativa o il punto di ebollizione di un metabolita può essere diminuito attraverso un'alterazione chimica, nota anche come derivatizzazione chimica. Il campione contenente metaboliti viene introdotto (iniettato) in una fase mobile, che nella gascromatografia è un gas inerte come l'elio. La fase mobile trasporta la miscela di campioni attraverso quella che viene definita una fase stazionaria. I metaboliti sono separati in base al loro adsorbimento alla fase stazionaria. La fase stazionaria è solitamente contenuta in una colonna di vetro o acciaio inossidabile e rappresenta una sostanza chimica in grado di adsorbire selettivamente i componenti in una miscela di campioni. Cambiando le caratteristiche della fase mobile e della fase stazionaria, è possibile separare diverse miscele di sostanze chimiche. Quando i singoli metaboliti eluiti dalla colonna GC, entrano nel rivelatore di ionizzazione elettronica (mass spec). Lì, sono bombardati da un flusso di elettroni ad alta energia (70 eV) che li fa rompere in frammenti. Questi frammenti possono essere pezzi grandi o piccoli delle molecole originali. Lo spettro di massa ottenuto per un dato composto chimico è fondamentalmente lo stesso ogni volta. Pertanto, lo spettro di massa è



2019-1-BG01-KA203-062371

un'impronta digitale per la molecola. Questa impronta digitale può essere utilizzata per identificare il composto.

- Cromatografia liquida-Spettrometria di massa (LC-MS) - La differenza principale tra GC-MS e LC-MS è che nella cromatografia liquida (LC), la fase mobile è un solvente. L'LC separa la miscela di metaboliti che sono in forma liquida, di solito contiene metanolo, acetonitrile e acqua. Utilizzando diversi imballaggi di colonne (diversa fase stazionaria) ad alta efficienza è possibile separare una piccola quantità di miscela complessa. Le colonne usate nell'HPLC sono costituite da imballaggi ottadecilici (C18), ottilici (C8), ciano, amminici, fenilici e generalmente la loro lunghezza è di circa 50mm a 300mm. Le colonne sono utilizzate sulla base della natura dei composti da separare. Questa miscela liquida contenente componenti viene trasferita nella fonte ioniche dello spettrometro di massa in cui avviene il processo di ionizzazione e si formano goccioline che trasportano un eccesso di carica elettrica positiva o negativa. La ionizzazione potrebbe essere ottenuta utilizzando diversi tipi di sorgenti e interfacce di ionizzazione. Dopo la ionizzazione gli ioni vengono trasferiti nell'analizzatore di massa dove la separazione degli ioni viene effettuata in base al loro rapporto massa-carica (m/z).

- Elettroforesi capillare-Spettrometria di massa (CE-MS) - questa tecnica analitica è anche chiamata elettroforesi a zona capillare-spettrometria di massa. CE-MS comporta la separazione delle specie ioniche in fase liquida attraverso l'applicazione di alte tensioni ed è solitamente accoppiato a una spettrometria di massa elettrospray. L'elettroforesi capillare separa i metaboliti in base alla loro mobilità elettroforetica in una soluzione di elettrolita liquido operante in un campo elettrico. La mobilità elettroforetica dipende dalla carica e dalle dimensioni del metabolita e quindi è possibile la separazione di metaboliti con dimensioni e/o cariche diverse. Tutti i metaboliti devono essere caricati per consentire la mobilità.

In GC-MS, LC-MS e CE-MS c'è una separazione dei metaboliti prima del loro rilevamento con spettrometria di massa. Questa peculiarità fornisce la capacità di rilevare metaboliti a basse concentrazioni, comunemente nanomoli/litro (nM/L) o micromoli/litro (µM/L).

- La spettroscopia a risonanza magnetica nucleare (NMR) è lo strumento analitico più utilizzato nella ricerca sulla metabolomica insieme alla spettrometria di massa. La spettroscopia NMR applica le proprietà magnetiche dei nuclei atomici in un metabolita. Solo alcuni atomi sono attivi NMR e includono ^1H , ^{13}C , ^{31}P ; protoni (^1H). La tecnica funziona posizionando un campione liquido in un piccolo tubo di diametro interno (ad esempio un tubo da 5 mm), o occasionalmente un pezzo di tessuto viene studiato direttamente utilizzando uno speciale portacampioni. Il campione viene pulsato con una gamma di frequenze radio che coprono tutte le energie possibili necessarie per eccitare il tipo selezionato di nuclei. I nuclei assorbono energia a diverse frequenze radio a seconda del loro ambiente chimico e quindi viene misurato il rilascio di questa energia, formando un decadimento a induzione libera (FID). Questo FID viene convertito da un set di dati del dominio del tempo a un dominio di frequenza - usando una trasformazione di Fourier - e uno spettro NMR è costruito come lo spostamento chimico (effettivamente l'energia di assorbimento) tracciato contro l'intensità di picco.



2019-1-BG01-KA203-062371

- Tecniche di spettroscopia vibrazionale come la spettroscopia trasformata di Fourier-infrarossa (FT-IR) e la spettroscopia Raman vengono applicate per l'analisi dei cambiamenti metabolici nei campioni biologici. Il principio delle tecniche si basa sulla trasmissione della luce ultravioletta o infrarossa attraverso un campione, o talvolta sul riflesso della luce dal campione, prima del suo rilevamento. Gli approcci misurano prevalentemente le vibrazioni e le rotazioni dei legami relativi a diversi gruppi funzionali chimici derivanti dall'interazione del campione con la luce ultravioletta o infrarossa. Queste tecniche di solito mancano della specificità per rilevare ogni metabolita separatamente, ma invece parti specifiche della molecola assorbono la luce ultravioletta o infrarossa a lunghezze d'onda specifiche. Ad esempio, nella FT-IR, le vibrazioni di allungamento C-H caratteristiche delle catene di acidi grassi si osservano tra l'intervallo del numero d'onda 3100 - 2800 cm^{-1} e la regione compresa tra 1800 e 1500 cm^{-1} è dominata da bande di ammidi I e ammidi II che indicano la predominanza delle strutture alfa elica o beta.

Nel campo della metabolomica, molte centinaia di campioni vengono regolarmente analizzati e di solito vengono rilevate almeno diverse centinaia di metaboliti. I dati ottenuti dall'individuazione dei metaboliti in un dato campione biologico sono seguiti da ulteriori analisi con metodi statistici multivariati (chemiometria) per estrarre informazioni biologiche, fisiologiche e clinicamente rilevanti.

VANTAGGI E SVANTAGGI DELLE TECNOLOGIE OMICHE

L'implementazione delle tecnologie omiche e l'integrazione dei dati omici è stata realizzata per un'ampia gamma di aree di ricerca, tra cui scienze degli alimenti e della nutrizione, microbiologia dei sistemi, analisi dei microbiomi, interazioni genotipo-fenotipo, biologia dei sistemi, scoperta di prodotti naturali e biologia delle malattie. Gli approcci basati sull'omica sono stati significativamente migliorati con l'aggiunta di nuovi concetti come l'esposoma/esposomica (ruolo dell'ambiente nelle malattie umane), l'adduttomica (studio dei composti che legano il DNA e causano danni e mutazioni), la volatilomica (studio dei composti organici volatili all'analisi metabolomica / lipidomica), nutrigenomica (studio di come gli alimenti influenzano i nostri geni), ecc. Tuttavia, le tecnologie omiche sono rivolte principalmente a quattro campi di ricerca omica - genomica, trascrittomica, proteomica e metabolomica. I settori di ricerca tradizionali come la genetica, la farmacogenetica e la tossicogenetica che non applicano il concetto di omica forniscono le sequenze statiche di geni e proteine. Al contrario, le tecnologie omiche forniscono contemporaneamente la misurazione del numero di proteine, geni, metaboliti e consentono di ottenere dati su larga scala in breve tempo. L'analisi collettiva dei processi biologici in un sistema biologico fornito da tutte le omiche è un vantaggio dei precedenti metodi tradizionali per una migliore comprensione dell'intero quadro del sistema biologico. Sebbene i dati omici su larga scala diventino più accessibili, l'integrazione di dati multi-omici reali è ancora molto impegnativa. Ciò è dovuto al fatto che molti degli strumenti analitici specifici e dei disegni sperimentali che sono convenzionalmente utilizzati per le singole discipline omiche (genomica, trascrittomica e



2019-1-BG01-KA203-062371

proteomica) non sono sufficientemente adatti per consentire confronti e integrazioni affidabili tra più discipline omiche. Ad esempio, i metodi per la raccolta e la conservazione dei campioni e la quantità e il tipo di campioni biologici richiesti negli studi di genomica spesso non sono compatibili con la metabolomica, la proteomica o la trascrittomica. Durante la raccolta dei dati, ci sono molti problemi come l'eterogeneità dei dati, le piccole dimensioni del campione rispetto a molti parametri, la conferma e l'interpretazione dei dati a causa di molte interazioni nel sistema biologico e informazioni carenti su tali sistemi. Quasi in tutti gli esperimenti dell'omica vengono misurate da centinaia a migliaia di molecole bersaglio e variabili (metaboliti, proteine, geni, trascrizioni e SNP). Ciò richiede lo studio di un numero sufficiente di campioni biologici allo scopo di fornire risultati statisticamente affidabili. Inoltre, i dati multi-omici devono essere generati dallo stesso insieme di campioni per consentire il confronto diretto nelle stesse condizioni. Tuttavia, ciò non è sempre possibile a causa delle limitazioni nella biomassa campione, nell'accesso ai campioni o nelle risorse finanziarie. Nei singoli esperimenti di omica, sono spesso necessarie dimensioni di campione più grandi per superare questo problema. Inoltre, i dati multi-omici integrati devono essere analizzati come singoli set di dati prima di essere depositati in database specifici dell'omica al fine di renderli disponibili al pubblico. Questi problemi sottolineano che per gli studi ad alta qualità su larga scala è necessario: 1) pianificare correttamente un esperimento, 2) raccogliere, preparare e memorizzare attentamente campioni biologici, 3) raccogliere dati multi-omici quantitativi e meta-dati associati con attenzione, 4) utilizzare strumenti adeguati per l'integrazione e l'interpretazione dei dati.

Alcuni dei vantaggi e degli svantaggi delle principali tecnologie economiche sono riassunti nella tabella 1.



2019-1-BG01-KA203-062371

Tabella 1. Vantaggi e svantaggi delle principali tecnologie di ricerca omica

Tecnologie Omiche	Vantaggio	Svantaggio
Genomica	<ul style="list-style-type: none"> • Analisi del genoma completo • Studio del polimorfismo genico tra gli individui • Attraverso l'identificazione dei SNPs viene fornita una preziosa informazione per la diagnosi precoce e il trattamento delle malattie • Tecniche di laboratorio facili da implementare 	<ul style="list-style-type: none"> • L'analisi del genoma eseguita non è sufficiente per prevedere l'effetto biologico finale del DNA a causa dell'epigenetica, dei cambiamenti post-trascrizionali e post-traslazionali • Richiede attrezzature da laboratorio specifiche e costose
Epigenomica	Fornire conoscenze sulla regolazione dell'espressione genica	<ul style="list-style-type: none"> • È difficile mettere in relazione i dati epigenetici ottenuti con l'espressione genica, perché la trascrizione può essere influenzata da più processi • Non tutte le regioni metilate possono essere rilevate con le tecniche applicate
Trascritomica	<ul style="list-style-type: none"> • Studio dell'insieme completo dell'mRNA (trascrittoma) • Identificazione dei principali percorsi coinvolti nella risposta e nella tossicità ai farmaci. • Tecniche di laboratorio facili da implementare 	<ul style="list-style-type: none"> • L'espressione proteica è influenzata da cambiamenti post-traslazionali che portano a dati errati • Richiede attrezzature costose e specifiche
Proteomica	<ul style="list-style-type: none"> • Studio dell'insieme completo delle proteine in un sistema biologico • Rilevazione di proteine sconosciute e inaspettate 	<ul style="list-style-type: none"> • Attrezzature costose e procedure dispendiose in termini di tempo non applicabili all'intero proteoma • Alcune proteine sono difficili da separare e purificare • L'applicazione della MS e l'interpretazione dei dati richiedono personale appositamente formato
Metabolomica	<ul style="list-style-type: none"> • Fornisce dati adeguati sulle modifiche nei processi metabolici nella cellula • I metaboliti endogeni sono inferiori a geni, trascrizioni e proteine, quindi meno dati devono essere interpretati • Individuazione dei biomarcatori delle malattie 	<ul style="list-style-type: none"> • Numero relativamente basso di metaboliti (poche migliaia), come si può misurare • Attrezzature specifiche e costose • L'uso di MS e NMR richiede personale appositamente formato



2019-1-BG01-KA203-062371

I set di dati Multi-omics possono fornire una maggiore profondità di comprensione in determinati scenari, ma questo non è privo di costi. Questi studi si basano spesso su un gran numero di confronti, sul tipo di dati corretti, sulle analisi statistiche pertinenti, sulle attrezzature specifiche e sul personale qualificato e su notevoli investimenti di tempo e denaro. Quando si progetta un esperimento si deve tener conto di quali tipi di dati omici possono e devono essere integrati per ottenere potenti intuizioni biologiche del sistema studiato. L'applicazione di piattaforme omiche ad alta produttività non è sempre necessaria per rispondere alla domanda di ricerca. I risultati ottenuti dalle tecniche tradizionali, come la reazione quantitativa a catena della polimerasi (qPCR), il saggio immunosorbente legato all'enzima (ELISA), l'immunoistochimica (IHC), a volte possono essere sufficienti a fornire comprensione per i meccanismi biologici. Le tecniche spesso integrano i risultati di uno studio omico più ampio in quanto vengono applicate per verificare che la molecola significativa identificata dai dati omici sia un vero risultato positivo.

Referenze:

Altaf-UI-Amin M, Afendi FM, Kiboi SK, Kanaya S. 2014. Systems Biology in the Context of Big Data and Networks, *BioMed Res Int*, 1:11.

Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH. 2017. Proteomics: Technologies and Their Applications. *J Chromatogr Sci*, 55(2).

Benavente L, Goldberg A, Myburg H, Chiera J, Zapata M & van Zyl L. 2015. The application of systems biology to biomanufacturing. *Pharm. Bioprocess*, 3(4):341–355.

Bertolaso M, Giuliani A, de Gara L. 2010. *Systems Biology Reveals Biology of systems*. Wiley Periodicals, Inc., 16 (6).

Bloom FE. 2001. What does it all mean to you? *J Neurosci*, 21:8304–8305.

Clish CB. 2015. Metabolomics: an emerging but powerful tool for precision medicine. *Cold Spring Harb Mol Case Stud*, 1: a000588

Coorsen JR. 2013. Proteomics. *Brenner's Encyclopedia of Genetics (Second Edition)*, p. 508-510.



2019-1-BG01-KA203-062371

Dos Santos EAF, Santa Cruz EC, Ribeiro HC, Barbosa LD, Zandonadi FS, Sussulini A. 2020. Multi-omics: An Opportunity to Dive into Systems Biology. *Braz J Anal Chem*, 7(29): 18-44.

Haas R, Zelezniak A, Iacovacci J, Kamrad S, Townsend S, Ralser M. 2017. Designing and interpreting ‘multi-omic’ experiments that may change our understanding of biology. *Curr Opin Syst Biol*, 6:37–45.

Han X, Aslanian A, Yates JR. 2008. Mass Spectrometry for Proteomics. *Curr Opin Chem Biol*, 12(5): 483–490.

Horgan RP, Kenny LC. 2011. SAC review ‘Omic’ technologies: genomics, transcriptomics, proteomics and metabolomics. *Obstet Gynecol*, 13:189–195

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng R, Bumgarner JK, Goodlett DR, Aebersold R, Hood L. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292:929–934.

Ishii N, Tomita M. 2009. Multi-Omics Data-Driven Systems Biology of *E. coli*. In: Lee S.Y. (eds) *Systems Biology and Biotechnology of Escherichia coli*. Springer, Dordrecht.

Kalavacharla V, (Kal), Subramani M, Ayyappan V, et al. 2017. Plant Epigenomics. *Handbook of Epigenetics*, 245–258.

Karahalil B. 2016. Overview of Systems Biology and Omics Technologies. *Curr Med Chem*, 23:1-10.

Karamperis K, Wadge S, Koromina M. 2020. Genetic testing. *Applied Genomics and Public Health (Translational and Applied Genomics)*, Elsevier, 189-207.

Laurence P. 2015. The case of the gene: Postgenomics between modernity and postmodernity. *EMBO Reports*, 16 (7): 777–781.

Liang KH. 2013. Transcriptomics. *Bioinformatics for Biomedical Science and Clinical Applications*. Woodhead Publishing, p. 49-82.

Lin, S., Fang, L., Li, C., Liu, G. 2019. Epigenetics and heritable phenotypic variations in livestock. In: Tollefsbol T. (editor). *Transgenerational Epigenetics*. 2nd edition. Amsterdam, Netherlands: Elsevier, p. 283-313

Milward EA, Shahandeh A, Heidari M, Johnstone DM, Daneshi N, Hondermarck H. 2016. Transcriptomics. *Encyclopedia of Cell Biology*. Elsevier, The Netherlands, 4: 160-165.



2019-1-BG01-KA203-062371

O'Donnell ST, Ross RP, Stanton C. 2020. The Progress of Multi-Omics Technologies: Determining Function in Lactic Acid Bacteria Using a Systems Level Approach. *Front. Microbiol*, 10: 1-17.

Pereira C, Adamec BJ. 2019. Metabolome Analysis. *Encyclopedia of Bioinformatics and Computational Biology*, 3:463-475

Pinu FR, Beale DJ, Paten AM, Kouremenos K, et al. 2019. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites*, 9(76).

Potters, G. 2010. Systems Biology of the Cell. *Nature Education*, 3(9):33.

Pratima NA, Gadikar R. 2018. Liquid Chromatography-Mass Spectrometry and Its Applications: A Brief Review. *Arc Org Inorg Chem Sci*, 1(1): 26-34.

Shah TR, Ambikanandan M. 2011. Proteomics. Challenges in Delivery of Therapeutic Genomics and Proteomics. Elsevier, p. 387-427.

Strange K. 2005. The end of “naïve reductionism”: rise of systems biology or renaissance of physiology? *Am J Physiol Cell Physiol*, 288: C968–C974.

Turunen TA, ... Ylä-Herttua S. 2018. Epigenomics. *Encyclopedia of Cardiovascular Research and Medicine*, p. 258-265.

Vlaanderen J, Moore LE, Smith MT, et al. 2010. Application of omics technologies in occupational and environmental health research: current status and projections. *Occup Environ Med*, 67: 136-143.

Yadav D, Tanveer A, Malviya N, Yadav S. 2018. Overview and Principles of Bioengineering: The Drivers of Omics Technologies and Bio-Engineering Towards Improving Quality of Life, p.3-23

Yang X. 2020. Multitissue Multiomics Systems Biology to Dissect Complex Diseases. *Trends Mol Med*, 26 (8).

A brief guide to genomics. <https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics>

Analytical Techniques applied in Metabolomics. <https://www.futurelearn.com/info/courses/metabolomics/0/steps/10710>

Boundless biology. <https://courses.lumenlearning.com/boundless-biology/chapter/the-genetic-code/>



Project website: www.digit-biotech.eu

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.