



METABOLOMICA E BIG DATA: DALLA CARATTERIZZAZIONE COMPLETA DI UN ORGANISMO ALL'IDENTIFICAZIONE DI BIOMARKER DIAGNOSTICI

La metabolomica rappresenta la disciplina più recente tra le scienze “omiche” e ha registrato una notevole crescita negli ultimi anni, grazie ai progressi nelle tecniche di misura e analisi. In particolare, negli ultimi anni, questa disciplina ha fatto enormi progressi, trovando ampie applicazioni in vari settori, tra cui la diagnostica, la biologia vegetale, la tossicologia, la microbiologia, l'industria agroalimentare e il settore medico-farmaceutico.

Introduzione

La metabolomica è la disciplina che si occupa dello studio delle piccole molecole chimiche, note come metaboliti. Quest'area di ricerca ha una vasta gamma di applicazioni che si estendono in numerosi settori della biologia e della biotecnologia. Esse comprendono la biochimica umana, vegetale e microbica, la scoperta di nuovi farmaci e biomarcatori, la tossicologia, la nutrizione e il controllo alimentare. Inoltre, la metabolomica ha contribuito alla scoperta di biomarcatori precoci di malattie umane come l'Alzheimer [1], il diabete di tipo 2 [2] e la progressione del cancro alla prostata [3]. Alcuni esempi del potenziale della metabolomica includono la comprensione di come le piante utilizzino piccole molecole nella loro difesa contro i patogeni [4] e come la modulazione della dieta possa influenzare il metabolismo [5]. Tuttavia, nonostante l'interesse diffuso per la metabolomica, questa disciplina non ha progredito allo stesso ritmo di altre scienze “omiche”. Ci sono diverse sfide analitiche da affrontare durante un'analisi metabolomica, dalla progettazione dello studio all'analisi dei dati. A differenza dei geni, dei trascritti o delle protei-

ne, che sono composti da sequenze ben definite di monomeri (nucleotidi e amminoacidi) e possono essere analizzati con tecnologie consolidate e strumenti bioinformatici, i metaboliti sono entità chimiche il cui sorgere non segue una sequenza predefinita all'interno delle cellule. La diversità chimico-fisica delle strutture dei metaboliti nei vari organismi deriva da una serie di trasformazioni chimiche, spesso catalizzate da enzimi.

Attualmente, l'analisi dei metaboliti può essere eseguita su una vasta gamma di campioni biologici, tra cui fluidi corporei umani e animali come urine, siero e plasma del sangue, tessuti ed estratti cellulari, tessuti vegetali e loro estratti, organismi modello come lieviti o batteri, studi cellulari *in vitro* e altri.

I metaboliti possono essere suddivisi in due categorie principali: endogeni, che sono prodotti naturalmente dall'organismo (ad esempio amminoacidi, acidi organici, zuccheri), ed esogeni, che includono farmaci, contaminanti ambientali, additivi alimentari, tossine e altre sostanze estranee assorbite dall'organismo attraverso l'ambiente. La grande diversità delle proprietà chimico-fisiche dei

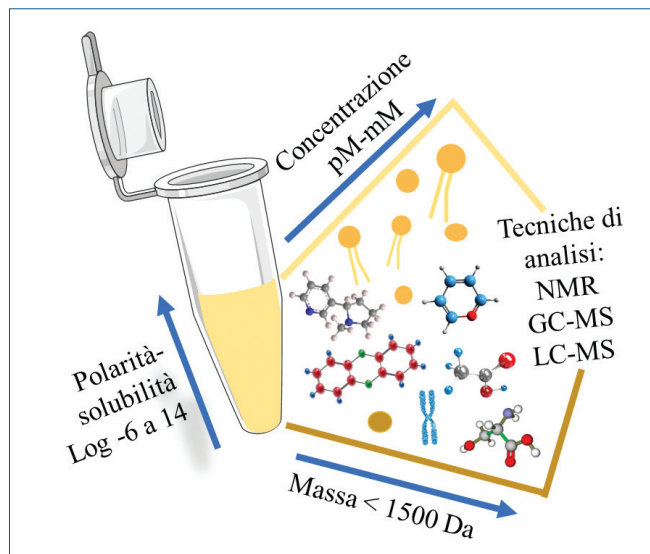


Fig. 1 - La gamma dinamica dei metaboliti è piuttosto ampia in termini di dimensioni, concentrazione e proprietà fisico-chimiche. Le dimensioni dei metaboliti variano di diverse grandezze d'ordine (~100-1500 Da), le concentrazioni vanno dai picomolari ai millimolari e le proprietà fisico-chimiche come la stabilità, l'idrofobicità, la stereochemica, la polarità, la volatilità e il pK_a possono mostrare notevoli differenze. Tipicamente, sono necessari diversi tipi di analisi analitiche per una copertura completa del metaboloma, anche assumendo che i metodi di estrazione siano ottimali

metaboliti si traduce in una vasta gamma di caratteristiche, come polarità, peso molecolare, gruppi funzionali, stereochemica, stabilità chimica e reattività. Questa diversità rende impossibile analizzare tutti i metaboliti di un campione con una singola tecnologia analitica. Pertanto, sono necessarie diverse configurazioni strumentali per coprire l'intero metaboloma, una sfida unica rispetto alla genomica o alla proteomica (Fig. 1).

Attualmente, due sono le piattaforme tecnologiche comunemente utilizzate per l'identificazione e la quantificazione dei metaboliti: la spettroscopia a risonanza magnetica nucleare (NMR) e la spettrometria di massa ad alta risoluzione (HRMS), spesso utilizzate insieme a tecniche cromatografiche, come la cromatografia liquida (LC) e la cromatografia gassosa (GC). Altre tecniche accoppiate, come l'elettroforesi capillare (CE-MS) sono utilizzate in misura minore. La metabolomica presenta sfide uniche, tra cui l'ampia variazione nella concentrazione dei metaboliti, la presenza di contaminanti nella fase mobile, il rumore strumentale nei profili metabolici, la presenza di numerosi composti non identificati in singoli campioni e la rapida variabili-

tà temporale e spaziale dei metaboliti cellulari. La sovrapposizione di alcuni metaboliti può causare interferenze chimiche che limitano la capacità di profilare altri metaboliti con successo. L'identificazione e la caratterizzazione dei metaboliti rimangono una sfida, ma in alcuni casi, l'analisi statistica dei dati può portare a informazioni utili, anche se non si identificano tutti i metaboliti presenti in un campione. Ad esempio, potrebbe emergere che un metabolita particolare, sconosciuto, è correlato a un tratto di interesse (malattia o specifico trattamento), e quindi essere presente in tutti i campioni test rispetto ai campioni di controllo. Questo approccio permette di ridurre il numero di metaboliti da identificare in modo significativo [6].

Work-flow di analisi

Per condurre un rigoroso esperimento di metabolomica, è di fondamentale importanza progettare un schema di lavoro che sia rigoroso. In Fig. 2, è mostrato un tipico work-flow di metabolomica. Brevemente, faremo un *excursus* delle varie fasi del lavoro.

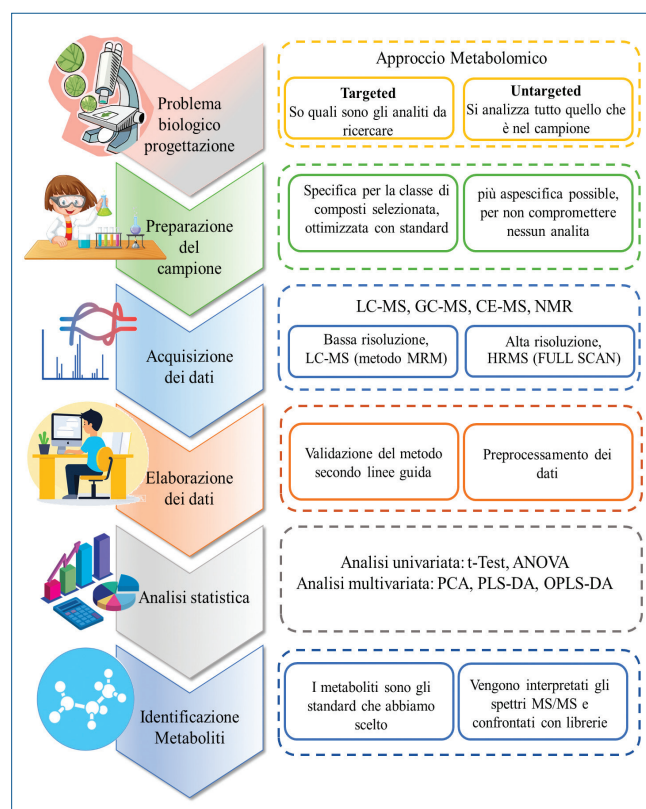


Fig. 2 - Work-flow analitico per uno studio metabolomico

Problema biologico e progettazione sperimentale

Il primo punto di fondamentale importanza riguarda la chiara e diretta formulazione del problema biologico da affrontare. Questa fase riveste una cruciale importanza in quanto orienterà la successiva progettazione sperimentale. In base al problema biologico identificato, si definiscono diversi aspetti, tra cui l'approccio metabolomico da adottare (*targeted* o *untargeted*), il tipo di campione da utilizzare (fluidi biologici, tessuti, cellule o organismi integri), le dimensioni del campione (il numero di campioni da valutare), le condizioni sperimentali a cui verranno sottoposti i campioni, la frequenza di raccolta dei campioni, le modalità di inattivazione dell'attività enzimatica (attraverso l'aggiunta di solventi organici o il congelamento immediato dei campioni tramite l'uso di ghiaccio secco o azoto liquido), le condizioni di conservazione (di solito si preferisce il mantenimento a lungo termine dei fluidi biologici a -80 °C), le piattaforme analitiche da utilizzare e le strategie di preparazione dei campioni. È importante sottolineare che gli studi di metabolomica sono intrinsecamente comparativi e, pertanto, comportano tipicamente la definizione di un gruppo di campioni di controllo (quelli che non sono stati sottoposti alla condizione investigata) e di campioni di test (che forniscono informazioni sulla condizione investigata) all'interno della progettazione sperimentale.

Preparazione del campione

Dopo aver definito il problema biologico e stabilito le condizioni sperimentali per la raccolta e la conservazione dei campioni, un passaggio successivo cruciale consiste nella preparazione dei campioni prima dell'analisi. La preparazione dei campioni è strettamente correlata al tipo di campione, all'approccio metabolomico scelto e alla piattaforma analitica impiegata. Nel caso della metabolomica *targeted*, la procedura di estrazione è solitamente ottimizzata per specifici metaboliti o classi chimiche di metaboliti in considerazione e può coinvolgere fasi di estrazione in fase solida per eliminare le interferenze della matrice del campione e/o per concentrare i metaboliti. Per quanto riguarda la metabolomica *untargeted* dei fluidi biologici, la preparazione del campione è generalmente mini-

male, coinvolgendo talvolta la precipitazione delle proteine come misura per preservare l'integrità del sistema cromatografico e massa-spettrometrico. In generale, spesso si effettua una semplice filtrazione seguita da una diluizione. Le preparazioni dei tessuti e/o delle cellule richiedono procedure di estrazione più complesse, generalmente effettuate attraverso estrazione su fase solida con solventi puri o miscele, seguite da centrifugazione e diluizione. Le analisi in GC dei fluidi biologici e degli estratti di cellule/tessuti richiedono ulteriori passaggi di derivatizzazione per convertire i metaboliti polari in addotti volatili. Questi passaggi, tuttavia, sono temporaneamente intensivi e soggetti a potenziali errori, il che limita il numero totale di campioni che possono essere elaborati in un singolo esperimento di metabolomica. Negli esperimenti di NMR, è di solito necessaria la diluizione del campione in opportuni solventi deuterati.

Acquisizione dei dati

A differenza di altre discipline "omiche", la metabolomica affronta sfide analitiche significative dovute all'ampia varietà di composti chimici presenti nei campioni biologici. Questi composti possono differire notevolmente nelle loro proprietà chimiche, caratteristiche strutturali, funzionalità e livelli di concentrazione. È importante notare che attualmente non esiste una singola piattaforma analitica in grado di identificare e quantificare l'intero insieme di metaboliti presenti in un sistema biologico. La diversità chimica del metaboloma, insieme alla sua vasta gamma dinamica, richiede l'impiego di diverse tecniche analitiche combinate al fine di ottenere risultati complementari che aumentino la copertura metabolica complessiva. Le tecniche analitiche comunemente utilizzate negli studi di metabolomica includono la spettroscopia NMR e la HRMS. La spettroscopia NMR può essere considerata una tecnica universale per la rilevazione dei metaboliti, poiché consente l'analisi diretta dei campioni con una manipolazione minima e la misura contemporanea di molte classi di piccoli metaboliti. Tuttavia, la spettroscopia NMR presenta svantaggi significativi in termini di sensibilità e complessità spettrale, con sovrapposizione di segnali in alcune regioni spettrali che possono compromettere l'identificazione chiara. La HRMS, sebbene più sen-



sibile e specifica rispetto alla spettroscopia NMR, richiede generalmente una ifenazione con tecniche di cromatografia o elettroforesi al fine di ridurre la complessità dei campioni e minimizzare gli effetti di soppressione della ionizzazione, aumentando così la sensibilità e la copertura del metaboloma.

Elaborazione dei dati - Nella metabolomica *untargeted*, i dati grezzi acquisiti vengono sottoposti a una fase di preprocessamento in base alla piattaforma analitica utilizzata. Nel caso della NMR, il trattamento dei dati include la correzione della linea di base, l'allineamento e la normalizzazione, utilizzando software e algoritmi come PERCH (PERCH Solution Ltd.), Chenomx NMR Suite (Chenomx Inc.), MestReNova (MestreLab Research), MetaboLab, AutoFit, TopSpin (Bruker Corp.) e MATLAB (The MathWorks Inc.). Per quanto riguarda le tecniche di HRMS, il trattamento dei dati comprende la deconvoluzione spettrale, la creazione di dataset, l'allineamento, la compilazione dei dati mancanti, la normalizzazione e la trasformazione. La normalizzazione dei vari campioni sarà basata sull'utilizzo di campioni di Quality Control (QC) che vengono ottenuti con piccole aliquote di tutti i campioni analizzati. Esistono diverse opzioni di software gratuiti e a pagamento per l'elaborazione dei dati MS, tra cui XCMS, Mass Profiler Professional (MPP, Agilent Technologies), MZmine, MetAlign, MassLynx (Waters Corp.) e AMDIS.

Analisi statistica - Le matrici di dati metabolomici raccolte, essendo estremamente complesse, richiedono l'uso di strumenti chemiometrici per l'identificazione di metaboliti che discriminino tra campioni di controllo e campioni di test. I metodi chemiometrici generalmente utilizzati sono basati su analisi *supervised* per evidenziare la presenza di variabili discriminanti nei gruppi di campioni a confronto. In particolare, mediante analisi discriminante con il metodo dei minimi quadrati (*partial least square - discriminant analysis*, PLS-DA) in *repeated double cross validation* (rDCV) sarà possibile determinare variabili in grado di differenziare i gruppi di campioni e allo stesso tempo minimizzare effetti di overfitting.

Identificazione dei metaboliti - L'identificazione dei metaboliti è necessaria solo negli studi di meta-

bolomica *untargeted*, poiché nella metabolomica *targeted* il metabolita o la classe di metaboliti di interesse è già definita. A tal fine, sono disponibili database e librerie gratuiti, come HMDB, KEGG, PubChem, Metlin, MassBank, LIPID MAPS e ChEBI ma anche l'utilizzo di dati di letteratura e regole di frammentazione di classi di composti strutturalmente correlati. I composti tentativamente identificati verranno confrontati con standard analitici certificati per conferma. Nella metabolomica *targeted*, la quantificazione o semiquantificazione degli analiti costituisce una parte significativa dell'elaborazione dei dati ed è solitamente effettuata tramite MS, piuttosto che NMR. Sono sviluppati metodi specifici, seguendo linee guida adeguate, per ottimizzare l'analisi.

Casi di studio - 1

Dagli inizi della metabolomica fino ad oggi, l'attenzione principale è stata incentrata sulla metabolomica delle piante. Tuttavia, con l'avvento recente della medicina personalizzata, la metabolomica clinica è diventata di primaria importanza grazie alla sua capacità di offrire una valutazione molecolare dei fluidi biologici, delle cellule e dei tessuti. In questo contesto, la metabolomica clinica sta guadagnando sempre più rilevanza nell'ambito della diagnosi delle malattie, della comprensione dei meccanismi patologici, dell'individuazione di nuovi obiettivi terapeutici, della personalizzazione dei trattamenti farmacologici e del monitoraggio degli esiti terapeutici. Poiché i metaboliti rappresentano i risultati finali dell'espressione genica e dell'attività cellulare, la metabolomica può offrire un approccio globale per comprendere il fenotipo di un organismo, svolgendo un ruolo cruciale nella biologia dei sistemi. La caratterizzazione dei fenotipi metabolici sostiene la medicina personalizzata identificando gli squilibri metabolici alla base delle malattie, individuando nuove possibilità di trattamento e indicando potenziali biomarcatori utilizzabili sia per la diagnosi che per il monitoraggio dell'efficacia delle terapie.

Il cancro alla prostata è il tumore più comunemente diagnosticato negli individui di sesso maschile, colpisce principalmente gli uomini di età superiore ai 50 anni ed è la principale causa di decessi correlati al cancro. Il metodo comunemente utilizzato

per la diagnosi del tumore alla prostata consiste nella misurazione del livello dell'antigene prostatico specifico nel sangue che, però, non permette di discriminare i casi di iperplasia prostatica benigna da quelli tumorali. Nello studio condotto nel nostro gruppo di ricerca: *"Comprehensive biomarker profiles and chemometric filtering of urinary metabolomics for effective discrimination of prostate carcinoma from benign hyperplasia"* [7], sono stati studiati 20 campioni di urina di pazienti con iperplasia benigna e 20 di pazienti con tumore prostatico per sviluppare una strategia metabolomica utile a distinguere la patologia maligna dall'iperplasia benigna. È stato effettuato un approccio *untargeted* UHPLC-HRMS per generare grandi matrici di dati (*big data*). Con l'ausilio di strumenti chemiometrici è stato costruito un modello PLS-DA rigorosamente convalidato; questo approccio ha, inoltre, filtrato i biomarcatori più significativi. Sono stati studiati per la prima volta un numero molto elevato di potenziali biomarcatori, 22 provenienti dalla ionizzazione massa-spettrometrica in positivo (ESI+) e 47

in negativo (ESI-). Il modello PLS-DA basato sui 22 biomarcatori in positivo ha fornito una sensibilità del $95\pm 1\%$ e una specificità dell' $83\pm 3\%$, mentre quello dei 47 biomarcatori in negativo ha prodotto una sensibilità dell' $88\pm 3\%$ e una specificità del $91\pm 2\%$. Questo sicuramente rappresenta un nuovo studio pionieristico per la scoperta di nuovi biomarcatori da testare su set più ampi di campioni per poi implementare nella pratica diagnostica di routine (Fig. 3).

Casi di studio - 2

L'origine del prodotto è sinonimo di garanzia, sicurezza e qualità. Di conseguenza, negli ultimi anni sono emerse certificazioni di prodotto di vario tipo per informare i consumatori sulle caratteristiche uniche del prodotto e per proteggere i produttori dalle frodi; le certificazioni di origine variano da Paese a Paese.

La metabolomica può essere utilizzata come sistema di impronte digitali chimiche applicabile alle identificazioni geografiche, il che potrebbe migliorare i processi di registrazione e la protezione contro le frodi. L'approccio metabolomico viene utilizzato negli studi su larga scala dei metaboliti e, quando applicato alle piante, comporta analisi ad alto rendimento di miscele di metaboliti complessi. Il metaboloma vegetale comprende metaboliti primari e secondari. Quelli primari (amminoacidi, peptidi, carboidrati, acidi grassi ecc.) sono importanti per la crescita e lo sviluppo delle piante; quelli secondari (flavonoidi, alcaloidi, tannini, terpeni ecc.) sono legati all'attrazione degli impollinatori, alla protezione contro parassiti, agenti patogeni ed erbivori e sono anche coinvolti nella percezione del colore, dell'aroma e del sapore della pianta. Molti studi metabolomici hanno come oggetto di studio l'obiettivo di verificare e confrontare le differenze metaboliche tra campioni differenti derivanti da diverse aree geografiche.

Nello studio condotto dal nostro gruppo di ricerca: *"Untargeted cannabinomics reveals the chemical differentiation of industrial hemp based on the cultivar and the geographical field location"* [8] si è sfruttato un approccio di metabolomica *untargeted-suspect screening*, su classi di composti selezionati (cannabinoidi e composti fenolici) per valutare sette distinte *cultivar* di canapa industria-

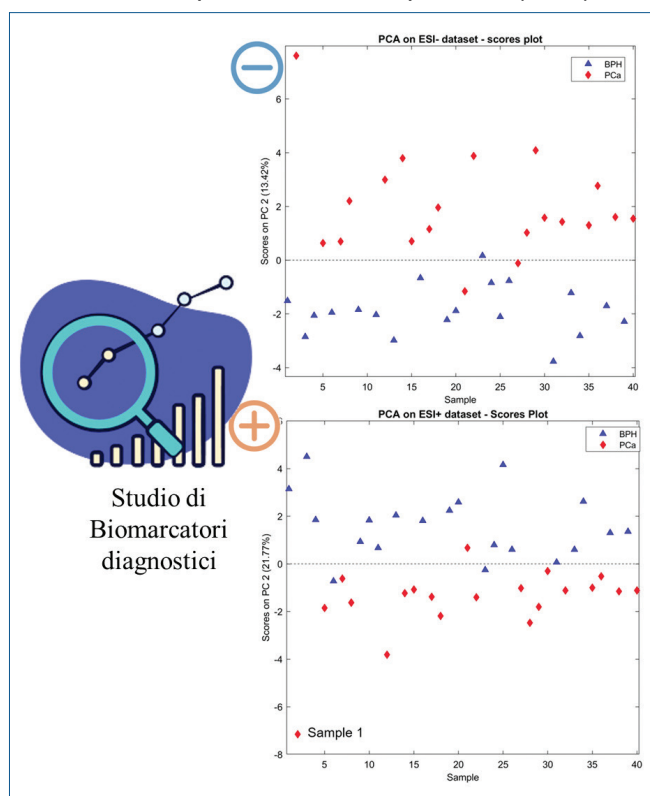


Fig. 3 - Profilo completo dei biomarcatori per una discriminazione efficace del carcinoma prostatico dall'iperplasia benigna

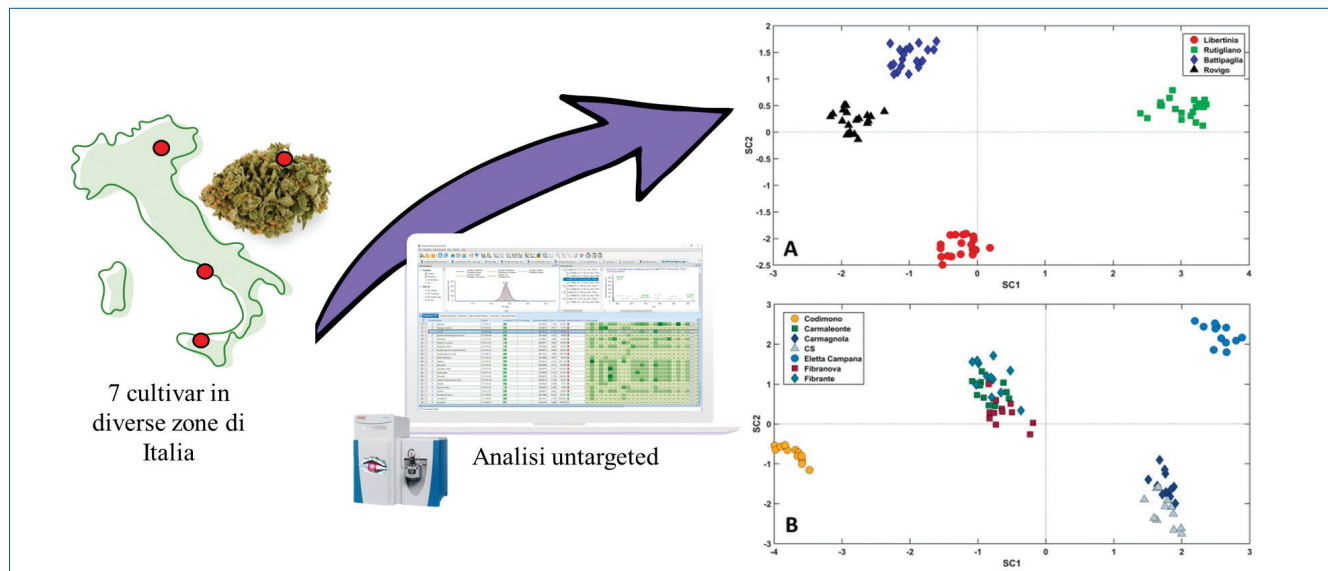


Fig. 4 - Schema riassuntivo di analisi, differenziazione mediante analisi *untargeted* di 7 varietà di canapa coltivate in diverse zone di Italia

le coltivate in quattro campi sperimentali nell'Italia settentrionale, meridionale e insulare.

La *Cannabis sativa* è stata a lungo raccolta per applicazioni industriali legate alle sue fibre, poiché conteneva una bassa concentrazione di sostanza psicoattiva; tale matrice è stata poco studiata in termini di contenuto in composti bioattivi. La metabolomica *untargeted* va oltre la caratterizzazione dei fitocomposti più studiati. Infatti caratterizzare in maniera completa tutti i fitocannabinoidi, i flavonoidi e gli acidi fenolici contenuti in queste *cultivar* è stato fondamentale per una buona differenziazione. Sono stati identificati un totale di 54 fitocannabinoidi, 134 flavonoidi e 77 acidi fenolici che hanno avuto un ruolo chiave nel distinguere i campioni di canapa in base alla posizione geografica del campo e alla *cultivar* (Fig. 4).

BIBLIOGRAFIA

- [1] M. Mapstone *et al.*, *Nature Medicine*, 2014, **20**(4), 415.
- [2] T.J. Wang *et al.*, *Nature Medicine*, 2011, **17**(4), 448.
- [3] A. Sreekumar *et al.*, *Nature*, 2009, **457**(7231), 910.
- [4] J.B. Rajniak, N.K. Barco *et al.*, *Nature*, 2015, **525**(7569), 376.
- [5] S.J. Mentch *et al.*, *Cell Metabolism*, 2015, **22**(5), 861.

- [6] R. Wehrens, R. Salek, *Advances in Experimental Medicine and Biology* ISBN 978-3-319-47655-1.
- [7] E. Amante *et al.*, *Scientific Reports*, 2022, **12**, 4361.
- [8] A. Cerrato *et al.*, *Analytica Chimica Acta*, 2023, **1278**, 341716.

Metabolomics and Big Data: from Comprehensive Organism Characterization to Diagnostic Biomarker Identification

Metabolomics represents the most recent discipline within the “omics” sciences and has experienced significant growth in recent years, owing to measurement and analytical techniques advancements. This term pertains to a field of study that enables the comprehensive characterization of small molecules, known as metabolites, produced by an organism at a given moment. In recent years, metabolomics has made substantial strides in diagnostics and research, thanks to the development of chemometric tools that allow for discriminative analysis of metabolomic data, thus facilitating differentiation among examined groups. Metabolomics has found extensive applications across various research sectors, including diagnostics, plant biology, toxicology, microbiology, the agro-food industry, and the pharmaceutical and medical sectors.