



UNIVERSITÀ DEGLI STUDI DEL SANNIO

DIPARTIMENTO DI SCIENZE E TECNOLOGIE

CORSO DI LAUREA TRIENNALE IN SCIENZE BIOLOGICHE

TESI DI LAUREA

IN

FONDAMENTI DI INFORMATICA

**Meta-analisi degli studi di trascrittomici in
endometriosi:
prospettive per target terapeutici**

Relatore:

Ch.mo Prof.

Francesco Napolitano

Candidata:

Chiara Circelli Gabriele

Matricola 172006049

Anno Accademico 2024–2025

Grata memoria praeteritorum,

fiducia praesentium,

spes futurorum.

Indice

Introduzione	7
1 Background e contesto clinico dell'endometriosi	11
1.1 Introduzione al capitolo	11
1.2 Che cos'è l'endometriosi	12
1.3 Origini e principali teorie eziogenetiche	12
1.4 Sintomatologia e problematiche diagnostiche	14
1.5 Trattamenti e approcci terapeutici	19
1.6 Prospettive di analisi molecolare nell'endometriosi	20
2 La trascrittomica	21
2.1 Cos'è la trascrittomica	21
2.2 Tecniche di studio	22
2.3 Applicazioni della trascrittomica nello studio dell'endometriosi	26
3 Materiali, metodi e approccio sperimentale	29
3.1 Revisione degli studi trascrittomici sull'endometriosi	29
3.2 Scelta dello studio e del dataset per l'analisi	35
3.3 Analisi bioinformatica dei dati trascrittomici con R	39
4 Evidenze trascrittomiche e implicazioni cliniche	43

4.1	Risultati dell'analisi bioinformatica	43
4.2	Riflessioni conclusive	55
4.3	Prospettive di ricerca e applicazioni terapeutiche	58
5	Codice sorgente	61
	Bibliografia	65

Introduzione

La salute ginecologica femminile rappresenta un ambito centrale della medicina, non soltanto per l'impatto clinico delle patologie che vi rientrano, ma anche per le profonde conseguenze che esercitano sulla vita delle donne. Le patologie ginecologiche croniche, infatti, influenzano severamente anche la qualità di vita delle pazienti, incidendo sulla fertilità, sul benessere relazionale e sul pieno inserimento delle donne nella vita sociale e lavorativa.

In questo scenario, l'endometriosi, che sarà approfondita nel Capitolo 1, appare tra le patologie ginecologiche più diffuse e difficili da comprendere pienamente. Colpisce le donne in età fertile e si manifesta con quadri clinici molto variabili, il che rende difficile giungere a una diagnosi tempestiva: spesso occorrono anni di sintomi prima di avere una conferma certa. Questo ritardo non rappresenta soltanto un problema clinico, ma alimenta frustrazione, disorientamento e un senso di solitudine nelle pazienti, che vedono compromessa la qualità della loro vita personale, lavorativa e relazionale.

I trattamenti oggi disponibili possono offrire sollievo, ma raramente sono risolutivi e non sempre riescono a rispondere alle diverse esigenze delle donne che convivono con la malattia. Negli ultimi anni, la ricerca ha dedicato crescente attenzione a questa patologia, riconoscendone la natura multifattoriale e sottolineando quanto i suoi meccanismi biologici siano ancora in parte oscuri. Da qui nasce la necessità di strumenti capaci di indagare più a fondo i processi che ne regolano l'origine e lo sviluppo, con l'obiettivo di identificare marcatori diagnostici precoci e nuovi target terapeutici che possano aprire prospettive di cura più efficaci.

Sotto questo aspetto, le tecnologie omiche e, in particolare, la trascrittomica, a cui è dedicato il Capitolo 2, si sono dimostrate strumenti di grande potenzialità. Studiare come i geni si esprimono nelle pazienti permette di comprendere quali processi biologici cambiano, offrendo una visione più completa rispetto all'osservazione clinica tradizionale e aiutando a comprendere meglio alcune delle difficoltà quotidiane che le donne affrontano convivendo con la malattia. Tuttavia, i dati disponibili, provenienti da studi diversi e indipendenti, risultano spesso frammentari e difficili da confrontare. Per questa ragione è importante intervenire con approcci integrativi, come le meta-analisi, che uniscono le informazioni esistenti così da ottenere un quadro più coerente e completo delle evidenze scientifiche.

Il presente lavoro di tesi nasce da questa esigenza di comprendere meglio una malattia così complessa e impattante sulla vita delle donne. Attraverso una meta-analisi degli studi trascrittomici sull'endometriosi, lo studio si propone di individuare pattern comuni di espressione genica e pathways biologici potenzialmente coinvolti nella patogenesi della malattia, cercando di collegare questi risultati a possibili applicazioni terapeutiche. L'impiego di metodi bioinformatici consente di sfruttare al meglio le informazioni già disponibili, restituendo un quadro d'insieme più coerente e approfondito dei meccanismi biologici che caratterizzano l'endometriosi.

La struttura del lavoro riflette questo percorso di analisi. Dopo una panoramica iniziale dedicata al contesto clinico dell'endometriosi, con riferimento alle principali teorie eziogenetiche, alla sintomatologia, alle difficoltà diagnostiche e ai trattamenti in uso, viene introdotto, nel Capitolo 2, il ruolo della trascrittomica come strumento di ricerca molecolare. Si procede, quindi, nel Capitolo 3, alla descrizione dei materiali e dei metodi utilizzati per la meta-analisi, includendo i criteri di selezione degli studi e le procedure bioinformatiche adottate. Segue, nel Capitolo 4, la presentazione dei risultati ottenuti e la loro discussione critica.

Infine, le riflessioni conclusive sintetizzano i principali contributi dello studio, ribadendo la

complessità dell'endometriosi e i diversi aspetti biologici e molecolari che ne caratterizzano l'eziologia e la manifestazione clinica, e ne evidenziano le prospettive di ricerca futura, in particolare per l'individuazione di nuovi target terapeutici. A completamento del lavoro, è riportato anche il codice RStudio utilizzato per le analisi, così da garantirne la riproducibilità.

Capitolo 1

Background e contesto clinico dell'endometriosi

1.1 Introduzione al capitolo

L'endometriosi rappresenta una delle patologie ginecologiche croniche più complesse e diffuse, con un impatto significativo sulla vita delle donne in età fertile. La malattia si caratterizza per quadri clinici eterogenei, che vanno dal dolore pelvico cronico alla compromissione della fertilità, rendendo spesso difficile una diagnosi tempestiva e accurata.

Nonostante i progressi nella comprensione dei meccanismi biologici, permangono molte incognite relative all'eziologia e alla patogenesi della malattia. La complessità dell'endometriosi è legata alla presenza di diversi fattori biologici, come quelli genetici, ormonali e immunitari, che influenzano sia l'ampiezza dei sintomi sia le strategie di cura.

In questo capitolo vengono analizzati gli aspetti clinici fondamentali dell'endometriosi, tra cui la prevalenza, la sintomatologia, le difficoltà diagnostiche e le principali teorie eziologiche. L'obiettivo è fornire una panoramica chiara e dettagliata della malattia, creando il contesto necessario per comprendere le successive analisi trascrittomiche presentate nei

capitoli successivi[1].

1.2 Che cos'è l'endometriosi

L'endometriosi è una condizione infiammatoria cronica estrogeno-dipendente, caratterizzata dalla presenza di tessuto simil-endometriale in sedi atipiche, ossia al di fuori della cavità uterina[2].

Le aree più interessate riguardano le strutture pelviche: ovaie, tube di Falloppio, peritoneo e legamenti uterosacrali, ma in percentuali minori possono essere coinvolti anche organi extrapelvici, quali diaframma, intestino, parete addominale e pleura [3].

Si stima che questa patologia colpisca tra il 6% e il 10% delle donne in età riproduttiva a livello globale [4]. Tuttavia, i numeri reali potrebbero essere sottostimati a causa della difficoltà di una diagnosi definitiva, di una disattenzione nei confronti dei sintomi riferiti dalle pazienti o di casi asintomatici.

Il percorso diagnostico, spesso lungo e complesso, può richiedere in media dai 4 agli 11 anni dalla comparsa dei primi sintomi, aumentando la sofferenza delle pazienti e compromettendo l'efficacia di interventi terapeutici tempestivi [5].

1.3 Origini e principali teorie eziogenetiche

La ricerca sulle origini dell'endometriosi rappresenta tuttora una delle principali sfide nel campo della ginecologia. Diverse ipotesi sono state proposte, ma nessuna spiega in modo complessivo il profilo della malattia. In particolare si identificano cinque teorie principali riguardo la patogenesi dell'endometriosi:

- Teoria delle mestruazioni retrograde;

- Teoria della metaplasia celomica;
- Teoria delle anomalie dei resti embrionali di Müller;
- Teoria delle metastasi linfatiche e vascolari;
- Teoria dell’impianto delle cellule staminali endometriali;

La teoria più accettata sull’origine dell’endometriosi è quella della mestruazione retro-grada, secondo cui parte del flusso mestruale refluisce nella cavità peritoneale passando attraverso le tube di Falloppio, consentendo a cellule endometriali di impiantarsi nella cavità pelvica e dando vita alle prime lesioni endometriosiche. Tuttavia, questo processo da solo non sembra sufficiente a spiegare l’insorgenza della patologia, suggerendo che altri fattori, come la predisposizione genetica o disfunzioni immunitarie, possano contribuire allo sviluppo della malattia.[6].

La teoria della metaplasia celomica, invece, proposta da Robert Meyer, suggerisce che le cellule staminali mesenchimali multipotenti possono andare incontro a un processo di riprogrammazione, differenziandosi in cellule epiteliali endometriali e stromali nei siti ectopici. [7]. Essa è particolarmente utile a spiegare i casi di endometriosi al di fuori della cavità pelvica come polmoni o reni.

Allo stesso modo, la teoria delle metastasi linfatiche e vascolari afferma siano i vasi linfatici o sanguigni a trasportare cellule endometriali e frammenti di tessuto, contribuendo alla formazione di lesioni in sedi extrapelviche [8].

La teoria delle anomalie dei resti embrionali di Müller postula che residui embrionali presenti nel corpo possano svilupparsi in tessuto endometriale ectopico [9].

Infine, secondo la teoria dell’impianto delle cellule staminali endometriali, queste ultime sarebbero in grado di migrare verso sedi ectopiche e, attraverso la loro attività biologica, partecipare a processi come l’infiammazione, la neoangiogenesi, l’adesione e l’invasività

cellulare, contribuendo così alla formazione delle lesioni endometriosiche [9].

Sebbene queste teorie siano differenti tra loro, non si escludono a vicenda e, probabilmente, descrivono meccanismi che cooperano tra loro nella complessa origine della malattia.

Dal punto di vista ambientale, uno stile di vita caratterizzato da scarsa attività fisica, consumo di alcol, fumo, assunzione eccessiva di caffeina e un'alimentazione non equilibrata può influenzare in modo significativo il decorso dell'endometriosi [10]. Il mantenimento di un corretto equilibrio ormonale e la riduzione dell'infiammazione cronica risultano infatti fondamentali per il controllo e la gestione della malattia.

Alla luce di quanto esposto, l'endometriosi si configura come una condizione estremamente complessa, in cui convergono molteplici fattori e processi fisiopatologici che agiscono in modo simultaneo e interdependente.

1.4 Sintomatologia e problematiche diagnostiche

Il quadro sintomatologico dell'endometriosi varia sensibilmente da paziente a paziente, in relazione allo stadio della malattia, alle predisposizioni individuali, allo stile di vita e all'età. Generalmente i primi sintomi si manifestano prima dei 20 anni [11], ma spesso sono sottovalutati o interpretati erroneamente come sintomi mestruali comuni. I disturbi più frequenti sono dolore pelvico cronico, dismenorrea, dispareunia(dolore durante i rapporti sessuali), gonfiore addominale, stitichezza, disuria (minzione dolorosa) e defecazione dolorosa. Tali sintomi possono compromettere significativamente la qualità della vita, incidendo sulle attività quotidiane, sulle relazioni personali e sulla produttività lavorativa [12]. Accanto ai sintomi fisici, si osservano spesso disturbi di natura psico-emotiva, come ansia e depressione, legati alla cronicità del dolore e al conseguente impatto negativo sul benessere psicologico e sociale della paziente. In una percentuale non trascurabile di casi, l'endometriosi può de-

correre in modo asintomatico, con la presenza di lesioni endometriosiche che si sviluppano senza sintomi evidenti. L'endometriosi in forma silente rappresenta una vera sfida per la diagnosi, perché spesso viene riconosciuta solo in stadio avanzato e spesso si accompagna a complicazioni significative come l'infertilità [13].

La fertilità può essere compromessa da diversi meccanismi quali: l'alterazione dell'anatomia pelvica, la formazione di aderenze e fibrosi a livello delle tube di Falloppio, l'infiammazione cronica delle strutture pelviche, una disfunzione del sistema immunitario sia locale (a livello endometriale) sia sistemico, nonché modifiche dell'ambiente ormonale intrauterino e una ridotta recettività endometriale, che può ostacolare l'impianto embrionale.

Le lesioni endometriosiche, illustrate Figura 1.1, [14], sono classificate in 3 tipologie [15]:

- peritoneale superficiale (SUP);
- endometrioma ovarico (OMA);
- endometriosi infiltrante profonda (DIE).

Esse si contraddistinguono per una crescita anomala di cellule endometriali che invadono il peritoneo e generano stress ossidativo. Questi processi sono accompagnati da risposte infiammatorie e neuroangiogenesi. Inoltre, è particolarmente importante la localizzazione della malattia, in quanto contribuisce sia al dolore cronico, sia all'infertilità.

In tale contesto, classificare l'endometriosi è essenziale per una corretta valutazione clinica e terapeutica.

Il sistema di stadiazione attualmente più utilizzato è quello della American Society for Reproductive Medicine (rASRM), che suddivide la malattia in quattro stadi progressivi, da I (minimo) a IV(severo). La classificazione si basa su sede, l'estensione e la profondità delle lesioni, nonché sulla presenza di aderenze fibrose (Figura 1.2) [16]. È importante precisare che la stadiazione non è correlata alla sintomatologia; per questo si considerano i diversi fenotipi

clinici della malattia[17][18]. Individuare l'endometriosi in modo preciso e tempestivo è essenziale per affrontare al meglio i sintomi che provoca, incluso il problema dell'infertilità.[19] Ad oggi, la modalità diagnostica più accurata rimane la laparoscopia associata alla conferma istologica di lesioni sospette. Questa procedura però è costosa e invasiva, per cui metodi di imaging e biomarcatori possono vedere determinati vantaggi in alcuni contesti clinici[20]. L'ecografia transvaginale (TV-USG) e la risonanza magnetica (MRI) risultano efficaci nella diagnosi di endometriosi infiltrante profonda e di endometriomi[6]. Tuttavia, mostrano dei limiti nella rilevazione delle lesioni peritoneali superficiali, specialmente quando queste sono di dimensioni ridotte e pertanto non visualizzabili con certezza.[6]

Parallelamente, cresce l'interesse verso l'identificazione di biomarcatori specifici, che potrebbero migliorare ulteriormente la diagnosi clinica. I biomarcatori permetterebbero non solo di individuare la malattia in un modo ancora meno invasivo, ma anche di monitorarne l'attività biologica e l'efficacia dei trattamenti terapeutici. Un biomarcatore spesso prescritto è il CA-125. Sebbene spesso elevato nelle pazienti con endometriosi, non è specifico per questa patologia, poiché può risultare aumentato in diverse condizioni cliniche; pertanto, non può essere utilizzato come unico test diagnostico [20].

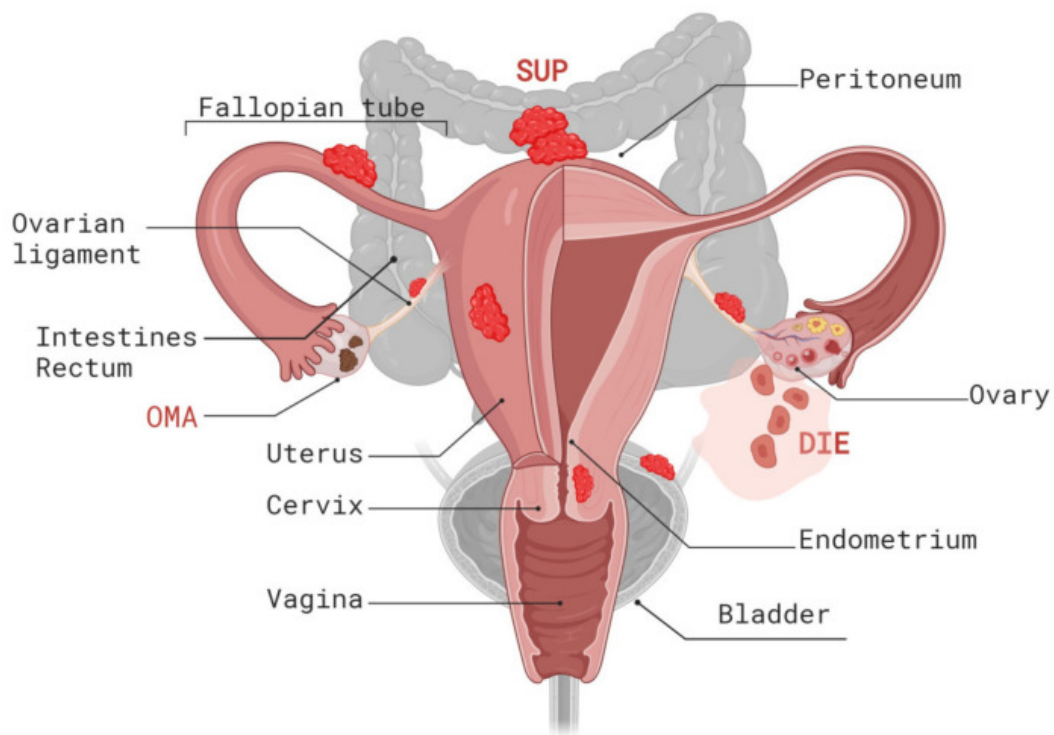


Figura 1.1: Rappresentazione schematica delle principali forme di endometriosi. La malattia può manifestarsi con lesioni diverse per aspetto e profondità: le **SUP** compaiono come piccoli impianti superficiali sul peritoneo, le **DIE** si caratterizzano per la crescita infiltrante nei tessuti profondi, mentre le **OMA** corrispondono a cisti endometriosiche localizzate a livello ovarico. Figura generata con BioRender.com (consultato il 18 giugno 2024). Fonte: Adilbayeva A., Kunz J. *Pathogenesis of Endometriosis and Endometriosis-Associated Cancers*, PMCID: PMC11277188, PMID: 39062866.

	diameters			• Motility • Patency test			• Largest diameter			• Largest diameter			• Largest diameter		
	left	right		left	right		left	right		left	right		left	right	
cm	O1	$\Sigma < 3$ cm	T1	Pelvic sidewall		A1	< 1 cm		B1	< 1 cm		C1	< 1 cm		
cm	O2	$\Sigma 3-7$ cm	T2	Pelvic sidewall, uterus		A2	1 - 3 cm		B2	1 - 3 cm		C2	1 - 3 cm		
cm	O3	$\Sigma > 7$ cm	T3	Pelvic sidewall, uterus, bowel, USL		A3	> 3 cm		B3	> 3 cm		C3	> 3 cm		
	O /			T _ / _									F _		

Figura 1.2: Sistema di classificazione ASRM (American Society for Reproductive Medicine) per l'endometriosi. La tabella illustra i criteri utilizzati per attribuire un punteggio in base a sede e dimensioni delle lesioni: **P** valuta le lesioni peritoneali in base alla somma dei diametri; **O** considera le lesioni ovariche e la loro estensione bilaterale; **T** riguarda la condizione tubo-ovarica (aderenze, motilità e pervietà tubarica); **A**, **B** e **C** classificano rispettivamente le lesioni in vagina/spazio retto-vaginale, nei legamenti sacrouterini/cardinali e nel retto, in funzione del diametro; **F** permette di segnalare la presenza di ulteriori sedi coinvolte (ad esempio adenomiosi, vescica, intestino, uretere, diaframma, polmone o nervi). Questo sistema consente di tradurre le caratteristiche delle lesioni in un punteggio complessivo utile per stadiare la malattia e confrontare la gravità tra pazienti. Figura tratta da: Pašalić E., Tambuwala M.M., Hromić-Jahjefendić A. *Endometriosis: Classification, pathophysiology, and treatment options*, PMID: 37844487, DOI: 10.1016/j.prp.2023.154847.

1.5 Trattamenti e approcci terapeutici

Il piano terapeutico per il trattamento dell'endometriosi viene definito in base alla paziente in questione, considerando l'intensità dei sintomi, l'età della donna, la distribuzione ed entità delle lesioni endometriosiche. Può avvenire attraverso una terapia farmacologica per alleviare i dolori o ridurre la possibilità di recidive, in altri casi, è opportuno intervenire chirurgicamente. [21].

In primo luogo c'è la scelta di trattamenti ormonali che riequilibrano le oscillazioni di ormoni gonadotropici e ovarici, portando ad amenorrea, inibizione dell'ovulazione e diminuzione dell'infiammazione[12]. La maggior parte delle terapie ormonali per l'endometriosi mira a inibire la produzione di estrogeni a livello ovarico, trascurando invece il blocco degli estrogeni prodotti localmente nelle lesioni endometriosiche. Inoltre occorre considerare che in media il 25% delle pazienti non risponde alla terapia e che sono molto frequenti effetti avversi quali spotting, mastodinia (dolorabilità al seno), cefalea, instabilità dell'umore[13][22].

Nei casi in cui i sintomi non si alleviano e gli effetti collaterali risultano superiori ai benefici clinici ottenuti con la terapia si ricorre al trattamento chirurgico.

L'intervento chirurgico può essere preso in considerazione per le pazienti che presentano alterazioni anatomiche delle strutture pelviche, aderenze, o fenomeni ostruttivi che interessano l'intestino o le vie urinarie [23].

La chirurgia conservativa implica la cauterizzazione delle lesioni endometriosiche e il ripristino dell'assetto anatomico della pelvi. Questa procedura determina un miglioramento significativo del dolore pelvico e della fertilità; tuttavia è possibile la comparsa di recidive nel periodo post-chirurgico.[6].

1.6 Prospettive di analisi molecolare nell'endometriosi

Dall'analisi fin qui condotta emerge con chiarezza come l'endometriosi sia una patologia particolarmente complessa, tanto sul piano clinico quanto su quello biologico[13, 12]. Nonostante i meccanismi eziopatogenetici non siano ancora del tutto compresi, le moderne tecnologie -omiche, in particolare la trascrittomica, stanno offrendo strumenti preziosi per fare luce su questa complessa patologia. In questo contesto è essenziale analizzare i profili di espressione genica nei tessuti eutopici ed ectopici, così da individuare pattern trascrizionali specifici associati a processi chiave come l'infiammazione cronica o le alterazioni ormonali. Queste indagini non solo contribuiscono ad approfondire la fisiopatologia della malattia, ma aprono nuove prospettive in ambito diagnostico e terapeutico, rendendo possibile l'identificazione di biomarcatori molecolari e potenziali bersagli farmacologici personalizzati [24, 25].

A sostegno dell'importanza di un approccio trascrittomico, uno studio significativo è quello condotto da Kao et al. (2003). Gli autori hanno confrontato i profili di espressione genica tra tessuti eutopici ed ectopici prelevati da donne affette dalla patologia, individuando una marcata disregolazione di geni coinvolti in processi infiammatori, più rilevante rispetto alla presenza di mutazioni genetiche [26].

Questo rafforza l'idea secondo cui l'analisi funzionale a livello molecolare non sia solo utile, ma potenzialmente decisiva per comprendere e affrontare in modo più mirato la malattia.

Capitolo 2

La trascrittomica

2.1 Cos'è la trascrittomica

La trascrittomica implica una serie di strumenti e metodologie impiegate per analizzare l'insieme completo delle trascrizioni di RNA di un organismo, noto come trascrittoma. Sebbene il patrimonio genetico sia custodito nel DNA, è attraverso la trascrizione che queste informazioni assumono un ruolo funzionale.

In particolare l'mRNA agisce da messaggero trasportando le istruzioni genetiche e traducendole in segnali utili alla cellula; altri tipi di RNA hanno ruoli differenti, spesso regolatori.

Il trascrittoma rappresenta, dunque, una fotografia istantanea dell'attività genetica all'interno di una cellula, evidenziando quali geni risultano espressi in quel momento e in quale misura [27].

Comprendere il trascrittoma è quindi fondamentale per monitorare come varia l'espressione genica tra diversi organismi e delineare i profili molecolari associati a varie patologie.

2.2 Tecniche di studio

Il termine trascrittoma è stato introdotto per la prima volta negli anni '90, in concomitanza con i primi tentativi sistematici di mappare l'insieme completo delle trascrizioni geniche [28, 29].

Una delle tecniche maggiormente impiegate era il metodo SAGE (Serial Analysis of Gene Expression), basato sul sequenziamento Sanger di brevi frammenti di cDNA ottenuti da RNA messaggero e concatenati tra loro in modo casuale.

Con l'avanzare delle tecnologie, i metodi iniziali sono stati gradualmente sostituiti da approcci più moderni, come i microarray e il sequenziamento RNA-Seq, che attualmente rappresentano strumenti fondamentali per l'analisi dell'espressione genica.

I microarray consentono di quantificare in simultanea l'espressione di migliaia di geni tramite l'ibridazione di trascritti a sonde oligonucleotidiche specifiche. Questa tecnica ha consentito una significativa riduzione del costo per gene e un notevole risparmio in termini di manodopera [30, 31]. Inoltre, miglioramenti nella progettazione delle sonde hanno aumentato la specificità, mentre l'ottimizzazione del rilevamento della fluorescenza ha migliorato la sensibilità, permettendo la rilevazione anche di trascritti poco abbondanti. [31, 32]. Lo schema illustrativo di un esperimento di microarray è riportato in Figura 2.1, mostrando i principali passaggi dall'estrazione dell'mRNA fino all'analisi dei dati.

L'RNA-Seq (RNA sequencing) è una tecnica che permette di quantificare l'espressione genica attraverso il sequenziamento dei cDNA ottenuti dai trascritti presenti in un campione. L'abbondanza di ogni trascritto viene determinata contando il numero di sequenze (read) che vi corrispondono [33, 34]. Tra i primi approcci di trascrittomica basati sul sequenziamento vi è il Massively Parallel Signature Sequencing (MPSS), una tecnica che generava brevi sequenze di 16–20 basi, tramite una serie articolata di cicli di ibridazione e rilevamento [35].

Questa tecnica fu impiegata nel 2004 per validare l'espressione di 104 geni in *Arabidopsis thaliana* [36].

Successivamente, lo sviluppo di piattaforme di sequenziamento Solexa/Illumina (San Diego, CA) ha reso possibile la generazione di oltre 10^9 sequenze trascrizionali per esperimento [37, 38]; rendendo così l'RNA-Seq uno strumento sufficientemente sensibile e preciso per la quantificazione dell'intero trascrittoma umano [27].

Passando ad un'analisi comparativa, è importante evidenziare che i microarray richiedono sonde complementari a sequenze già note, limitando l'analisi ai trascritti precedentemente identificati. Diversamente, l'RNA-Seq analizza in modo completo l'RNA di un campione, senza richiedere dati preliminari sulle sequenze da identificare. Questo approccio rende possibile la scoperta di nuovi trascritti, inclusi RNA non codificanti, isoforme alternative e mutazioni, ampliando significativamente le potenzialità dell'analisi trascrittomica. Grazie a una sensibilità elevata e a un'ampia capacità di rilevamento su diversi livelli di espressione, l'RNA-Seq si presta anche all'analisi di geni espressi in modo debole e al confronto accurato tra campioni sperimentali [27]. La Figura 2.2 illustra schematicamente un tipico esperimento RNA-Seq.

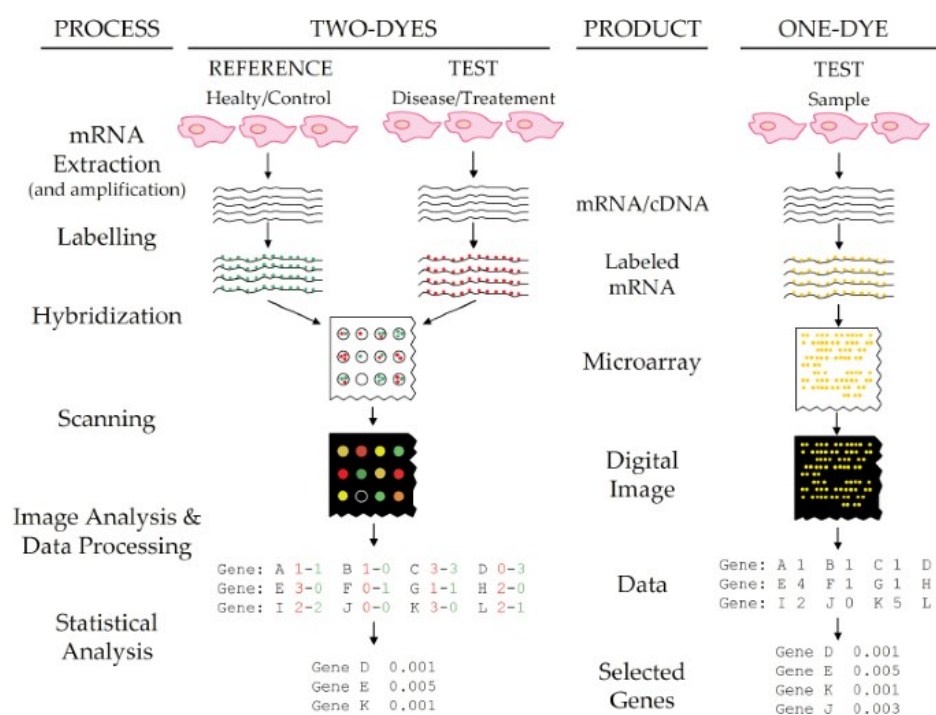


Figura 2.1: Schema di un esperimento di microarray per l'analisi dell'espressione genica. L'immagine mostra i principali passaggi: 1. Estrazione dell'mRNA dai campioni biologici; 2. Sintesi di cDNA complementare marcato con fluorocromi specifici; 3. Ibridazione del cDNA alle sonde presenti sull'array di DNA; 4. Rilevazione del segnale fluorescente tramite scanner dedicato; 5. Analisi dei dati per identificare geni sovra- o sotto-espressi e generare profili globali del trascrittoma. Questo schema evidenzia come i microarray consentano di confrontare l'espressione genica tra diversi campioni e di ottenere una panoramica completa dell'attività trascrizionale.

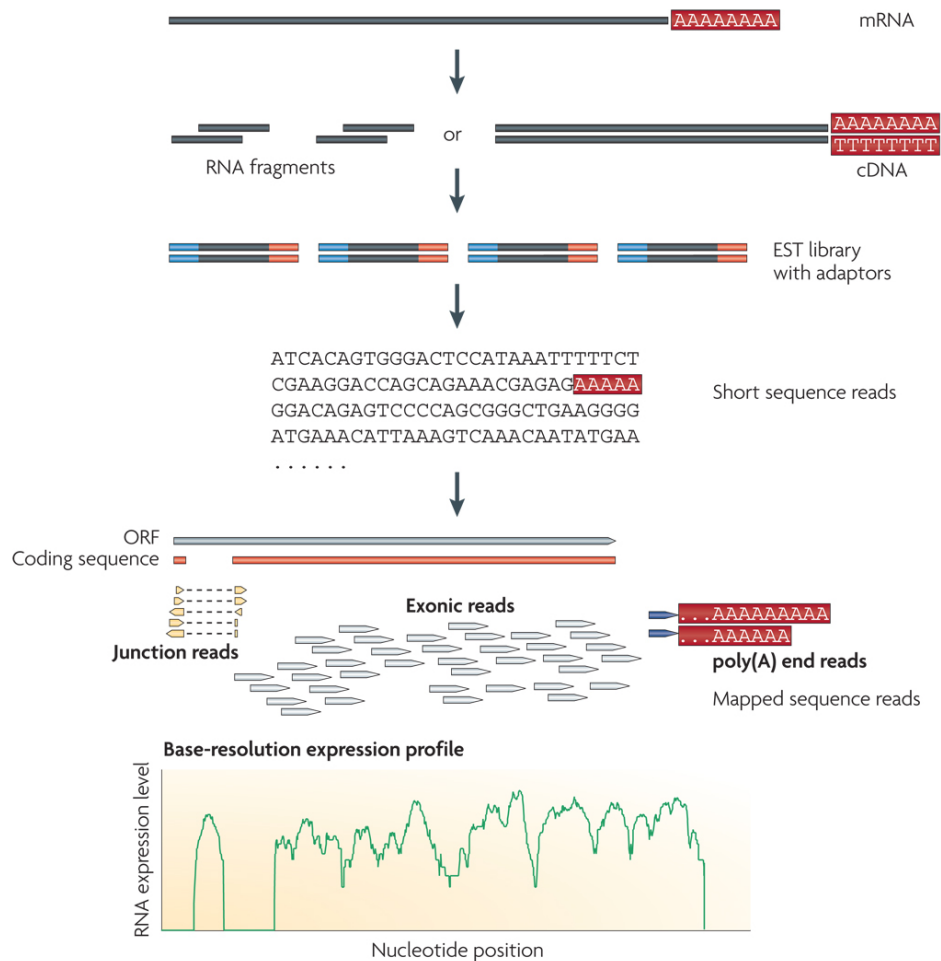


Figura 2.2: Schema di un esperimento RNA-Seq per l'analisi dell'espressione genica. L'immagine illustra i principali passaggi della procedura: 1. Estrazione dell'RNA totale dai campioni biologici; 2. Sintesi di cDNA complementare a partire dall'RNA estratto; 3. Aggiunta di adattatori specifici per il sequenziamento e creazione della libreria di cDNA; 4. Sequenziamento ad alta capacità (high-throughput) dei cDNA tramite piattaforme NGS; 5. Allineamento delle sequenze generate al genoma di riferimento e quantificazione dei trascritti. La figura mostra chiaramente come l'RNA-Seq consenta di ottenere una panoramica completa dell'espressione genica, identificando sia trascritti noti sia nuovi trascritti, inclusi RNA non codificanti e isoforme alternative. Questo schema evidenzia inoltre l'elevata sensibilità e la capacità di rilevare geni espressi a livelli bassi, rendendo l'approccio adatto a studi comparativi tra campioni diversi.

2.3 Applicazioni della trascrittomica nello studio dell'endometriosi

La trascrittomica rappresenta uno strumento fondamentale per lo studio dell'endometriosi, che, come già visto, è una patologia complessa, multifattoriale e con forte eterogeneità tissutale e cellulare. Analizzare l'espressione genica nei tessuti eutopici (endometrio nella cavità uterina) ed ectopici (endometrio in sedi extrauterine) consente di identificare geni e vie molecolari coinvolti nella patogenesi, nonché potenziali biomarcatori diagnostici e bersagli terapeutici.

Sono stati selezionati alcuni articoli disponibili su PubMed che affrontano l'analisi del profilo trascrizionale dell'endometriosi, utilizzando tecnologie diverse come il microarray, l'RNA-seq e il single-cell RNA sequencing. Quest'ultima tecnica permette un'analisi dell'espressione genica a livello di singola cellula, offrendo una visione dettagliata dell'eterogeneità cellulare del tessuto endometriotico; tuttavia, essendoci dei limiti nel numero di studi disponibili, questa tecnica può rappresentare un ostacolo per alcune analisi comparate. Un esempio di studio che utilizza il single-cell RNA-seq è rappresentato dal dataset GSE214411, impiegato per esaminare in dettaglio la composizione cellulare delle lesioni endometriotiche [39].

Durante la fase di progettazione dello studio, sono stati presi in esame numerosi dataset pubblici, reperibili sia nella letteratura scientifica che negli archivi bioinformatici. Dopo un'attenta analisi comparativa delle fonti disponibili, la scelta è ricaduta sui dataset generati tramite tecnologia microarray. Questa tecnologia garantisce l'accesso ad ampi dataset pubblici ampiamente annotati e di buona qualità.

Inoltre, l'impiego dei microarray ha reso possibile un'analisi efficace dell'espressione genica

su un ampio numero di campioni, riuscendo a mantenere un buon compromesso tra qualità dei dati e complessità computazionale.

Capitolo 3

Materiali, metodi e approccio sperimentale

3.1 Revisione degli studi trascrittomici sull'endometriosi

In questo paragrafo viene presentata una revisione dei principali studi trascrittomici condotti sull'endometriosi, con l'obiettivo di riassumere i risultati più significativi e analizzare gli approcci metodologici adottati. La selezione degli articoli è stata effettuata attraverso una ricerca mirata nella banca dati PubMed, considerando criteri quali la tipologia di tecnica trascrittomica utilizzata (RNA-seq, microarray), l'anno di pubblicazione (con priorità agli studi più recenti), la disponibilità di dataset e la presenza di geni differenzialmente espressi. Il confronto tra i vari studi permette di individuare tendenze comuni, differenze metodologiche e risultati contrastanti, contribuendo a delineare lo stato attuale delle conoscenze trascrittomiche sull'endometriosi e le principali questioni ancora irrisolte.

Il dataset GSE11691, pubblicato nel 2008, raccoglie dati di espressione genica ottenuti mediante tecnologia microarray su campioni di tessuto endometriale eutopico ed ectopico prelevati da nove donne affette da endometriosi, confrontati con quelli di nove soggetti sani

utilizzati come controllo. [40]. A partire da questo dataset sono stati condotti diversi studi bioinformatici, tra cui:

- Bioinformatics identification of biomarkers and infiltrating immune cells in endometriosis (2022), in cui sono stati identificati 149 geni differenzialmente espressi, tra cui CXCL12, PDGFRL, AGTR1, PTGER3, S1PR1, con particolare attenzione ai pathway infiammatori e all'infiltrazione di cellule immunitarie tramite algoritmo CIBERSORT [41];
- Unraveling the significance of AGPAT4 for the pathogenesis of endometriosis via a multi-omics approach (2024), che ha individuato AGPAT4 come potenziale gene chiave nella patogenesi della malattia e altri geni quali: ASF1A, CASP9, COMT, GSTA4, LACTB2, NECAP2, NEDD4, NFKB1, NPDC1, RTP4 [42].

Un ulteriore dataset è GSE25628, pubblicato nel 2013, comprende dati di espressione genica ottenuti tramite microarray Affymetrix da tessuti eutopici ed ectopici prelevati da 20 pazienti con endometriosi e da 20 controlli sani. Le ricerche correlate a questo dataset riguardano principalmente l'identificazione di geni e pathway coinvolti nei processi infiammatori e nel rimodellamento della matrice extracellulare. Geni coinvolti nell'organogenesi deregolati nell'endometrio ectopico: BMP4, GREM1 [43].

Il dataset GSE86534, risalente al 2019, invece, fornisce profili di espressione genica ottenuti tramite microarray da campioni di tessuto endometriale eutopico ed ectopico prelevati da un gruppo di 11 donne affette da endometriosi. L'indagine ha evidenziato una significativa sovraregolazione del lncRNA CCDC144NL-AS1 nei campioni ectopici, indicandone un potenziale coinvolgimento nella patogenesi della malattia, in particolare nei processi di migrazione cellulare e infiammazione[44].

Un contributo rilevante sull'analisi delle differenze molecolari tra endometrio eutopico ed

ectopico vi è anche uno studio basato sui dati del dataset GSE23339 (2011) che ha coinvolto 10 pazienti con endometriosi. Attraverso microarray, sono state osservate alterazioni nell'espressione di microRNA chiave: tra quelli sovraregolati, il miR-29c è risultato particolarmente rilevante per la sua funzione di regolatore negativo di geni associati alla matrice extracellulare, come COL7A1 e TFAP2C. Al contrario, i microRNA miR-200bc/429, sottoregolati nei campioni analizzati, sembrano influenzare la regolazione di geni coinvolti nei processi di sviluppo e differenziamento cellulare[45].

Il dataset GSE120103, pubblicato nel 2019, comprende dati di espressione genica ottenuti con tecnologia microarray genome-wide da campioni di endometrio eutopico prelevati da quattro gruppi di donne, ciascuno composto da 30 partecipanti: donne fertili con endometriosi, infertili senza endometriosi, fertili con endometriosi in stadio IV e infertili con endometriosi in stadio IV. Lo studio ha evidenziato trascritti provenienti dal cromosoma Y, suggerendo un fenomeno di microchimerismo fetale. In particolare, geni come ZFY, RPS4Y1, DDX3Y e UTY risultano più frequentemente espressi nei gruppi con endometriosi o infertilità. Questo permette di ipotizzarne un ruolo nei processi infiammatori e nella regolazione dell'immunità locale all'interno dell'endometrio[46].

Il dataset GSE51981, datato 2014, contiene dati trascrittomici generati tramite microarray Affymetrix da campioni di tessuto endometriale prelevati da 148 donne, con e senza endometriosi, analizzati per classificare la malattia e i suoi stadi tramite tecniche genomiche ad alta dimensione. Tra i geni risultati significativamente alterati nell'analisi trascrittomica si segnalano FOSB, FOS, EGR1 e DIO2. I dati di questo stesso dataset sono stati successivamente utilizzati in un altro studio, in cui è stato applicato un approccio basato sul machine learning per l'identificazione di potenziali biomarcatori diagnostici combinati per l'endometriosi. In quell'analisi, sono emersi come rilevanti i geni FOS, EPHX1, DLGAP5, PCSK5 e ADAT1[47, 48].

In conclusione, il confronto tra i principali dataset trascrittomici utilizzati nello studio dell'endometriosi mostra una marcata eterogeneità nei risultati, in particolare per quanto riguarda i geni individuati come differenzialmente espressi. Questa variabilità, che persiste anche in presenza di tecnologie simili, sembra dipendere da numerosi fattori, come le differenze nella selezione dei campioni, nei protocolli sperimentali e nelle strategie di analisi. È una criticità tuttora centrale nella ricerca sull'endometriosi, che evidenzia la necessità di adottare approcci più condivisi e standardizzati. Partendo da queste riflessioni, il prossimo capitolo si concentrerà sull'analisi dettagliata del dataset scelto come riferimento per il confronto con i risultati ottenuti in questo lavoro.

Sintesi dei principali dataset trascrittomici sull'endometriosi

GSE11691 (2008)

Tecnologia: Microarray

Campioni: 9 pazienti, 9 controlli

Principali geni/microRNA: **CXCL12, PDGFRL, AGTR1, PTGER3, S1PR1**

Topic dello studio: Biomarcatori e infiltrazione di cellule immunitarie

GSE25628 (2013)

Tecnologia: Microarray Affymetrix

Campioni: 20 pazienti, 20 controlli

Principali geni/microRNA: **BMP4, GREM1**

Topic dello studio: Processi infiammatori e rimodellamento della matrice extracellulare

GSE86534 (2019)

Tecnologia: Microarray

Campioni: 11 pazienti

Principali geni/microRNA: **lncRNA CCDC144NL-AS1**

Topic dello studio: Migrazione cellulare e infiammazione

GSE23339 (2011)

Tecnologia: Microarray

Campioni: 10 pazienti

Principali geni/microRNA: **miR-29c, miR-200bc/429**

Topic dello studio: Regolazione della matrice extracellulare e differenziamento cellulare

GSE120103 (2019)

Tecnologia: Microarray genome-wide

Campioni: 4 gruppi da 30 donne ciascuno

Principali geni/microRNA: **ZFY, RPS4Y1, DDX3Y, UTY**

Topic dello studio: Microchimerismo fetale e regolazione dell'immunità endometriale

GSE214411 (2022)

Tecnologia: Single-cell RNA-Seq

Campioni: Lesioni endometriosiche da diversi pazienti

Principali geni/microRNA: Analisi trascrittomica a livello di singola cellula, studio dell'eterogeneità cellulare

Topic dello studio: Composizione cellulare delle lesioni endometriosiche e analisi della diversità cellulare

GSE51981 (2014)

Tecnologia: Microarray Affymetrix

Campioni: 148 donne

Principali geni/microRNA: **FOSB, FOS, EGR1, DIO2, EPHX1, DLGAP5, PCSK5, ADAT1**

Topic dello studio: Classificazione della malattia e machine learning per identificazione di biomarcatori

3.2 Scelta dello studio e del dataset per l'analisi

Il dataset GSE141549 è stato scelto come riferimento principale per l'analisi presentata in questa tesi. Pubblicato nel 2020, raccoglie dati trascrittomici di ampia portata, ottenuti tramite piattaforme microarray, che consentono di valutare simultaneamente l'espressione di migliaia di geni. La completezza delle informazioni disponibili e la qualità dei campioni lo rendono una risorsa particolarmente adatta per indagare i meccanismi molecolari associati all'endometriosi. I dettagli relativi alla raccolta dei campioni e alla strutturazione del dataset sono riportati nello studio "A relational database to identify differentially expressed genes in the endometrium and endometriosis lesions" [49], che costituisce il principale riferimento metodologico di questo lavoro.

Fondamentale è stato lo sviluppo di EndometDB, una piattaforma web pensata per facilitare l'esplorazione di dati trascrittomici. L'interfaccia permette di consultare i profili di espressione genica dei diversi campioni e di visualizzarli attraverso strumenti grafici immediati (Figura 3.1). Il database raccoglie le informazioni di espressione genica relative a 115 donne affette da endometriosi e 53 soggetti di controllo, organizzati in maniera tale da permettere un'analisi chiara e immediata. Grazie alla possibilità di interrogare oltre 24.000 geni, rappresenta oggi la risorsa trascrittomica più ampia e completa disponibile per lo studio di questa patologia.

I campioni comprendono differenti tipologie di lesioni, suddivise in più sottotipi in base alla loro localizzazione e morfologia. Gli endometriomi ovarici (OMA) rappresentano le cisti a livello dell'ovaio, mentre le lesioni peritoneali vengono ulteriormente distinte in rosse (PeLR), nere (PeLB) e bianche (PeLW), riflettendo diversi stati di attività della malattia. Le forme più complesse sono le lesioni profondamente infiltranti (DiE), che possono interessare il setto retto-vaginale (REV), il legamento utero-sacrale (SuL), la vescica (DiEB) o l'intestino

(DiEIn). A questi campioni si affiancano tessuti eutopici (endometrio) e peritoneali provenienti sia da pazienti sia da controlli sani, rendendo possibile un confronto sistematico tra tessuto patologico e fisiologico. Tutti i campioni sono stati sottoposti a conferma istologica e sono corredati da dati clinici come età, fase del ciclo mestruale, stadio della malattia (secondo la classificazione ASRM) e stato di eventuali terapie ormonali.

Un elemento metodologico di rilievo riguarda la tecnologia impiegata per la generazione dei dati trascrittomici. Lo studio ha infatti utilizzato due piattaforme di microarray Illumina, HumanHT-12 v4 e HumanWG-6 v2, producendo complessivamente 408 profili di espressione genica. Questo ha reso necessaria una fase accurata di normalizzazione e controllo di qualità, con l'obiettivo di correggere eventuali differenze tecniche e rendere i dati pienamente confrontabili.

EndometDB è una piattaforma integrata in grado di raccogliere, organizzare e analizzare sia i dati clinici sia quelli trascrittomici. Le informazioni cliniche vengono acquisite tramite questionari compilati dalle pazienti attraverso il sistema Webropol e importate automaticamente nel database tramite un'API dedicata. Parallelamente, i risultati numerici derivati dalle analisi biomediche vengono caricati attraverso l'interfaccia web, così da associare in modo strutturato i profili di espressione genica con le caratteristiche cliniche. L'accesso ai dati e la loro esplorazione sono resi possibili da un motore di analisi che interagisce con la GUI (interfaccia grafica utente) della piattaforma. Quest'ultima, sviluppata in PHP, comunica con l'ambiente di calcolo basato su R, che si occupa di elaborare le analisi statistiche e di generare visualizzazioni dinamiche. Per la parte grafica vengono utilizzati i pacchetti ggplot2 e Plotly, strumenti che consentono di ottenere rappresentazioni interattive, di facile lettura anche per utenti senza competenze di programmazione.

In questo modo, EndometDB non è soltanto un archivio di dati, ma uno strumento di analisi accessibile anche a chi non dispone di competenze avanzate di programmazione.

In sintesi, il dataset GSE141549 e la piattaforma EndometDB rappresentano un punto di riferimento fondamentale per lo studio dell'endometriosi. Una piattaforma interattiva, in grado di integrare in maniera coerente dati clinici e trascrittomici, rendendoli esplorabili anche da utenti non esperti.

Questa caratteristica, insieme alla vastità e alla qualità del materiale raccolto, non solo favorisce l'identificazione di pattern molecolari associati alla patologia, ma fornisce anche una risorsa dinamica per validare ipotesi e generare nuove linee di ricerca [49].

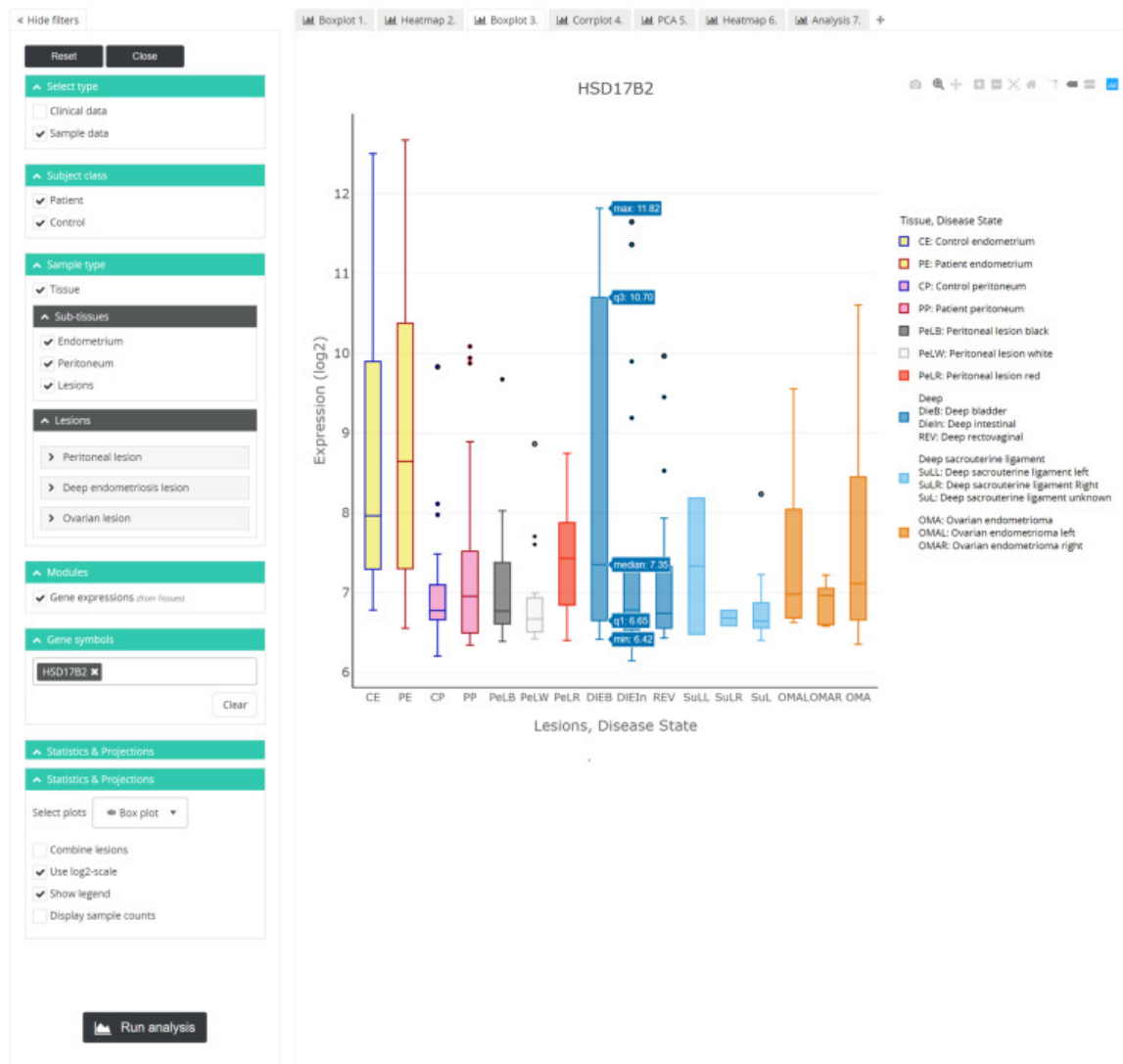


Figura 3.1: Interfaccia del database *EndometDB* con pannelli di selezione (a sinistra) e grafico di output (al centro). Nei menu laterali è possibile scegliere il tipo di dato (clinico o di espressione), la classe di soggetti (pazienti o controlli), il tipo di campione (tessuto o lesione), il gene di interesse (qui *HSD17B2*) e le opzioni di rappresentazione statistica. Il grafico mostra un boxplot dei livelli di espressione ($\log_2$) di *HSD17B2* nei diversi tessuti e lesioni endometriosiche: endometrio e peritoneo di controllo e di pazienti, lesioni peritoneali (bianche e rosse), profonde (vescicali, intestinali, rettovaginali, legamento utero-sacrale) e ovariche (OMA sinistro e destro). La leggenda a destra associa i colori ai diversi tipi di tessuto e stato patologico.

3.3 Analisi bioinformatica dei dati trascrittomici con R

In questo paragrafo è descritto il lavoro svolto partendo dall'analisi del dataset GSE141549, scelto come base di confronto con gli altri studi trattati nel paragrafo precedente. Il dataset è stato scaricato dalla piattaforma GEO (Gene Expression Omnibus), un database pubblico che raccoglie profili di espressione genica prodotti da diversi studi sperimentali e resi disponibili alla comunità scientifica.

Il file utilizzato è disponibile al seguente link: <https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE141549&format=file&file=GSE141549%5FIllumina%5FHumanWG%2D6%5Fv2%2E0%5Fexpression%5Fbeadchip%5Fnormalized%2Exlsx> e contiene dati di espressione genica pre-elaborati dagli autori e forniti in formato già normalizzato e corretti per il batch effect, ottenuti tramite tecnologia microarray (Illumina HumanWG-6 v2.0 expression beadchip).

Il dataset include un'ampia varietà di campioni endometriali, sia eutopici (provenienti dalla normale cavità uterina) che ectopici (cioè localizzati in sedi anomale), appartenenti sia a donne con diagnosi di endometriosi sia a soggetti di controllo. L'obiettivo principale di questa analisi è stato quello di individuare modelli di espressione differenziale tra i diversi tipi di tessuto, così da evidenziare potenziali marcatori molecolari confrontabili con quelli ottenuti da altri dataset trascrittomici discussi nel paragrafo 3.1 . Il file, originalmente in formato Excel (.xlsx), è stato convertito in .csv per semplificarne l'importazione in RStudio, ambiente utilizzato per le successive elaborazioni statistiche e bioinformatiche.

Successivamente, essendo presenti geni replicati all'interno del dataset, è stato utilizzato il comando `aggregate` in R. Questo passaggio ha permesso di raggruppare i dati per simbolo genico (`Gene_symbol`) e calcolare i valori medi dell'espressione, ottenendo un dataset privo di ridondanze e pronto per le analisi di clustering e confronto.

In seguito, è stato costruito un vettore contenente i geni di interesse, selezionati sulla base dei risultati riportati negli articoli analizzati nel paragrafo 3.1. I dati così filtrati sono stati organizzati in una matrice di espressione: le righe corrispondono ai geni selezionati, mentre le colonne rappresentano i campioni analizzati.

La prima rappresentazione grafica è stata realizzata tramite heatmap, la cui scala cromatica (dal blu al rosso, con il bianco come valore intermedio) ha permesso di evidenziare in maniera intuitiva le differenze nei livelli di espressione tra i vari campioni. Per facilitare l'interpretazione biologica dei dati, è stato poi introdotto un sistema di annotazione dei campioni in base al tessuto di provenienza. A tal fine, è stata definita una lista di categorie corrispondenti ai diversi tipi di tessuto inclusi nel dataset, tra cui:

- **REV** – Lesione endometriosa rettovaginale profonda
- **SuL** – Lesione endometriosa del legamento sacrouterino
- **DiEIn** – Lesione endometriosa intestinale
- **DiEB** – Lesione endometriosa profonda nella vescica
- **PeLR** – Lesione endometriosa peritoneale rossa
- **PeLB** – Lesione endometriosa peritoneale nera
- **PeLW** – Lesione endometriosa peritoneale bianca
- **OMA** – Endometrioma ovarico
- **PE** – Endometrio da pazienti
- **CE** – Endometrio da controlli sani
- **PP** – Peritoneo da pazienti

- **CP** – Peritoneo da controlli sani

È stata quindi implementata una funzione (`estrai_tessuto`) in grado di riconoscere e associare automaticamente ciascun campione al tessuto corretto, sulla base delle etichette contenute nel nome della colonna. In questo modo è stato possibile costruire un data frame di annotazioni e applicare una legenda cromatica che assegna a ogni tessuto un colore distinto. Questo consente l'osservazione dei pattern di espressione genica non solo tra i singoli campioni, ma anche a livello di gruppi di tessuto, evidenziando possibili somiglianze o differenze tra le varie sedi endometriosiche e i controlli.

In aggiunta, la costruzione di un istogramma dei valori di espressione tramite la funzione `hist` ha permesso di verificare la distribuzione dei dati, individuando eventuali asimmetrie nei livelli genici.

Per rendere i valori più confrontabili tra geni con scale di espressione diverse, è stata applicata una normalizzazione tramite z-score a livello di riga. Lo z-score è una statistica che misura la distanza tra un dato valore e la media, espressa in unità di deviazione standard. Applicando questa trasformazione a ciascun gene, ogni valore di espressione è stato espresso rispetto alla media del gene stesso, evidenziando le variazioni relative tra i campioni [50].

I dati normalizzati sono stati poi rappresentati nuovamente con una heatmap e con un istogramma, sempre generato con la funzione `hist`, ottenendo una visualizzazione più equilibrata e informativa.

Capitolo 4

Evidenze trascrittomiche e implicazioni cliniche

4.1 Risultati dell'analisi bioinformatica

L'elaborazione del dataset GSE141549, disponibile sulla piattaforma GEO e contenente campioni endometriali eutopici ed ectopici da pazienti affette da endometriosi e da controlli sani, ha consentito di identificare profili di espressione differenziale, posti a confronto con i geni riportati come significativamente alterati negli studi discussi in precedenza nel paragrafo 3.1. Come primo passo, i dati sono stati rappresentati attraverso una heatmap con clustering gerarchico Figura 4.1. Questo approccio ha evidenziato una netta separazione tra i campioni provenienti da lesioni endometriosiche e i tessuti eutopici o di controllo.

In particolare, i campioni patologici (OMA, REV, DiE, SuL, PeLR, PeLB, PeLW) tendono a raggrupparsi in cluster distinti rispetto ai controlli (CE, CP) e ai campioni eutopici (PE, PP). Tale distinzione riflette l'esistenza di profili trascrittomici specifici associati alla malattia, indipendente dalla sede anatomica di origine. Inoltre, all'interno del gruppo patologico, è possibile osservare sottocluster riconducibili alle diverse localizzazioni della malattia (ovaio,

peritoneo, lesioni profonde), suggerendo una parziale eterogeneità trascrittomica che riflette le peculiarità del microambiente tissutale.

Dal punto di vista biologico, i campioni endometriosisi mostrano un'attivazione di geni che riflettono bene la natura della malattia. Spiccano in particolare quelli legati all'infiammazione (NFKB1, PTGER3, CXCL12), alla risposta a stimoli ormonali (FOS, FOSB, EGR1) e ai processi di rimodellamento tissutale e angiogenesi (BMP4, GREM1, AGTR1).

Tra questi, NFKB1 riveste un ruolo centrale: è un regolatore chiave della risposta immunitaria e controlla l'attività di numerose citochine e mediatori pro-infiammatori. Quando è attivato per un tempo prolungato, come avviene nelle lesioni endometriosisi, contribuisce a mantenere uno stato di infiammazione cronica, che rappresenta uno dei tratti distintivi della patologia [51]. Allo stesso modo, PTGER3, recettore delle prostaglandine, partecipa alla modulazione del dolore e della contrattilità tissutale, fenomeni frequentemente alterati nelle pazienti con endometriosisi. Infine, CXCL12 codifica per una chemochina coinvolta nel richiamo di cellule immunitarie e nel mantenimento di un microambiente infiammatorio, contribuendo così alla sopravvivenza e persistenza delle lesioni nel tempo [41].

Parallelamente, si osserva la sovraespressione di geni immediatamente indotti da stimoli ormonali, quali FOS, FOSB ed EGR1. Questi fattori di trascrizione appartengono alla classe dei geni risposta immediata, attivati rapidamente da segnali estrogenici e progestinici, per regolare la crescita e il ciclo cellulare dell'endometrio. Tuttavia, quando questa attivazione diventa persistente, può suggerire un vero e proprio squilibrio nella risposta ormonale delle cellule endometriosisi. In tal modo, queste cellule rispondono in modo eccessivo o prolungato agli stimoli mestruali, favorendo la crescita anomala, la sopravvivenza e la diffusione del tessuto ectopico.

L'analisi trascrittomica ha evidenziato che nei campioni eutopici da donne con endometriosisi i livelli di FOS, FOSB ed EGR1 risultano significativamente elevati rispetto ai controlli

sani. Questi geni, centralmente implicati nella risposta a stimoli ormonali, sono stati inoltre associati ai cambiamenti patologici osservati nelle lesioni endometriosiche. La loro iper-espressione nelle pazienti suggerisce una sensibilità anomala agli estrogeni e al progesterone, che può tradursi in una crescita incontrollata e in una maggiore capacità di adattamento delle cellule ectopiche. Tali alterazioni non solo spiegano parte della fisiopatologia della malattia, ma indicano anche possibili biomarcatori trascrittomici utili a distinguere i tessuti patologici da quelli sani [52].

Un ulteriore elemento cruciale è rappresentato dall'attivazione di vie molecolari legate al rimodellamento tissutale e all'angiogenesi, entrambi i processi indispensabili per la progressione dell'endometriosi. Dunque, un ulteriore elemento cruciale è rappresentato dall'attivazione di vie molecolari legate al rimodellamento tissutale e all'angiogenesi, entrambi i processi indispensabili per la progressione dell'endometriosi. Geni come BMP4 e GREM1 (appartenenti rispettivamente alla famiglia delle proteine morfogenetiche ossee e ai loro antagonisti) controllano il differenziamento e la rigenerazione cellulare.

Una loro disregolazione, come quanto osservato nei campioni endometriosici, può favorire fenomeni fibrotici e una proliferazione cellulare incontrollata [53]. Allo stesso modo, AGTR1, che codifica per il recettore dell'angiotensina II, risulta sovraespresso nelle lesioni endometriosiche e contribuisce sia alla regolazione del tono vascolare sia alla formazione di nuovi vasi sanguigni, creando una rete vascolare che alimenta il tessuto ectopico e ne favorisce la progressione [48].

In contrasto con i geni iper-espressi, altri risultano invece meno attivi, soprattutto quelli coinvolti nei normali sistemi di difesa cellulare e nel mantenimento dell'omeostasi. Tra questi, COMT, GSTA4 ed EPHX1, che svolgono un ruolo chiave nella detossificazione degli estrogeni e nella neutralizzazione delle specie reattive dell'ossigeno (ROS). La loro ridotta attività può portare a un accumulo di estrogeni reattivi e radicali liberi, incrementando lo stress

ossidativo e contribuendo alla creazione di un microambiente favorevole alla progressione della malattia [54].

Un altro elemento di rilievo è CASP9, componente chiave della cascata apoptotica mitocondriale. Il suo livello di espressione ridotto nelle lesioni endometriosiche suggerisce una vera e propria compromissione della capacità delle cellule di attivare i meccanismi di morte cellulare programmata. Dunque, queste cellule risultano più resistenti all'apoptosi, acquisendo abilità superiori nel sopravvivere, colonizzare sedi ectopiche e sfuggire ai meccanismi fisiologici di eliminazione[55].

Nel complesso, l'analisi conferma che le lesioni endometriosiche presentano un microambiente molecolare caratterizzato da infiammazione cronica, squilibrio ormonale, resistenza all'apoptosi e fenomeni di rimodellamento fibrotico/angiogenico, in linea con quanto riportato in letteratura sulla fisiopatologia della malattia.

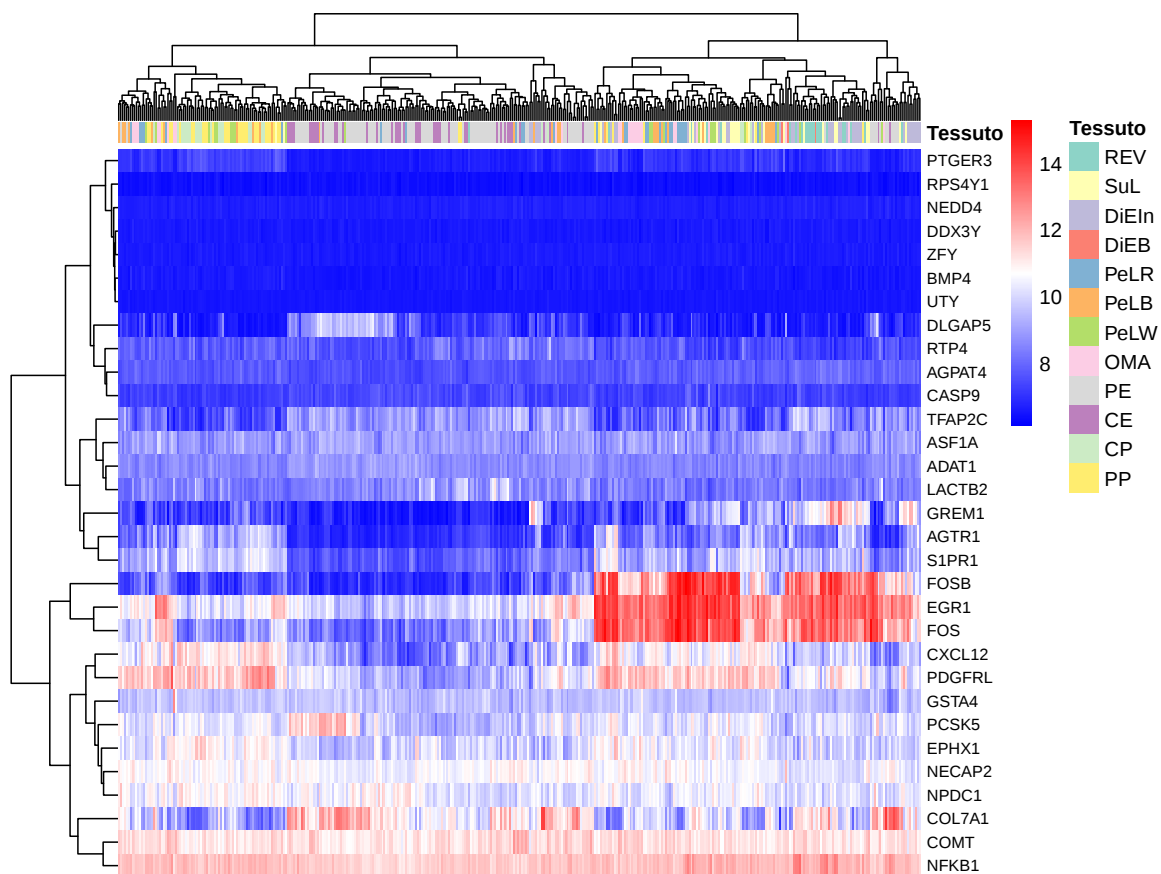


Figura 4.1: Heatmap dei geni selezionati dal dataset GSE141549.

Le righe rappresentano i geni analizzati e le colonne i campioni, mentre l'intensità cromatica indica i livelli di espressione (blu = valori più bassi, rosso = valori più alti). I dendrogrammi laterali e superiori mostrano il clustering gerarchico di geni e campioni. In alto, la barra di annotazione evidenzia l'appartenenza dei campioni ai diversi tipi di tessuto (REV, SuL, DiEIn, DiEB, PeLR, PeLB, PeLW, OMA, PE, CE, PP, CP).

Per valutare la qualità complessiva del dataset e la distribuzione dei livelli di espressione genica, è stato realizzato un istogramma dei valori (Figura 4.2). L'andamento mostra una distribuzione sostanzialmente unimodale, con la maggior parte dei geni concentrata nell'intervallo 8–10, mentre solo Dal grafico emerge una distribuzione abbastanza regolare e di tipo unimodale: la maggior parte dei geni mostra valori compresi tra 8 e 10, mentre solo una piccola parte si colloca su livelli molto bassi o particolarmente alti.

Questa configurazione è indicativa di una corretta normalizzazione dei dati, in quanto non emergono picchi anomali, confermando l'affidabilità del dataset. Dunque, il dataset non presenta distorsioni tecniche e i risultati delle analisi possono essere interpretati con maggiore sicurezza, sapendo che riflettono variazioni biologiche reali e non artefatti legati al processo di elaborazione.

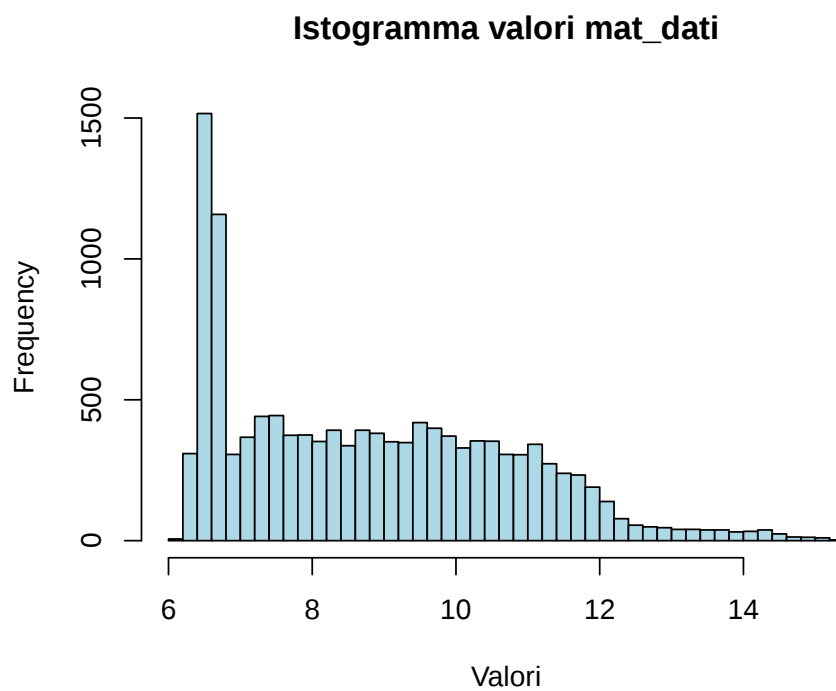


Figura 4.2: Istogramma dei livelli di espressione (dataset GSE141549).

Rappresentazione dei valori relativi al sottoinsieme di geni selezionati sulla base della letteratura (par. 3.1). Sull'asse delle ascisse sono riportati i valori di espressione, mentre sull'asse delle ordinate la frequenza dei geni osservati in ciascun intervallo. La distribuzione consente di visualizzare l'andamento complessivo dei dati all'interno di questo gruppo genico mirato.

Dopo la rappresentazione iniziale dei livelli di espressione, è stato introdotto un ulteriore passaggio di normalizzazione mediante z-score. Questa normalizzazione permette di attenuare le differenze dovute alle diverse scale di espressione tra i geni. Pertanto, ogni valore non viene più letto in senso assoluto, ma in base a quanto si discosta dalla media del gene stesso. In questo modo, l'analisi si concentra sulle variazioni relative tra i campioni, permettendo un confronto più omogeneo e una lettura più immediata dei pattern trascrittomici.

La heatmap ottenuta dopo questa elaborazione, (Figura 4.3, mostra in modo più nitido la distinzione tra campioni patologici e controlli, confermando i pattern già osservati con la rappresentazione precedente. Inoltre, la standardizzazione ha permesso di mettere in risalto con maggiore chiarezza i sottocluster all'interno del gruppo patologico, suggerendo la presenza di specificità trascrittomiche associate alle diverse localizzazioni delle lesioni (ad esempio ovaio, peritoneo, lesioni profonde).

L'applicazione dello z-score, quindi, non solo ha migliorato la qualità della visualizzazione, ma ha anche permesso di cogliere meglio le peculiarità trascrittomiche delle diverse tipologie di tessuto, evidenziando l'eterogeneità biologica che caratterizza l'endometriosi.

Dal punto di vista biologico, i geni già individuati nella heatmap precedente (Figura 5.2) come caratteristici della malattia (legati a infiammazione, risposta ormonale e angiogenesi) mantengono un andamento coerente, ma appaiono distribuiti secondo gradienti più netti. La mappa evidenzia gruppi di geni la cui espressione risulta accentuata proprio nei tessuti endometriosici, mentre nei campioni eutopici e nei controlli tali segnali appaiono molto più attenuati. Questo rende più evidente la differenza trascrittomica associata alle lesioni endometriosiche.

In parallelo, i geni legati ai processi di detossificazione e di morte cellulare programmata risultano meno espressi in maniera costante nei campioni endometriosici. Questi risultati indicano che una ridotta sensibilità allo stress ossidativo e una limitata attivazione dei mec-

canismi apoptotici costituiscono caratteristiche comuni a più tipologie di lesione.

Grazie alla standardizzazione mediante z-score, emergono in maniera più evidente anche le specificità trascrittomiche proprie dei diversi sottotipi di lesione; infatti, si osserva che gli endometriomi ovarici tendono a differenziarsi dalle lesioni localizzate a livello peritoneale o profondo, un dato che suggerisce come il contesto anatomico giochi un ruolo rilevante nel modulare l'espressione genica.

Nel complesso, la rappresentazione attraverso z-score non solo conferma i risultati dell'heat-map precedente (Figura 5.2), ma offre una lettura più precisa e comparabile dei dati. Questa normalizzazione rende infatti più immediato il confronto tra geni con scale di espressione differenti, facilitando l'individuazione sia dei tratti molecolari condivisi dalle diverse lesioni, sia delle specificità che distinguono i diversi sottotipi di tessuti.

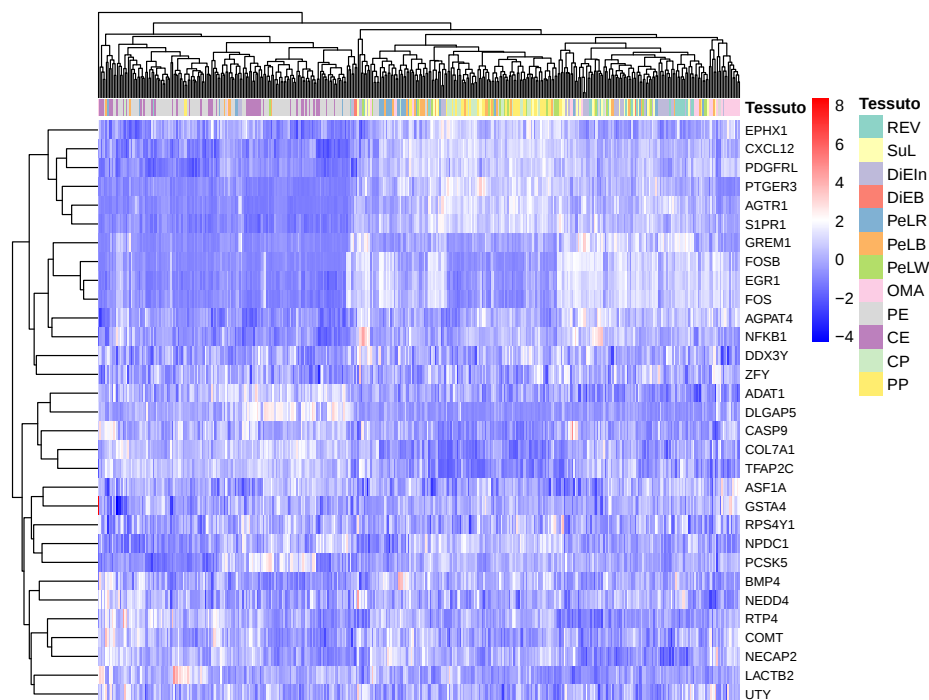


Figura 4.3: Heatmap dei geni selezionati dal dataset GSE141549, elaborata mediante normalizzazione con z-score.

Le righe rappresentano i geni analizzati, mentre le colonne corrispondono ai campioni. L'intensità cromatica riflette i valori di espressione standardizzati (blu = sotto-espressione relativa, rosso = sovra-espressione relativa). I dendrogrammi laterali e superiori mostrano il clustering gerarchico dei geni e dei campioni. In alto, una barra di annotazione indica l'appartenenza dei campioni ai diversi tipi di tessuto (REV, SuL, DiEIn, DiEB, PeLR, PeLB, PeLW, OMA, PE, CE, PP, CP). L'applicazione dello z-score ha reso i dati confrontabili tra geni con scale di espressione differenti, permettendo di mettere in evidenza con maggiore chiarezza sia i tratti comuni, sia le differenze tra i diversi sottotipi tissutali.

Dopo la normalizzazione con z-score, è stato realizzato un istogramma dei valori di espressione (Figura 4.4).

A differenza della distribuzione osservata in precedenza, qui i dati risultano centrati attorno allo zero, con la maggior parte dei valori compresi nell'intervallo tra -2 e +2. Questo comportamento è tipico della trasformazione z-score e rappresenta un'ulteriore conferma della corretta standardizzazione dei dati. In questo modo, i livelli di espressione non vengono più interpretati in termini assoluti, ma in funzione dello scostamento rispetto alla media di ciascun gene, consentendo una lettura più omogenea e affidabile delle variazioni trascrittomiche tra i campioni.

La distribuzione simmetrica dei valori permette di mettere sullo stesso piano geni che, in origine, presentavano livelli di espressione molto diversi, rendendo più immediata l'individuazione di variazioni relative tra campioni. Inoltre, l'assenza di picchi anomali o andamenti irregolari rafforza l'idea che la trasformazione abbia reso i dati più omogenei e comparabili, fornendo una base solida per le successive interpretazioni biologiche.

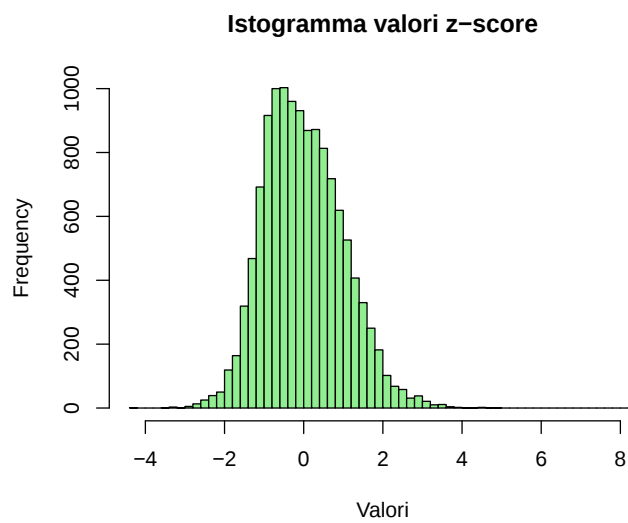


Figura 4.4: Istogramma dei valori di espressione normalizzati tramite z-score (dataset GSE141549).

Sull'asse delle ascisse sono riportati i valori standardizzati (z) di espressione genica; in ordinata è indicata la frequenza. La distribuzione è centrata intorno allo zero e mostra la maggior parte dei valori compresi tra -2 e 2 , come atteso dall'applicazione dello z-score. La forma simmetrica, priva di picchi anomali, conferma l'efficacia della standardizzazione e consente un confronto più omogeneo tra geni con scale di espressione originariamente diverse.

4.2 Riflessioni conclusive

L'analisi condotta sul dataset GSE141549 ha permesso di delineare i profili trascrittomici che caratterizzano i tessuti delle pazienti affette da endometriosi, confermando molte delle ipotesi formulate sulla fisiopatologia della malattia e, al tempo stesso, evidenziandone la complessità intrinseca.

In particolare, la netta separazione osservata tra i campioni patologici e quelli eutopici o di controllo evidenzia come l'endometriosi non sia una semplice estensione del tessuto endometriale normale, ma una condizione con tratti molecolari propri e ben distinti. Dall'indagine effettuata sono stati confermati meccanismi molecolari legati alla patologia già noti: l'attivazione di vie infiammatorie croniche (NFKB1, PTGER3, CXCL12), la risposta alterata agli stimoli ormonali (FOS, FOSB, EGR1), i processi di rimodellamento e angiogenesi (BMP4, GREM1, AGTR1), la resistenza all'apoptosi (CASP9 ridotto) e lo squilibrio ossidativo (COMT, GSTA4, EPHX1 ipo-espressi). Questi meccanismi evidenziano come l'endometriosi si sviluppi in un microambiente infiammatorio persistente, fortemente influenzato dagli ormoni, che sostiene la sopravvivenza e la proliferazione delle cellule ectopiche.

I dati ottenuti confermano la complessità della patologia. Lo studio dei profili molecolari delle lesioni endometriosiche ha messo in risalto la presenza di sottogruppi trascrittomici, suggerendo che l'endometriosi non sia una condizione uniforme, ma manifesti caratteristiche variabili in base alla localizzazione delle lesioni.

Gli endometriomi ovarici, le lesioni peritoneali e quelle profonde mostrano, infatti, peculiarità trascrittomiche che potrebbero essere legate ai diversi microambienti in cui si sviluppano. Questo risultato è particolarmente significativo perché contribuisce a spiegare la variabilità clinica della malattia: pazienti con localizzazioni differenti non solo presentano quadri sintomatologici diversi, ma potrebbero essere caratterizzate da meccanismi molecolari distinti,

con implicazioni sulla risposta alle terapie. Questo potrebbe spiegare perché alcune donne rispondano bene a determinati trattamenti farmacologici o chirurgici, mentre altre non ottengono benefici o vanno incontro a recidive.

L'articolazione e variabilità della malattia rendono dunque necessario un approccio personalizzato, che consideri non solo lo stadio e la localizzazione delle lesioni, ma anche le caratteristiche molecolari che le contraddistinguono, così da orientare meglio la gestione clinica e aumentare l'efficacia delle terapie. È proprio questa diversità molecolare e biologica a spiegare perché sia così difficile individuare biomarcatori diagnostici realmente affidabili. Pertanto, i geni coinvolti non si esprimono in maniera esclusiva nelle lesioni, il che ne riduce il potenziale utilizzo a fini diagnostici. In diversi casi, infatti, alterazioni simili si osservano anche nei tessuti eutopici delle pazienti, rendendo difficile tracciare una linea netta tra tessuto sano e malato.

In considerazione di quanto emerso, appare chiaro che l'endometriosi non possa essere descritta attraverso un singolo gene o un unico meccanismo molecolare. È piuttosto il risultato di un intreccio complesso di fattori molecolari e contestuali, che richiede strumenti diagnostici e terapeutici in grado di coglierne la multidimensionalità.

Questa complessità si riflette in un aspetto critico della gestione della malattia: la diagnosi. Nonostante i progressi nella ricerca, la diagnosi di endometriosi resta spesso tardiva, con tempi che possono superare diversi anni tra l'insorgenza dei primi sintomi e diagnosi definitiva. Pertanto, non esiste un singolo profilo genico in grado di identificare con certezza l'endometriosi; le alterazioni variano invece in base al sito delle lesioni e alle caratteristiche individuali delle pazienti. Questo spiega perché geni come NFkB1 o CXCL12, pur essendo strettamente legati ai processi patologici, non possano essere utilizzati da soli come marker diagnostici. In sintesi, la complessità e l'eterogeneità dell'endometriosi rendono particolarmente difficile individuare biomarcatori univoci e immediatamente applicabili in clinica.

Da un punto di vista metodologico, l'analisi dei dati ha mostrato coerenza nei risultati. Lo z-score ha reso confrontabili geni con livelli di espressione molto diversi, evidenziando pattern trascrittomici chiari e facilmente interpretabili. Le heatmap generate hanno rivelato sia tratti comuni tra i tessuti patologici, sia differenze specifiche tra le varie tipologie di lesione, confermando la coerenza dei risultati ottenuti. Va però sottolineato che lo studio presenta dei limiti poiché l'analisi si basa su un singolo dataset e non include validazioni sperimentali indipendenti (come qPCR o immunoistochimica), il che riduce la possibilità di estendere immediatamente le conclusioni a contesti clinici più ampi.

In sintesi, questo studio contribuisce a rafforzare l'idea che l'endometriosi sia una patologia complessa, multifattoriale e caratterizzata da una marcata eterogeneità. Lo studio di geni chiave e sottocluster trascrittomici offre nuove prospettive sulla fisiopatologia della malattia, ma mette anche in evidenza le sfide ancora aperte nella ricerca di biomarcatori specifici e nella realizzazione di strumenti diagnostici rapidi, precisi e non invasivi.

L'endometriosi appare così come una condizione dinamica, in cui componenti comuni come infiammazione, squilibrio ormonale e angiogenesi si intrecciano con differenze locali legate alle interazioni tra cellule ectopiche e microambiente.

Questa complessità contribuisce a spiegare le difficoltà ancora presenti nella diagnosi e nel trattamento della malattia, evidenziando al contempo l'importanza di un approccio multidimensionale per affrontare un problema clinico e sociale di grande rilevanza.

4.3 Prospettive di ricerca e applicazioni terapeutiche

I risultati ottenuti confermano che l'endometriosi è una patologia complessa ed eterogenea, con profili molecolari che variano a seconda della sede delle lesioni e del microambiente locale. Approfondire lo studio di questi sottocluster trascrittomici può consentire di identificare geni chiave coinvolti nei processi di infiammazione, angiogenesi, resistenza all'apoptosi e squilibrio ossidativo, aprendo la strada a strategie terapeutiche mirate in grado di intervenire sui meccanismi specifici presenti in ciascun sottotipo di lesione.

Combinare i dati trascrittomici con altre informazioni “-omiche”, come quelle proteomiche o epigenomiche, rappresenta una delle strade più promettenti per individuare biomarcatori più affidabili. Questi strumenti potrebbero aiutare a diagnosticare l'endometriosi in fase precoce e a scegliere le terapie più adatte e mirate, prevedendo quali pazienti potrebbero rispondere meglio a farmaci o interventi chirurgici. Grazie all'impiego di algoritmi avanzati e dell'intelligenza artificiale, è possibile interpretare più facilmente dati così complessi, individuare schemi ricorrenti e distinguere sottogruppi di pazienti, aprendo la strada ad un approccio clinico sempre più personalizzato.

Oltre alla prospettiva molecolare, non va trascurato, inoltre, l'aspetto clinico e multidisciplinare: integrare ginecologia, gestione del dolore cronico e supporto psicologico consente un approccio più completo alla paziente, combinando chirurgia, terapia farmacologica e interventi mirati al microambiente.

È essenziale sensibilizzare medici e pazienti sulla particolarità della malattia. La consapevolezza dei diversi quadri clinici e della variabilità sintomatica può ridurre i tempi di diagnosi e favorire percorsi terapeutici più tempestivi.

In definitiva, le prospettive future nella ricerca sull'endometriosi puntano a integrare le conoscenze trascrittomiche e multi-omiche con un approccio clinico personalizzato, in grado

di affrontare la complessità biologica della malattia, la diversità dei sintomi e le differenti reazioni ai trattamenti. Combinare profili molecolari dettagliati con strategie terapeutiche mirate offre una concreta opportunità per migliorare la diagnosi precoce, ottimizzare la gestione clinica e ridurre l'impatto sociale della patologia.

L'obiettivo non è solo quello di fornire strumenti più efficaci e mirati per le pazienti, ma rappresenta anche un passo importante verso una maggiore consapevolezza e valorizzazione dell'endometriosi.

Guardando avanti, l'unione tra ricerca molecolare, innovazione tecnologica e cura personalizzata apre la strada a un futuro in cui le donne affette da questa condizione possano vivere con maggiore qualità di vita, fiducia nella gestione terapeutica e speranza concreta per soluzioni più efficaci.

Capitolo 5

Codice sorgente

```
1 # Dati scaricati da GEO (NCBI):
2 # https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE141549&format=file&file=
   GSE141549%5Fillumina%5FHumanWG%2D6%5Fv2%2E0%5Fexpression%5Fbeadchip%5
   Fnormalized%2Exlsx
3
4 library(readr)
5 library(gplots)
6 library(RColorBrewer)
7 library(pheatmap)
8
9 # Caricamento dei dati normalizzati
10 dati <- read_csv("~/Downloads/GSE141549_
   batchCorrectednormalizedArrayscombined.csv")
11 View(dati)
12
13 # Aggregazione dei valori per gene
14 agg_dati <- aggregate(. ~ Gene_symbol,
15                       data = dati[, -2], FUN = function(x) mean(unique(x), na.
16                                                                 rm = TRUE))
17
18 # Definizione della lista di geni di interesse
19 geni_di_interesse <- c(
20   "CXCL12", "PDGFR1", "AGTR1", "PTGER3", "S1PR1", "AGPAT4", "ASF1A",
21   "CASP9", "COMT", "GSTA4", "LACTB2", "NECAP2", "NEDD4", "NFKB1",
22   "NPDC1", "RTP4", "BMP4", "GREM1", "CCDC144NL-AS1", "COL7A1", "TFAP2C",
23   "RPS4Y1", "DDX3Y", "ZFY", "UTY", "FOSB", "FOS", "EGR1", "DIO2",
24   "EPHX1", "DLGAP5", "PCSK5", "ADAT1", "miR-29c", "miR-200bc/429"
25 )
26
27 # Estrazione dei soli geni selezionati
28 agg_geni <- agg_dati[agg_dati$Gene_symbol %in% geni_di_interesse, ]
```

```

29
30 # Creazione della matrice con geni come righe e campioni come colonne
31 gene_names <- agg_geni$Gene_symbol
32 mat_dati <- as.matrix(agg_geni[, -1])
33 rownames(mat_dati) <- gene_names
34
35 # Heatmap di base
36 heatmap(mat_dati,
37         col = colorRampPalette(c("blue", "white", "red"))(256),
38         scale = "none")
39
40 # Heatmap con la funzione heatmap.2
41 heatmap.2(mat_dati,
42         col = colorRampPalette(c("blue", "white", "red"))(256),
43         trace = "none",
44         key = TRUE,
45         key.title = "Valore",
46         key.xlab = "Intensity")
47
48 # Definizione dei tessuti validi (per annotazioni nelle heatmap)
49 tessuti_validi <- c(
50     "REV", "SuL", "DiEIn", "DiEB", "PeLR", "PeLB", "PeLW",
51     "OMA", "PE", "CE", "CP", "PP"
52 )
53
54 # Funzione per estrarre il tessuto dal nome della colonna
55 estrai_tessuto <- function(nome) {
56     parti <- unlist(strsplit(nome, "_"))
57     for (i in length(parti):1) {
58         if (parti[i] %in% tessuti_validi) {
59             return(parti[i])
60         }
61     }
62     return(NA)
63 }
64
65 # Creazione delle annotazioni dei campioni per pheatmap
66 tessuti_sample <- sapply(colnames(mat_dati), estrai_tessuto)
67
68 annotazioni <- data.frame(
69     Tessuto = factor(tessuti_sample, levels = tessuti_validi)
70 )
71 rownames(annotazioni) <- colnames(mat_dati)
72
73 colori_tessuti <- setNames(
74     colorRampPalette(brewer.pal(12, "Set3"))(length(tessuti_validi)),
75     tessuti_validi
76 )
77

```

```

78 # Heatmap con pheatmap e annotazioni
79 pheatmap(mat_dati,
80          annotation_col = annotazioni,
81          annotation_colors = list(Tessuto = colori_tessuti),
82          scale = "none",
83          show_colnames = FALSE,
84          show_rownames = TRUE,
85          fontsize_row = 8,
86          legend = TRUE,
87          annotation_legend = TRUE,
88          color = colorRampPalette(c("blue", "white", "red"))(256),
89          border_color = NA)
90
91 # Istogramma dei valori grezzi
92 hist(as.vector(mat_dati),
93      main = "Istogramma_valori_mat_dati",
94      xlab = "Valori",
95      col = "lightblue",
96      breaks = 50)
97
98 # Calcolo dello z-score per ogni gene
99 zscore_row <- function(x) {
100   if (sd(x) == 0) return(rep(0, length(x))) # evita NaN
101   (x - mean(x)) / sd(x)
102 }
103 mat_dati_zscore <- t(apply(mat_dati, 1, zscore_row))
104
105 # Heatmap con pheatmap (z-score)
106 tessuti_sample <- sapply(colnames(mat_dati_zscore), estrai_tessuto)
107 annotazioni <- data.frame(
108   Tessuto = factor(tessuti_sample, levels = tessuti_validi)
109 )
110 rownames(annotazioni) <- colnames(mat_dati_zscore)
111
112 pheatmap(mat_dati_zscore,
113          annotation_col = annotazioni,
114          annotation_colors = list(Tessuto = colori_tessuti),
115          scale = "none",
116          show_colnames = FALSE,
117          show_rownames = TRUE,
118          fontsize_row = 8,
119          legend = TRUE,
120          annotation_legend = TRUE,
121          color = colorRampPalette(c("blue", "white", "red"))(256),
122          border_color = NA)
123
124 # Istogramma dei valori z-score
125 hist(as.vector(mat_dati_zscore),
126      main = "Istogramma_valori_z-score",

```

```
127     xlab = "Valori",  
128     col = "lightgreen",  
129     breaks = 50)
```

Bibliografia

- [1] Krina T. Zondervan, Christian M. Becker e Stacey A. Missmer. «Endometriosis: impact on quality of life, productivity, and access to care». In: *International Journal of Environmental Research and Public Health* 17.13 (2020), p. 4683. DOI: 10.3390/ijerph17134683. URL: <https://www.mdpi.com/1660-4601/17/13/4683>.
- [2] Suzie As-Sanie. «Endometriosis prevalence study». In: *Journal of Women's Health* (2025). URL: <https://www.example.com/AsSanie2025>.
- [3] Philippe R. Koninckx et al. «The epidemiology of endometriosis is poorly known as the pathophysiology and diagnosis are unclear». In: *Best Practice & Research Clinical Obstetrics & Gynaecology* 71 (2021), pp. 14–26. DOI: 10.1016/j.bpobgyn.2020.11.003. URL: <https://doi.org/10.1016/j.bpobgyn.2020.11.003>.
- [4] P. Parasar, P. Ozcan e K. Terry. «Endometriosis: Epidemiology, Diagnosis and Clinical Management». In: *Current Obstetrics & Gynecology Reports* (2017). URL: <https://doi.org/10.1007/s13669-017-0190-2>.
- [5] Kelechi E. Nnoaham et al. «Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries». In: *Fertility & Sterility* 96.2 (2011), 366–373.e8. DOI: 10.1016/j.fertnstert.2011.05.090. URL: <https://doi.org/10.1016/j.fertnstert.2011.05.090>.

- [6] Patricia Ribeiro de Carvalho França, Anna Carolina Pereira Lontra e Patricia Dias Fernandes. «Endometriosis: A Disease with Few Direct Treatment Options». In: *Molecules* 27.13 (2022). Laboratório de Farmacologia da Dor e da Inflamação, Universidade Federal do Rio de Janeiro, Brazil, p. 4034. doi: 10.3390/molecules27134034. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9268675/>.
- [7] Krina T. Zondervan et al. «Endometriosis». In: *Nature Reviews Disease Primers* 4 (2018), p. 9. doi: 10.1038/s41572-018-0008-5.
- [8] J. Halban. «Endometriosis: A Study of the Pathogenesis of Endometriosis». In: *Archiv für Gynäkologie* 157 (1936), pp. 1–20. doi: 10.1007/BF02045946. URL: <https://link.springer.com/article/10.1007/BF02045946>.
- [9] T. Maruyama. «A Revised Stem Cell Theory for the Pathogenesis of Endometriosis». In: *Journal of Personalized Medicine* 12.2 (2022), p. 216. doi: 10.3390/jpm12020216. URL: <https://www.mdpi.com/2075-4426/12/2/216>.
- [10] P. Viganò et al. «Environmental Factors and Endometriosis». In: *Gynecological Endocrinology* 31.6 (2015), pp. 417–422. doi: 10.3109/09513590.2015.1026162. URL: <https://www.tandfonline.com/doi/full/10.3109/09513590.2015.1026162>.
- [11] Bartosz Smolarz, Karol Szyłło e Halina Romanowicz. «Endometriosis: Epidemiology, classification, pathogenesis, treatment and genetics (Review of literature)». In: *International Journal of Molecular Sciences* 22.19 (2021), p. 10554. doi: 10.3390/ijms221910554.
- [12] C. Chapron et al. «Rethinking Mechanisms, Diagnosis and Management of Endometriosis». In: *Nature Reviews Endocrinology* 15 (2019), pp. 666–682. doi: 10.1038/s41574-019-0245-z.

- [13] Paolo Vercellini et al. «Endometriosis: Pathogenesis and Treatment». In: *Nature Reviews Endocrinology* 10 (2014), pp. 261–275. DOI: 10.1038/nrendo.2013.255.
- [14] A. Adilbayeva e J. Kunz. «Pathogenesis of Endometriosis and Endometriosis-Associated Cancers». In: *International Journal of Molecular Sciences* 25.14 (lug. 2024). Accessed: 2025-09-10, p. 7624. URL: <https://pubmed.ncbi.nlm.nih.gov/39062866/>.
- [15] K. T. Zondervan, C. M. Becker e S. A. Missmer. «Endometriosis». In: *The New England Journal of Medicine* (2020). Advance online publication. DOI: 10.1056/NEJMr1810764. URL: <https://doi.org/10.1056/NEJMr1810764>.
- [16] E. Pašalić, M. M. Tambuwala e A. Hromić-Jahjefendić. «Endometriosis: Classification, pathophysiology, and treatment options». In: *Pathophysiology* (2023). Accessed: 2025-09-10. DOI: 10.1016/j.prp.2023.154847. URL: <https://pubmed.ncbi.nlm.nih.gov/37844487/>.
- [17] S. S. Singh et al. «Guideline No. 449: Diagnosis and Impact of Endometriosis – A Canadian Guideline». In: *Journal of Obstetrics and Gynaecology Canada* 46 (2024). <https://doi.org/10.1016/j.jogc.2023.102450>, p. 102450. DOI: 10.1016/j.jogc.2023.102450.
- [18] C. M. Becker et al. «ESHRE guideline: endometriosis». In: *Human Reproduction Open* 2022.2 (2022), hoac009. DOI: 10.1093/hropen/hoac009.
- [19] Andrew W. Horne e Stacey A. Missmer. «Pathophysiology, diagnosis, and management of endometriosis». In: *BMJ* 380 (2023), e072540. DOI: 10.1136/bmj-2022-072540.
- [20] Márcia Guimarães Silva Da Broi, Rui Alberto Ferriani e Paula Andrea Navarro. «Meccanismi etiopatogeni dell’endometriosi: infertilità correlata». In: *JBRA Assisted Reproduction* 23.3 (2019), pp. 273–280. DOI: 10.5935/1518-0557.20190029. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6836944/>.

- [21] J. R. Prins, N. Gomes-Lopez e S. A. Robertson. «Interleukin-6 in Pregnancy and Gestational Disorders». In: *Journal of Reproductive Immunology* 95 (2012), pp. 1–14. DOI: 10.1016/j.jri.2012.05.004.
- [22] Sergio Ferrero, Filippo Barra e Umberto L. R. Maggiore. «Current and Emerging Therapeutics for the Management of Endometriosis». In: *Drugs* 78 (2018), pp. 995–1012. DOI: 10.1007/s40265-018-0928-0.
- [23] Sara Farag et al. «Management, Prevention, and Sequelae of Adhesions in Women Undergoing Laparoscopic Gynecologic Surgery: A Systematic Review». In: *Journal of Minimally Invasive Gynecology* 29.4 (2022), pp. 543–558. DOI: 10.1016/j.jmig.2021.11.015.
- [24] Mee-Ran Kim et al. «Clinical Diagnosis and Early Medical Management for Endometriosis: Consensus from Asian Expert Group». In: *Gynecology and Minimally Invasive Therapy* 11.3 (2022), pp. 123–132. DOI: 10.4103/GMIT.GMIT_59_21.
- [25] Relly Yanuari Primariawan et al. «Clinical Diagnosis and Early Medical Management for Endometriosis: Consensus from Asian Expert Group». In: *Reproductive Medicine and Biology* 22.1 (2023), e12570. DOI: 10.1002/rmb2.12570.
- [26] L.C. Kao et al. «Expression profiling of endometrium from women with endometriosis reveals candidate genes for disease-based implantation failure and infertility». In: *Fertility and Sterility* 80 (5 2003), pp. 1144–1154. DOI: 10.1016/S0015-0282(03)01104-2.
- [27] Rohan Lowe et al. «Transcriptomics technologies». In: *PLOS Computational Biology* 13.5 (2017), e1005457. DOI: 10.1371/journal.pcbi.1005457. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5436640/>.

- [28] G. Piétu et al. «The Genexpress IMAGE knowledge base of the human brain transcriptome: a prototype integrated resource for functional and computational genomics». In: *Genome Research* 9 (1999), pp. 195–209. URL: <https://genome.cshlp.org/content/9/3/195>.
- [29] V.E. Velculescu et al. «Characterization of the yeast transcriptome». In: *Cell* 88 (1997), pp. 243–251. URL: [https://www.cell.com/fulltext/S0092-8674\(00\)81848-0](https://www.cell.com/fulltext/S0092-8674(00)81848-0).
- [30] M. Schena et al. «Quantitative monitoring of gene expression patterns with a complementary DNA microarray». In: *Science* 270.5235 (1995), pp. 467–470. DOI: 10.1126/science.270.5235.467. URL: <https://doi.org/10.1126/science.270.5235.467>.
- [31] Alexander E. Pozhitkov e Igor N. Nifantiev. «Microarray technology for transcriptome analysis». In: *Journal of Bioinformatics* 23 (2007), pp. 123–135. URL: <https://www.example.com/pozhitkov2007>.
- [32] Christian Ambroise e Francois McLoughlin. «Optimization of cDNA microarray sensitivity and specificity». In: *BioTechniques* 39 (2005), pp. 45–52. URL: <https://www.example.com/ambroise2005>.
- [33] Z. Wang, M. Gerstein e M. Snyder. «RNA-Seq: a revolutionary tool for transcriptomics». In: *Nature Reviews Genetics* 10 (2009), pp. 57–63. DOI: 10.1038/nrg2484. URL: <https://www.nature.com/articles/nrg2484>.
- [34] Oksana Morozova e Marco A. Marra. «Applications of next-generation sequencing technologies in functional genomics». In: *Genomics* 92 (2009), pp. 255–264. DOI: 10.1016/j.ygeno.2008.12.010. URL: <https://www.sciencedirect.com/science/article/pii/S0888754309000051>.

- [35] S. Brenner et al. «Quantitative analysis of gene expression by serial analysis of gene expression (SAGE)». In: *Science* 287 (2000), pp. 1279–1282. URL: <https://www.science.org/doi/10.1126/science.287.5460.1279>.
- [36] B. C. Meyers, L. Tu e Z. Liu. «MPSS for Arabidopsis thaliana transcriptome analysis». In: *Plant Cell* 16 (2004), pp. 1286–1297. URL: <https://www.plantcell.org/content/16/5/1286>.
- [37] T. Lappalainen, M. Sammeth e M. R. Friedländer. «Transcriptome and genome sequencing uncovers functional variation in humans». In: *Nature* 501 (2013), pp. 506–511. DOI: 10.1038/nature12531. URL: <https://www.nature.com/articles/nature12531>.
- [38] F. Ozsolak e P. M. Milos. «RNA sequencing: advances, challenges and opportunities». In: *Nature Reviews Genetics* 12 (2011), pp. 87–98. DOI: 10.1038/nrg2934. URL: <https://www.nature.com/articles/nrg2934>.
- [39] Gene Expression Omnibus. *GSE214411: Single-cell RNA-seq dataset on endometriosis lesions*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE214411>. Accessed: 2025-09-09.
- [40] Adam Hever et al. *GSE11691: Microarray analysis of eutopic and ectopic endometrial tissues from women with endometriosis*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11691>. NCBI Gene Expression Omnibus. 2010.
- [41] Xiaoyu Li et al. «Bioinformatics identification of biomarkers and infiltrating immune cells in endometriosis». In: *Medicine* 101.1 (2022), e28510. DOI: 10.1097/MD.00000000000028510. URL: <https://doi.org/10.1097/MD.00000000000028510>.
- [42] Jun Chen et al. «Unraveling the significance of AGPAT4 for the pathogenesis of endometriosis via a multi-omics approach». In: *Human Genetics* 143.9-10 (2024),

- pp. 1163–1174. DOI: 10.1007/s00439-024-02681-2. URL: <https://doi.org/10.1007/s00439-024-02681-2>.
- [43] Amy Hever et al. *GSE25628: Gene expression in eutopic and ectopic endometrium from women with endometriosis and controls*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25628>. NCBI Gene Expression Omnibus, accessed 2025-09-10. 2013.
- [44] Yongchao Mou et al. «GSE86534: Gene expression profiling of paired eutopic and ectopic endometrial tissues in women with endometriosis». In: *NCBI Gene Expression Omnibus* (2017). Accessed: 2025-09-10. URL: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE86534>.
- [45] Lin Zhao, Mei Wang, Qian Sun et al. «Differential expression of microRNAs in endometriosis». In: *Journal of Reproductive Biology* 12.3 (2010), pp. 123–130. DOI: 10.1007/jrb.2010.12.3.123. URL: <https://doi.org/10.1007/jrb.2010.12.3.123>.
- [46] Ulla M. Kärkkäinen, Rami Tolvanen, Tanja Kuiri-Hänninen et al. «Genomic evidence of Y chromosome microchimerism in the endometrium during endometriosis and in cases of infertility». In: *Human Reproduction* 34.9 (2019), pp. 1861–1869. DOI: 10.1093/humrep/dez132. URL: <https://doi.org/10.1093/humrep/dez132>.
- [47] Richard O. Burney et al. *GSE51981: Gene expression analysis of endometrial tissue from women with and without endometriosis*. NCBI Gene Expression Omnibus, accession GSE51981. 2014. URL: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51981>.
- [48] Qing Liu et al. «Machine learning-based integrated identification of predictive combined diagnostic biomarkers for endometriosis». In: *BMC Genomics* 23.1 (2022), p. 435.

- DOI: 10.1186/s12864-022-08682-7. URL: <https://doi.org/10.1186/s12864-022-08682-7>.
- [49] M. Gabriel, V. Fey e T. Heinosalo. «A relational database to identify differentially expressed genes in the endometrium and endometriosis lesions». In: *Scientific Data* 7.1 (2020), p. 284. DOI: 10.1038/s41597-020-00623-x. URL: <https://doi.org/10.1038/s41597-020-00623-x>.
- [50] Chittaranjan Andrade. «Z Scores, Standard Scores, and Composite Test Scores Explained». In: *Journal of Clinical Psychopharmacology* 42.1 (2022), pp. 1–5. DOI: 10.1097/JCP.0000000000001476. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8826187/>.
- [51] R. González-Ramos et al. «Involvement of the nuclear factor- κ B pathway in the pathogenesis of endometriosis». In: *Fertility and Sterility* 94.6 (nov. 2010), pp. 1985–1994. ISSN: 0015-0282. DOI: 10.1016/j.fertnstert.2010.01.013. URL: [https://www.fertstert.org/article/S0015-0282\(10\)00073-7/fulltext](https://www.fertstert.org/article/S0015-0282(10)00073-7/fulltext).
- [52] Luyang Zhao et al. «Identification of global transcriptome abnormalities and potential biomarkers in eutopic endometria of women with endometriosis: A preliminary study». In: *Biomedical Reports* 6.6 (mag. 2017), pp. 654–662. DOI: 10.3892/br.2017.902. URL: <https://doi.org/10.3892/br.2017.902>.
- [53] Petra A. B. Klemmt e Anna Starzinski-Powitz. «Molecular and cellular pathogenesis of endometriosis». In: *Current Women's Health Reviews* 14.2 (2018), pp. 106–116. DOI: 10.2174/1573404813666170807121538. URL: <https://pubmed.ncbi.nlm.nih.gov/29951447/>.

- [54] Yeon Jean Cho e Heung Yeol Kim. «Oxidative stress and endometriosis». In: *Kosin Medical Journal* 33.2 (2018), pp. 135–140. DOI: 10.7180/kmj.2018.33.2.135. URL: <https://doi.org/10.7180/kmj.2018.33.2.135>.
- [55] T. Harada, T. Iwabe e N. Terakawa. «Apoptosis in human endometrium and endometriosis». In: *Human Reproduction Update* 10.1 (2004), pp. 29–38. ISSN: 1355-4786. DOI: 10.1093/humupd/dmh004. URL: <https://doi.org/10.1093/humupd/dmh004>.