

UNIVERSITÀ DEGLI STUDI DEL SANNIO

Dipartimento di Scienze e Tecnologie

Corso di Laurea in Biotecnologie

Tesi di Laurea in
Bioinformatica

Trascrittomica spaziale

Effetti del trattamento con eme sull'espressione
genica nel tessuto cerebrale murino in funzione della
distanza dal sito di iniezione

Relatore:

Ch.mo Prof.
Francesco Napolitano

Candidata:

Felicità Pia Masone
Matr. 171002962

Anno accademico:

2021/2022

*La materia fluisce di luogo in luogo
e si coagula momentaneamente a formare noi.*

*Qualunque cosa siamo, quindi,
non siamo la materia di cui siamo fatti.*

- Steve Grand

Indice

Introduzione	1
1 Contesto biologico	5
1.1 Circolazione cerebrale	5
1.1.1 Circolo cerebrale anteriore	6
1.1.2 Circolo cerebrale posteriore	6
1.2 Ictus	7
1.2.1 Cause e fattori di rischio	7
1.3 Ictus emorragico intraparenchimale	8
1.3.1 Lesioni cerebrali secondarie	8
1.3.1.1 Il ruolo dell'eme nei SBI	9
1.4 Meccanismi molecolari indotti dall'eme	9
1.4.0.1 Aptoglobina ed emopessina	11
1.5 L'uso del trascrittoma come <i>readout</i>	12
2 Trascrittomica Spaziale	13
2.1 Genoma e trascrittoma	13
2.2 Trascrittomica: lo studio del trascrittoma	15
2.2.1 RNAseq workflow	16
2.2.2 Bulk, single-cell e spatial RNAseq	17
2.3 Analisi di trascrittomica spaziale	18
2.3.1 10x Visium Genomics	19
2.4 Pipeline di analisi dei dati di TS	21
2.4.1 La classe SpatialExperiment	21
2.4.2 Step di pre-elaborazione	21
2.4.3 Quality Control e normalizzazione	23
2.4.4 Batch correction	24
2.4.5 Riduzione della dimensionalità	25
2.4.6 Espressione genica differenziale	25
2.4.7 Spatially Variable Genes	26
2.5 L'uso della trascrittomica nell'ambito della Drug Discovery . . .	27

3	RGDIST: un nuovo metodo per il drug discovery con la trascrittomica spaziale	30
3.1	Dataset	30
3.2	RGDIST	31
3.3	Confronto tra metodi	32
3.4	Analisi dei risultati	35
3.4.1	Geni significativi	35
	Il33	37
	Gjb1	37
	Cdkn1a	37
	Il11	38
	Fosl1	38
	Gdf15	38
	Procr	39
	Aldh3b2	39
	Gm9732	39
3.4.2	Analisi di arricchimento delle pathway	40
	Conclusioni	43
	Bibliografia	45

Introduzione

L'informatica e la biologia molecolare sono due discipline scientifiche che sembrano agli antipodi, ma che negli ultimi decenni hanno trovato un punto di incontro nella bioinformatica. Quest'ultima, inoltre, si relaziona anche ad altre branche come la matematica e la statistica [1], con l'obiettivo di sviluppare strumenti e risorse che aiutino nell'analisi e nell'interpretazione dei dati in modo significativo. La bioinformatica utilizza algoritmi e tecniche per una vasta gamma di applicazioni, tra cui l'analisi di sequenze biologiche, lo studio della struttura e della funzione delle proteine, l'analisi delle interazioni molecolari all'interno di cellule e tessuti, la progettazione di nuovi farmaci e di terapie personalizzate in base ai profili genetici dei pazienti [2].

Per comprendere in maniera sempre più accurata la complessità dei meccanismi molecolari, è necessario analizzarli su diversi livelli; i quali, definiti nel loro insieme "scienze omiche", con l'avvento e lo sviluppo delle tecnologie di sequenziamento sono diventati fondamentali [3]. Tali livelli si suddividono in:

1. genoma, l'insieme del patrimonio genetico che caratterizza ogni organismo vivente;
2. trascrittoma, il prodotto iniziale dell'espressione del genoma. È l'insieme di tutti i trascritti di RNA, compresi quelli codificanti e non codificanti, la cui informazione biologica è richiesta dalla cellula in un determinato momento;
3. proteoma, il risultato finale dell'espressione del genoma che deriva dal processo di traduzione degli RNA messaggeri. Tale prodotto rappresenta il repertorio delle proteine, il quale determina la natura delle reazioni biochimiche che la cellula è in grado di svolgere [4];
4. epigenoma, che racchiude tutti i fenomeni di regolazione genica che non provocano alterazioni della sequenza nucleotidica, come la metilazione della citosina sul quinto residuo carbonioso (5mC) o in genere sui dinucleotidi citosina-guanina (CpG) e le modifiche sulle code amino-terminali delle proteine istoniche (metilazioni, acetilazioni, fosforilazioni, ubiquitilazioni e sumoilazioni) [5];
5. metaboloma, l'insieme di tutti i piccoli metaboliti presenti in una cellula che formano una grande rete di reazioni metaboliche, in cui i prodotti di una sono i substrati per le successive [6].

Le scienze omiche sono interconnesse e richiedono l'uso di tecnologie e strumenti bioinformatici avanzati per l'analisi e l'interpretazione dei dati.

La trascrittomica esamina i livelli di RNA sia da un punto di vista qualitativo che da un punto di vista quantitativo cioè analizza sia la presenza che l'abbondanza dei trascritti [7]. L'analisi del trascrittoma è uno step cruciale per gli studi di espressione genica ed è possibile grazie a tecniche di sequenziamento ad high-throughput come l'RNAseq. Quest'ultimo è basato sul next-generation sequencing (NGS), una tecnologia di sequenziamento del DNA su larga scala che consente di ottenere informazioni dettagliate a livello dell'intero trascrittoma con risoluzione fino alla singola coppia di basi [8, 9]. La tecnica di RNA-seq, in generale, consiste nell'isolamento dell'RNA totale da tessuti o cellule di interesse, seguito dalla costruzione di librerie di DNA complementare (cDNA), le quali vengono poi sequenziate per le successive analisi. L'RNAseq è fondamentale per analizzare risposte cellulari a stimoli fisiologici, effetti di perturbazioni sperimentali su geni e specifiche pathway o malfunzionamento della regolazione genica in condizioni patologiche [10]. L'analisi dell'espressione genica differenziale (DGE) è una delle applicazioni più comuni dei dati di RNA-seq. Questo processo consente di individuare i geni espressi in modo differenziale in due o più condizioni biologiche, con l'obiettivo di identificare i geni che presentano un'espressione alterata dipendente dalla condizione. Un esempio è l'analisi di cellule trattate rispetto a cellule non trattate [11].

Nonostante tutte le cellule del nostro corpo condividano pressoché lo stesso genotipo, il trascrittoma in ciascuna cellula riflette l'attività solo di un sottoinsieme di geni, infatti, i diversi tipi di cellule nel nostro corpo esprimono un trascrittoma unico e perciò l'RNAseq convenzionale eseguito sull'intera popolazione cellulare può fornire solo il segnale di espressione medio. I recenti progressi tecnologici hanno permesso di profilare il trascrittoma in singole cellule e si parla in questo caso di *single-cell RNA sequencing* (scRNAseq). Gli step di analisi sono analoghi all'RNAseq, ma prevedono un passaggio preliminare che consiste nell'isolamento delle singole cellule e un successivo step di identificazione del tipo di cellula grazie all'uso di marker cellulari noti [12].

Il scRNAseq ha permesso di estendere lo studio del trascrittoma alle singole cellule, ma per sua natura questa tecnica perde l'informazione spaziale relativa alla posizione delle cellule all'interno del tessuto. La trascrittomica spaziale (TS) è una tecnologia innovativa che consente l'analisi dell'espressione genica mantenendo integra l'informazione spaziale [13]. Esistono diversi approcci per questo tipo di analisi che differiscono in termini di risoluzione e di *multiplexing*,

in generale le analisi sul trascrittoma si basano su tecniche di ibridazione *in situ* mediante fluorescenza o sequenziamento [14]. Quest’ultima prende il nome di *spatial RNA sequencing* [15]. L’analisi dei dati di TS esplora tipicamente le differenze interne tra le regioni spaziali di un tessuto. Ad esempio, è possibile cercare geni che presentano dei pattern di espressione spaziale, chiamati *spatially variable genes* (SVG, [16]). Nel caso dell’analisi differenziale, i dataset di spRNAseq sono tipicamente analizzati singolarmente per estrarre i geni rilevanti da ciascuna condizione e in seguito i diversi risultati vengono messi a confronto. Tuttavia, informazioni spaziali potenzialmente utili che potrebbero essere sfruttate in un contesto di TS, come il sito di iniezione di un trattamento, in genere vengono ignorate.

L’obiettivo di questa tesi è di esaminare gli effetti di un trattamento sull’espressione genica all’interno di un tessuto biologico, prendendo in considerazione la distanza dal sito in cui viene somministrato. Si propone RGDIST (Responsive Genes based on Distance from Injection Site), un metodo che prende in esame le informazioni relative al sito di iniezione di un trattamento, al fine di identificare i geni la cui espressione varia in funzione della distanza, che sono dunque influenzati, positivamente o negativamente, dal trattamento. A tal proposito è stato analizzato un dataset di TS ottenuto dopo iniezioni stereotassiche di eme ed emopessina nel cervello di topo [17]. Molte malattie cerebrali come l’ictus, l’ipertensione cronica, l’angiopatia amiloide, le malformazioni vascolari e i tumori cerebrali possono causare emorragie intracraniche [18], in seguito alle quali l’eme viene rilasciato dall’emoglobina liberata dalle cellule e può causare lesioni cerebrali secondarie nell’area del sanguinamento.

Partendo da tali premesse, l’elaborato è articolato in tre capitoli: il primo capitolo presenta un quadro sul contesto biologico del dataset preso in esame e fornisce una panoramica sui meccanismi molecolari coinvolti nell’ictus, esplorando le cause della patologia e i fattori di rischio, con un’attenzione particolare ai danni secondari causati dalle componenti del sangue. Studi in vitro del trascrittoma possono rivelarsi molto utili per comprendere i processi molecolari che seguono un’emorragia intracerebrale; pertanto il secondo capitolo segue il percorso dal genoma all’analisi dei dati di espressione genica, partendo dalla definizione di trascritto per arrivare a una comprensione più approfondita delle pipeline bioinformatiche. Grazie all’integrazione di informazioni spaziali nella profilazione molecolare, è possibile caratterizzare meglio l’espressione genica in un contesto morfologico e valutare gli effetti di trattamenti farmacologici o con sostanze estranee. Il terzo capitolo illustra il metodo RGDIST e i risultati otte-

nuti dall'applicazione di questo metodo, con particolare attenzione su alcuni geni ritenuti significativi. Le analisi sono state eseguite con il software R, versione 4.2.2.

Capitolo 1

Contesto biologico

Al fine di comprendere meglio il contesto biologico del dataset preso in analisi questo capitolo offre una panoramica sulla circolazione cerebrale e sui meccanismi molecolari coinvolti nelle patologie associate, in particolare l'ictus. Verranno esplorate le cause dell'ictus e i fattori di rischio, con una particolare attenzione ai danni secondari causati da alcune componenti del sangue.

1.1 Circolazione cerebrale

La vascolarizzazione dell'encefalo (Figura 1.1) si realizza attraverso il circolo cerebrale arterioso, costituito dal circolo anteriore, che apporta circa il 70% del sangue e dal circolo posteriore, che fornisce circa il 30% del sangue. La vascolarizzazione cerebrale è definita di tipo terminale poiché oltre un certo livello non esistono anastomosi efficienti. Ciò significa che, nel caso di una lesione a carico di un ramo terminale, la zona da esso vascolarizzata non riceve più alcun apporto di sangue e va incontro a danni spesso irreversibili. Questa possibilità ha portato all'evoluzione di un sistema di circolazione di mutuo soccorso, noto come Poligono di Willis (Figura 1.2). Questo sistema di anastomosi arteriose si

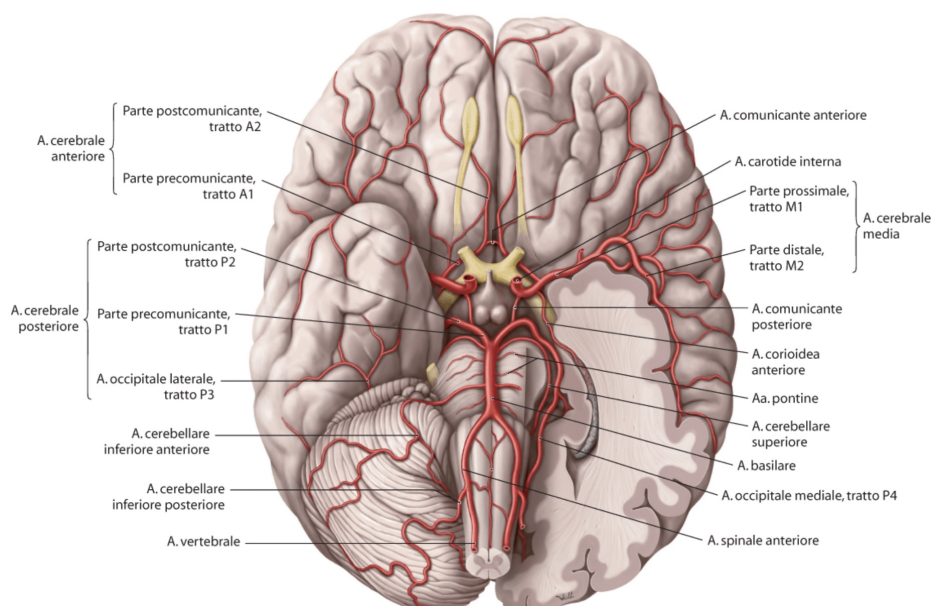


Figura 1.1: Veduta inferiore dell'encefalo che mostra la distribuzione delle arterie.

sviluppa alla base dell'encefalo tra il circolo anteriore e posteriore e garantisce una distribuzione uniforme del flusso sanguigno.

1.1.1 Circolo cerebrale anteriore

Il circolo cerebrale anteriore è formato dalle arterie carotidi interne, le quali entrano nel cranio attraverso i canali carotici delle ossa temporali e risalgono fino al nervo ottico, dove ognuna si divide in tre rami: arteria oftalmica, che vascolarizza l'occhio, arteria cerebrale anteriore, che irrori i lobi frontali e parietali dell'encefalo e arteria cerebrale media, destinata al mesencefalo e alle superfici laterali degli emisferi cerebrali. Da quest'ultima si dipartono numerose ramificazioni, note come vasi perforanti, che vanno a vascolarizzare le zone più profonde dell'encefalo ossia i nuclei della base e il talamo (Figura 1.2).

1.1.2 Circolo cerebrale posteriore

Il circolo cerebrale posteriore è formato dal sistema delle arterie vertebrali, che hanno origine dalle arterie succlavie. Le arterie vertebrali risalgono lungo i fori trasversi delle vertebre cervicali e entrano nel cranio attraverso il forame magno, dove si uniscono per formare l'arteria basilare. Quest'ultima si estende sulla superficie ventrale dell'encefalo, passando lungo il ponte, e si dirama diverse volte prima di dividersi nelle arterie cerebrali posteriori, che a loro volta si suddividono nelle arterie comunicanti posteriori.

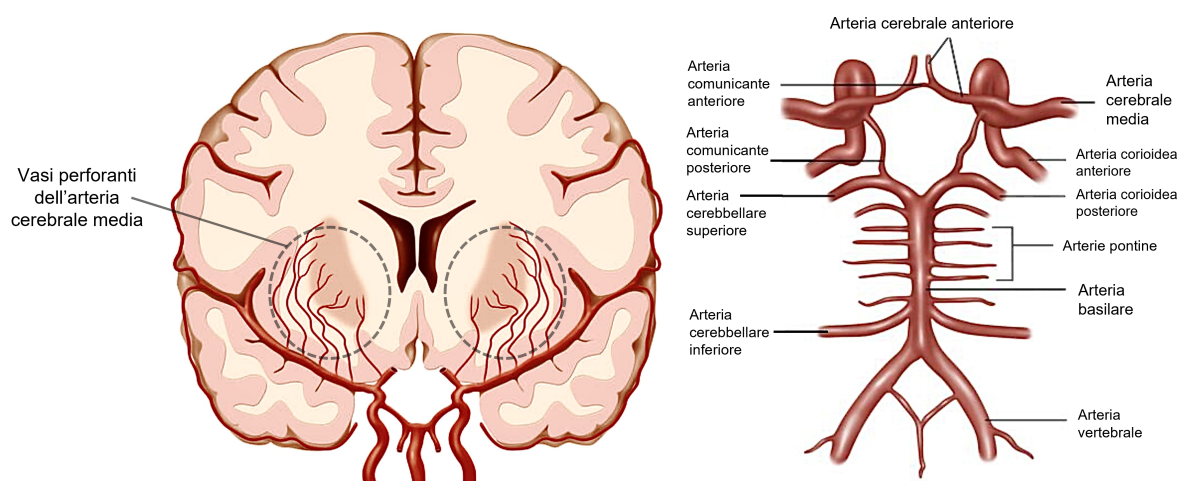


Figura 1.2: Sezione coronale dell'encefalo che mostra i vasi perforanti. A destra, Poligono di Willis.

1.2 Ictus

L'ictus (o *stroke*), dal latino “colpo”, è un evento cerebrovascolare causato dall'improvvisa occlusione o rottura di un vaso con conseguente danno al tessuto cerebrale e rappresenta la seconda causa di morte nel mondo, la seconda causa di demenza e una delle principali cause di invalidità [19]. Si tratta di una categoria non omogenea che comprende (Figura 1.3):

1. ictus ischemico (IS), si verifica quando un'arteria che irrorava l'encefalo viene ostruita dalla formazione di una placca aterosclerotica, da un coagulo di sangue che si forma sopra la placca stessa oppure da un coagulo di sangue proveniente da un altro distretto vascolare. Circa l'85% di tutti gli *stroke* è ischemico;
2. ictus emorragico da emorragia intraparenchimale (ICH), si verifica quando un'arteria situata nell'encefalo si rompe, provocando così un'emorragia intracerebrale non traumatica. Questa condizione rappresenta circa il 10% degli *stroke*;
3. ictus emorragico da emorragia subaracnoidea (SAH), si verifica quando il sanguinamento è localizzato tra il cervello e le meningi. La causa più frequente è il trauma in quanto la sua forza può causare lo stiramento delle pareti dei vasi sanguigni, che si rompono e generano sanguinamento all'interno dell'aracnoide, diffondendosi poi in tutto il cervello. Rappresenta circa il 5% di tutti gli ictus [18].

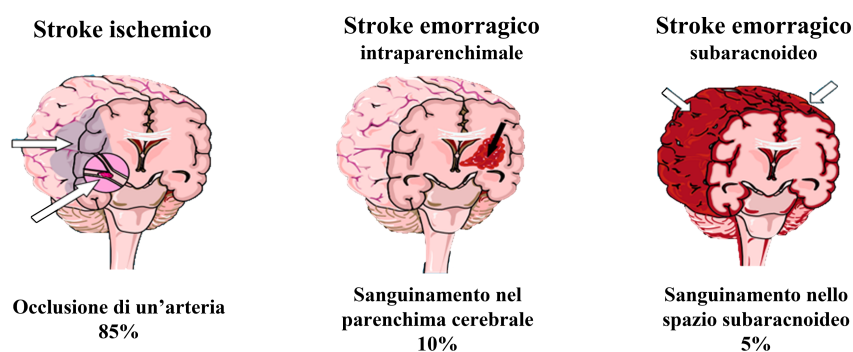


Figura 1.3: Illustrazione delle tre principali categorie di stroke.

1.2.1 Cause e fattori di rischio

La possibilità di sviluppare un ictus aumenta con l'avanzare dell'età e raddoppia dopo i 55 anni, sia negli uomini che nelle donne. I fattori di rischio possono essere suddivisi in non modificabili, che includono età, sesso, etnia, attacco ischemico transitorio (TIA) e caratteristiche ereditarie, e modificabili, che

includono ipertensione, diabete, fibrillazione atriale, iperlipidemia, abuso di alcol e sostanze, fumo e scarsa attività fisica [20].

1.3 Ictus emorragico intraparenchimale

Lo *stroke* emorragico intraparenchimale è un sottotipo di ictus grave, con alti tassi di morbidità e mortalità, causato dalla rottura spontanea di un'arteria cerebrale o dalla conversione emorragica di un ictus ischemico con conseguente emorragia intracranica (ICH) [21]. Si distinguono due tipi di emorragie intraparenchimali:

- tipica, emorragia profonda dovuta alla rottura dei vasi perforanti. Si localizza specificamente intorno ai ventricoli, nei gangli basali (50%), nei lobi cerebrali (10%-20%), nel talamo (15%), a livello del ponte e del tronco encefalico (10%-20%) o nel cervelletto (10%) [22];
- atipica, qualsiasi emorragia che si localizza in posizione diversa da quella tipica. Può essere dovuta alla rottura di vasi propri di una neoplasia cerebrale o ad una malformazione vascolare [18].

Il danno (o lesione) primario si verifica da pochi minuti a poche ore dall'inizio del sanguinamento e rappresenta la principale causa di mortalità dell'ictus emorragico; è dovuto principalmente all'elevata ipertensione endocranica: il sangue fuoriesce dal vaso in maniera rapida e violenta provocando la compressione e il danneggiamento del tessuto neuronale circostante [23]. In seguito all'espansione dell'emorragia vascolare, si verifica una serie di eventi patologici che provoca i cosiddetti danni cerebrali secondari [22].

1.3.1 Lesioni cerebrali secondarie

Se il paziente sopravvive all'ictus, l'ematoma formatosi nel parenchima cerebrale si espande e innesca una serie di eventi avversi che causano lesioni cerebrali secondarie (SBI) e gravi deficit neurologici [23]. I danni secondari sono causati da diversi eventi patologici quali infiammazione, rottura della barriera emato-encefalica (BEE), edema, sintesi eccessiva di radicali liberi come le specie reattive dell'ossigeno (ROS), eccitotossicità indotta dal glutammato e rilascio di emoglobina e ferro dalle emazie [22]. L'espansione dell'ematoma, che si verifica come conseguenza dell'effetto di massa della lesione primaria, è una delle principali cause di deterioramento neurologico e di esito clinico sfavorevole dopo un'ICH;

le varie componenti del sangue fuoriescono dai vasi e continuano a diffondersi nel tessuto cerebrale provocando lesioni secondarie sempre più estese [24].

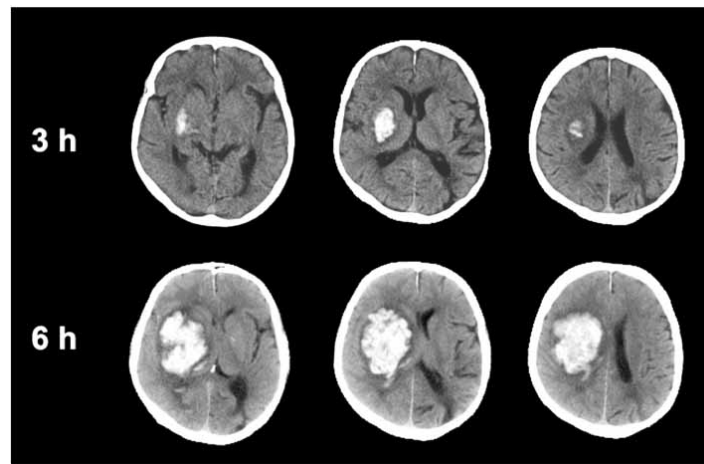


Figura 1.4: Espansione dell'ematoma. La tomografia computerizzata seriale, 3 e 6 ore dopo l'esordio dei sintomi, rivela una significativa espansione di un'emorragia intracerebrale primaria localizzata nel putamen destro di un paziente di 76 anni affetto da ipertensione acuta. Immagine di *Bornes and Butcher* [25].

1.3.1.1 Il ruolo dell'eme nei SBI

A livello dell'ematoma, i globuli rossi si decompongono lentamente e riversano il loro contenuto nell'area circostante. Il 95% delle proteine dell'eritrocita è rappresentato dall'emoglobina (Hb), una molecola tetrameric deputata al trasporto dell'ossigeno; ciascuna delle quattro subunità è legata covalentemente all'eme, un gruppo prostetico costituito da una molecola di protoporfirina IX coordinata con uno ione ferro Fe^{2+} , a formare emoglobina ferrosa (Hb^{2+}). Gli eritrociti contengono emoglobina a una concentrazione di 15 mM di eme-equivalenti, quantità che supera di oltre 1000 volte la soglia di tossicità [17]. Nella fase iniziale, i tetrameri di Hb (64 kDa) si dissociano in piccoli dimeri (32 kDa); inoltre, in un ambiente ossidativo, l' Hb^{2+} viene rapidamente convertita in emoglobina ferrica (Hb^{3+}) o nella forma ferrile (Hb^{4+}). Entrambe queste forme sono notevolmente reattive e, diffondendo nei tessuti, provocano gravi conseguenze come vasospasmi, vasocostrizione, ischemia e insufficienza cardiaca [26, 27]. L'eme libero da solo può indurre molti degli effetti dannosi osservati per l'emoglobina libera da cellule.

1.4 Meccanismi molecolari indotti dall'eme

L'emoglobina libera e i suoi componenti contribuiscono al danno tissutale attraverso diversi meccanismi che includono:

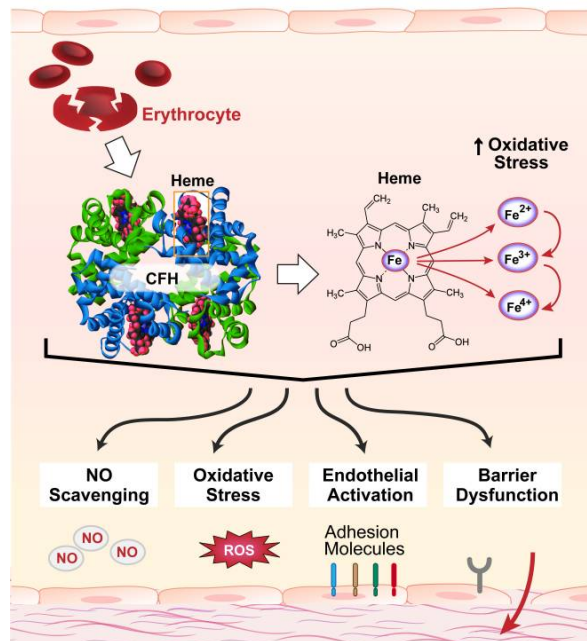


Figura 1.5: Struttura tridimensionale dell'Hb e del gruppo eme. L'Hb e l'eme rilasciati dai globuli rossi danneggiati sono responsabili di lesioni endoteliali attraverso deplezione di NO, stress ossidativo, attivazione endoteliale e disfunzione della barriera endoteliale. CFH = Cell Free Hemoglobin [27].

- deplezione dell'ossido nitrico (NO), l'emoglobina ha la capacità di legare l'NO, che è essenziale per l'omeostasi funzionale e metabolica del cervello [26]. La rimozione dell'NO, che causa inizialmente vasocostrizione, scatena una serie di eventi endoteliali che includono la riduzione del flusso sanguigno, l'aumento dell'infiammazione, l'attivazione delle piastrine e danni agli organi;
- stress ossidativo, in quanto l'emoglobina reagisce con lipidi e lipoproteine formando radicali liberi reattivi e ROS che vanno ad attivare o danneggiare altre cellule, tra cui leucociti e globuli rossi, rendendo l'ambiente sempre più ossidativo. Si instaura così un meccanismo di retroazione positiva che ha come risultato l'estensione del danno. A livello molecolare, i ROS inducono l'attivazione della via di segnalazione di NLRP3, un inflammasoma in grado di attivare le citochine pro-infiammatorie [28], e alla produzione dell'IL-1 β , un mediatore leucocitico endogeno. Per contrastare questo stress ossidativo, le cellule attivano meccanismi molecolari che portano alla produzione di molecole citoprotettive, tra cui l'eme ossigenasi 1 (HO-1), la catena pesante della ferritina (FHC) e il fattore indotto dall'ipossia (HIF) [27];
- attivazione endoteliale, avviene attraverso la pathway TLR4/MyD88, che a sua volta attiva i sistemi di segnalazione MAPK e NF- κ B, nonché il sistema del complemento. Ciò porta all'upregulation di citochine pro-infiammatorie, molecole di adesione e mediatori della coagulazione [29];
- disfunzione della barriera endoteliale, caratterizzata dalla perdita di integri-

tà delle giunzioni cellula-cellula (aderenti e serrate) e dalla riorganizzazione dell'actina del citoscheletro che forma le cosiddette fibre di stress. I meccanismi di segnalazione che sembrano essere coinvolti includono l'*up-regulation* delle MAPK, Src, Rho, MMPs e la segnalazione di pMLC. In diversi esperimenti *in vivo*, l'iniezione di emoglobina ha provocato la rottura della barriera emato-encefalica [30].

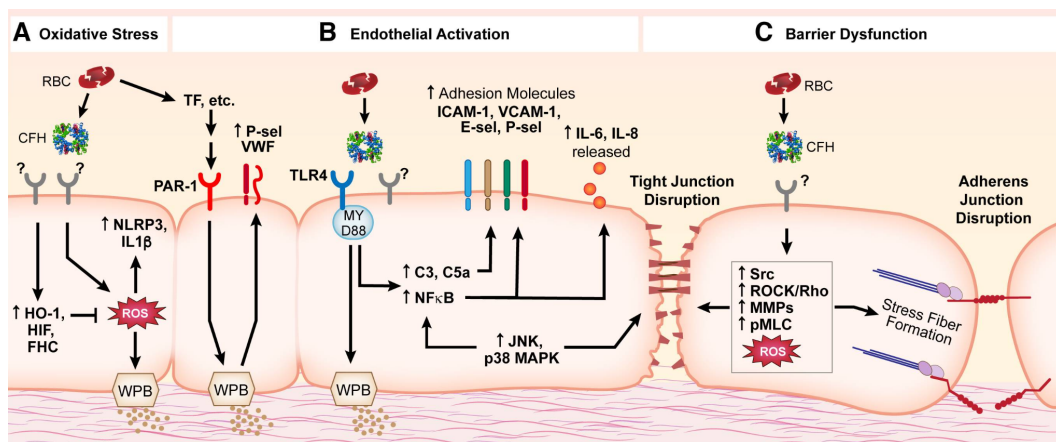


Figura 1.6: Potenziali meccanismi di segnalazione coinvolti nel danno endoteliale mediato da emoglobina libera [27].

1.4.0.1 Aptoglobina ed emopessina

Due proteine endogene, l'aptoglobina e l'emopessina, agiscono per combattere la tossicità provocata dall'emoglobina e dall'eme dopo un'ICH [23]. L'aptoglobina (Hp) è una proteina presente in grandi quantità nel plasma sanguigno, prodotta principalmente dagli epatociti e in minor misura da altri tessuti. La sua funzione principale è quella di neutralizzare l'emoglobina tossica in caso di emolisi intravascolare. Questo avviene attraverso la formazione di complessi tra l'Hp e l'emoglobina, che vengono poi rimossi dalla circolazione da una sottoclasse specializzata di macrofagi che esprimono il recettore CD163. In condizioni normali, l'Hp svolge un'importante funzione protettiva contro la tossicità dell'emoglobina. Attraverso saggi di immunofluorescenza, lo studio di Zhao et al. ha dimostrato che gli oligodendrociti nel cervello di topo producono Hp dopo un'emorragia intracerebrale e la sua espressione aumenta significativamente intorno all'ematoma entro poche ore dall'inizio dell'ICH. Poiché gli oligodendrociti sono abbondanti nella sostanza bianca e sono presenti in tutta la sostanza grigia, la produzione locale di Hp da parte di queste cellule rappresenta probabilmente un importante meccanismo endogeno di protezione del cervello contro la tossicità dell'Hb extravascolare. Oltre al sistema di rimozione Haptoglobina-emoglobina-Hb//CD163, esiste un sistema indipendente che aiuta a rimuovere i prodotti di

rottura dell'emoglobina, l'eme e il ferro. Questo sistema coinvolge l'emopessina (Hpx), una glicoproteina presente nel plasma sanguigno sintetizzata principalmente dagli epatociti. L'Hpx si lega all'eme e forma complessi Hpx-eme, che vengono facilmente endocitati dai macrofagi che esprimono il recettore CD91. In condizioni fisiologiche, il CD91 ha un ruolo nel riciclaggio del ferro. Tuttavia, in risposta all'emolisi extravascolare nei tessuti colpiti dall'ematoma, il sistema Hpx-eme/CD91 può facilitare la rimozione dell'eme pro-ossidativo da parte della microglia e dei macrofagi [32].

1.5 L'uso del trascrittoma come *readout*

Identificare i meccanismi specifici di danno cellulare e tissutale causati dall'emoglobina può essere difficile a causa dei suoi molteplici stati di ossidazione e del potenziale rilascio di eme e ferro. Questi fattori aumentano la complessità dell'analisi, specialmente *in vivo* [27]. Gli studi *in vitro* del trascrittoma possono rivelarsi molto utili per comprendere i processi molecolari che seguono un'ICH. Inoltre, grazie alla trascrittomica spaziale è possibile valutare l'estensione dei danni a livello molecolare e analizzare la distribuzione in specifiche regioni del tessuto dei geni e delle proteine coinvolte, permettendo anche l'identificazione dei tipi cellulari. Grazie a questi studi, sarà possibile migliorare la conoscenza dei processi molecolari associati all'ICH e sviluppare nuove terapie funzionali per il trattamento della patologia.

Nel prossimo capitolo, verrà fornita una spiegazione dettagliata del trascrittoma e delle relative analisi, partendo da una breve descrizione del genoma per una migliore comprensione del concetto di trascritto.

Capitolo 2

Trascrittomica Spaziale

Studi in vitro del trascrittoma possono rivelarsi molto utili per comprendere i processi molecolari che seguono un'ICH. Inoltre, con la trascrittomica spaziale è possibile valutare l'estensione dei danni a livello molecolare e analizzare la distribuzione in specifiche regioni del tessuto dei geni e delle proteine coinvolte. Questo capitolo segue il percorso dal genoma all'analisi dei dati di espressione genica, partendo dalla definizione di trascritto per arrivare a una comprensione più approfondita delle *pipeline* bioinformatiche.

2.1 Genoma e trascrittoma

Il genoma è l'insieme completo del materiale genetico di un organismo, che contiene tutte le informazioni necessarie per lo sviluppo, la crescita e il funzionamento dello stesso. Per la maggior parte degli organismi è costituito dal DNA (acido deossiribonucleico), un polimero organico formato da nucleotidi, composti da un gruppo fosfato, uno zucchero (deossiribosio) e una delle quattro basi azotate: adenina, timina, citosina e guanina. La specificità genetica del DNA risiede nelle sue sequenze nucleotidiche, che sono uniche per ogni individuo. Infatti, nonostante il DNA abbia un alfabeto ridotto a quattro lettere, ha un numero di sequenze possibili estremamente elevato, consentendo combinazioni pressoché infinite (4^N , N = numero di lettere nella sequenza). Il genoma

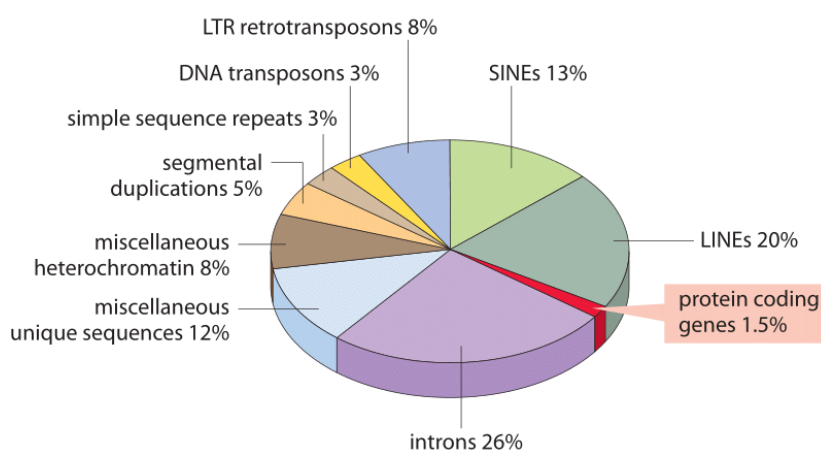


Figura 2.1: Grafico a torta che mostra la composizione del genoma umano in termini di tipologia di sequenze, [33]. La porzione codificante è solo l'1.5%, tutto il resto è costituito da sequenze non codificanti.

si suddivide in due porzioni principali: una codificante, costituita dai geni che contengono informazioni per la sintesi di proteine, e una non codificante. Le regioni non codificanti concorrono alla complessità genetica e fenotipica degli organismi, rappresentando la maggior parte del DNA negli eucarioti, infatti, anche all'interno dei geni, ci sono sequenze non codificanti chiamate introni. Il DNA è usato dalla cellula come modello per la sintesi di RNA, una molecola simile al DNA con alcune differenze strutturali: contiene il ribosio al posto del deossiribosio, l'uracile al posto della timina e si presenta come singolo filamento. Il processo di sintesi di RNA è chiamato trascrizione e inizialmente si pensava che solo il DNA codificante venisse trascritto, mentre la porzione non codificante era considerata inutile. I progressi nelle tecniche di biologia molecolare hanno permesso di caratterizzare meglio questa porzione ed è emerso che il DNA non codificante è fondamentale per il corretto funzionamento delle cellule, infatti anche esso viene trascritto. L'RNA è una molecola molto versatile ed eterogenea sia per quanto riguarda la composizione e le strutture che può formare, sia per le funzioni che svolge all'interno della cellula, come mostrato nella Figura 2.2:

- RNA messaggero, mRNA, contiene la sequenza nucleotidica che codifica per una specifica proteina e viene utilizzato come modello per la sintesi proteica durante la traduzione;
- RNA transfer, tRNA, molecola relativamente piccola che funge da adattatore molecolare tra l'mRNA e l'amminoacido corrispondente. Il tRNA lega specificamente un amminoacido e lo trasferisce al sito di sintesi proteica sul ribosoma;
- RNA ribosomiale, rRNA, costituisce la maggior parte della massa degli RNA cellulari e rappresenta il componente strutturale del ribosoma, il complesso macromolecolare responsabile della sintesi proteica;
- piccoli RNA nucleari, snRNA, localizzati nel nucleo delle cellule eucariotiche dove si associano con proteine specifiche a formare complessi chiamati snRNP (*small nuclear ribonucleoprotein*), coinvolti nei processi di maturazione dell'mRNA, come lo *splicing*, e nella regolazione dell'espressione genica;
- piccoli RNA nucleolari, snoRNA, si trovano principalmente nel nucleolo delle cellule eucariotiche e sono coinvolti nelle modificazioni chimiche durante il processo di maturazione degli rRNA e dei tRNA;
- microRNA e piccoli RNA interferenti, miRNA e siRNA, piccoli frammenti di RNA che regolano l'espressione genica a livello post-trascrizionale. Si distin-

guono sulla base della loro origine, possono legare l'mRNA e impedirne la traduzione o degradarlo, influenzando così la quantità di proteina prodotta;

- lunghi RNA non-codificanti, lncRNA, svolgono funzioni regolatorie dell'espressione genica, come la modulazione della trascrizione e della stabilità dell'mRNA.

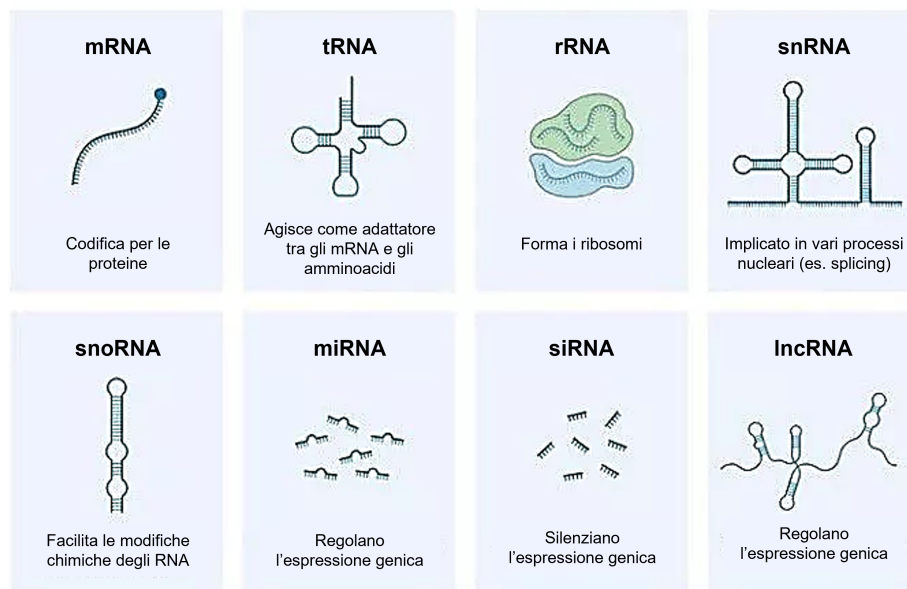


Figura 2.2: Illustrazione dei diversi tipi di RNA che si trovano nelle cellule.

L'insieme di tutti gli RNA sintetizzati da una cellula prende il nome di *trascrittoma* e con il termine *trascritto* si indica una qualsiasi molecola di RNA. Talvolta questi termini fanno riferimento solo al sottoinsieme di RNA codificanti, gli mRNA, in quanto da essi deriva il proteoma della cellula. Una caratteristica unica degli mRNA è la loro coda poli-adenilata, poli-A, che ne permette l'isolamento.

2.2 Trascrittomica: lo studio del trascrittoma

La trascrittomica mira all'analisi sia qualitativa che quantitativa di interi profili di espressione. Tali studi sono fondamentali per individuare gli elementi funzionali del genoma, approfondire la complessità dell'espressione genica e per chiarire i meccanismi alla base dello sviluppo e delle malattie [7]. Nel corso degli anni sono state sviluppate varie tecniche per le analisi di espressione genica che differiscono nei metodi e nella risoluzione dei dati generati, tra cui:

- PCR quantitativa (qPCR), prevede la conversione dell'RNA in DNA complementare (cDNA) e l'amplificazione mediante PCR utilizzando sonde fluorescenti specifiche per il gene di interesse. La quantità di cDNA amplificato

viene poi misurata in tempo reale durante la PCR, consentendo di quantificare l'espressione dei geni di interesse;

- ibridazione *in situ*, utilizza delle sonde a fluorescenza che si legano in modo selettivo per complementarità ad alcune specifiche regioni del cromosoma. Per individuare il sito di legame tra sonda e cromosoma si utilizzano tecniche di microscopia a fluorescenza;
- Microarray, permette di analizzare migliaia di geni contemporaneamente. I campioni di RNA vengono marcati con fluorocromi e ibridati su una matrice contenente sonde complementari ai geni di interesse. I segnali di fluorescenza sono poi rilevati e analizzati;
- RNA sequencing (RNAseq), permette di quantificare l'espressione genica a livello di singolo trascritto. L'RNA viene estratto e convertito in cDNA, che viene poi sequenziato utilizzando una piattaforma ad *high throughput*. I dati di sequenziamento vengono poi analizzati per identificare i trascritti e quantificarne l'espressione.

L'RNAseq è diventata rapidamente la tecnologia preferita per lo studio dell'espressione genica grazie alla sua maggiore risoluzione e alla maggiore riproducibilità rispetto alle altre. Consente una migliore identificazione degli eventi di *splicing* alternativo, di nuovi geni e trascritti; inoltre è in grado di rilevare anche bassi livelli di espressione di trascritti rari o di geni poco espressi, il che aumenta la sensibilità e la specificità dei dati generati [34].

2.2.1 RNAseq workflow

Un prerequisito fondamentale per uno studio di RNAseq è quello di definire un design sperimentale. Questo comporta innanzitutto la scelta del tipo di libreria, *single-end* (SE) o *paired-end* (PE) a seconda dello scopo delle analisi. La tecnica SE prevede il sequenziamento da una sola estremità, mentre la PE consente di sequenziare entrambe le estremità di un frammento. In generale, le brevi *read* SE sono sufficienti per studi sui livelli di espressione genica in organismi i cui genomi sono ben annotati, mentre le più lunghe PE sono preferibili per caratterizzare trascrittomi scarsamente annotati. Altro requisito indispensabile è la scelta della profondità di sequenziamento, o *library size*, cioè il numero di letture sequenziate per un determinato campione; anche in questo caso dipende dall'obiettivo dell'esperimento. Infine è importante valutare il numero di repliche, sia tecniche che biologiche, appropriate per il sistema in studio, per garantire che l'acquisizione dei dati non sia contaminata da *bias* inutili [35].

Una volta definito il progetto sperimentale, si procede seguendo i protocolli di analisi che prevedono quattro step principali (Figura 2.3): il preprocessamento dei campioni, la preparazione delle librerie, il sequenziamento e l'analisi dati.

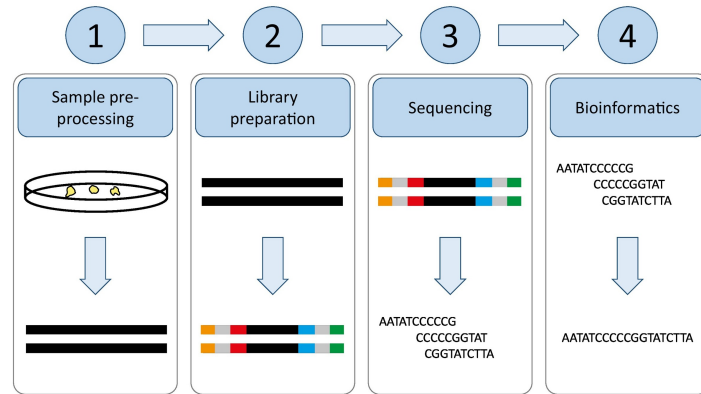


Figura 2.3: Schema generale del *workflow* per l'RNAseq.

In primo luogo, l'acido nucleico deve essere estratto dal campione che si intende analizzare; uno step cruciale di questa fase è la rimozione dell'rRNA che, vista la sua abbondanza, ridurrebbe il segnale d'interesse. L'RNA è una molecola relativamente poco stabile, per questo viene convertita in cDNA attraverso il processo di *trascrizione inversa*. La preparazione della libreria è realizzata con sequenze di oligonucleotidi definite in modo da essere compatibili con la tecnica di sequenziamento applicata e produce una popolazione di frammenti di cDNA di lunghezza stabilita. Successivamente, ha luogo il sequenziamento vero e proprio con il rispettivo sistema e i dati così ottenuti vengono elaborati seguendo pipeline di analisi bioinformatiche per ottenere le informazioni pertinenti [36].

2.2.2 Bulk, single-cell e spatial RNAseq

L'analisi globale del trascrittoma di un campione è anche chiamata *bulk*-RNAseq ed è un metodo efficace per lo studio dell'espressione genica che però ha delle limitazioni, come la perdita di informazioni sulle differenze tra diversi tipi cellulari all'interno di un campione misto. Per superare questo problema, i progressi tecnologici hanno permesso lo sviluppo del cosiddetto *single-cell* RNAseq, grazie al quale è possibile profilare il trascrittoma di singole cellule; infatti, all'interno di uno stesso tessuto o in un ambiente tumorale la popolazione cellulare è eterogenea. Le analisi in questo caso prevedono uno step in più che consiste nell'isolamento delle singole cellule. Tuttavia, anche il scRNAseq ha i suoi limiti, infatti esso perde una parte di informazioni che è molto utile per la comprensione della funzionalità cellulare e dei cambiamenti patologici ossia la posizione delle cellule all'interno del tessuto. A partire da questi limiti è stato

sviluppato lo *spatial*-RNAseq che combina i punti di forza dell'analisi di RNAseq e dell'ibridazione *in situ*, fornendo dati sull'intero trascrittoma con informazioni spaziali [15].

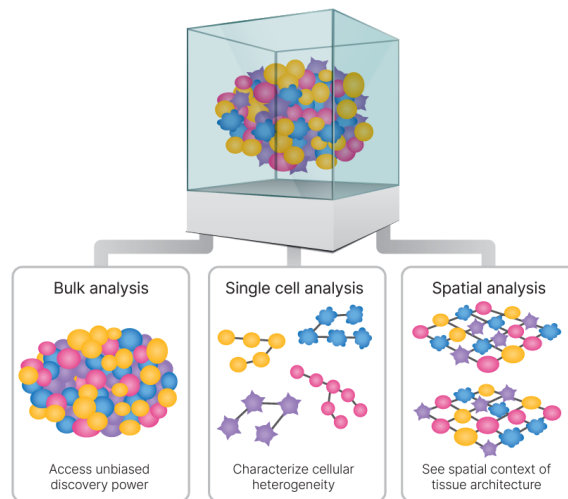


Figura 2.4: Schema delle tre categorie di RNAseq.

2.3 Analisi di trascrittomica spaziale

La trascrittomica spaziale fornisce una visione, finora inaccessibile, dell'intero trascrittoma in un contesto morfologico. Negli ultimi anni si è verificata una rapida crescita dei metodi per la profilazione molecolare spaziale, che variano per risoluzione, scala e *multiplexing*. Tali tecnologie si dividono in cinque categorie a seconda del metodo di acquisizione delle informazioni spaziali [13]:

1. regione di interesse (ROI), consiste nella selezione e nell'isolamento di specifiche porzioni di tessuto attraverso microdissezione fisica o ottica. Le ROI isolate possono essere analizzate con microarray o RNAseq, o dissociate in singole cellule per il scRNAseq;
2. ibridazione in situ a fluorescenza, FISH, tecnica di ibridazione altamente multiplexata che utilizza sonde marcate con fluorofori per rilevare i trascritti di RNA a livello di singola molecola;
3. *in situ sequencing* (ISS), utilizza sonde marcate con fluorofori che legano specifiche sequenze di RNA. Le sonde sono quindi sequenziate *in situ*, cioè all'interno del tessuto, con marcatori fluorescenti. Possono essere sia mirati, utilizzando barcodes genici specifici, sia non mirati, utilizzando brevi frammenti di cDNA;

4. NGS con barcode spaziali, prevede l'utilizzo di slide con degli array contenenti migliaia di barcode, identificatori molecolari unici (UMI) e oligonucleotidi poli-T che permettono di catturare i trascritti poli-adenilati e sequenziarli;
5. metodi che non richiedono conoscenze spaziali a priori, ma che permettono di preservare informazioni utili a ricostruire una mappa spaziale; un esempio è la DNA *microscopy*.

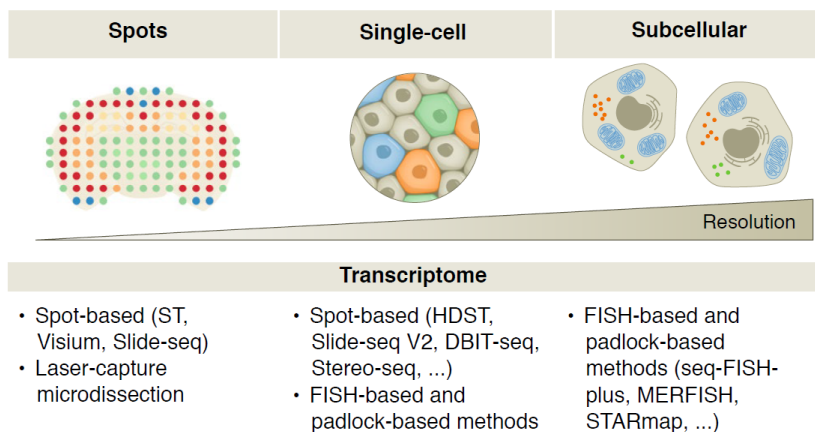


Figura 2.5: Risoluzione delle diverse tecniche di trascrittomiche spaziale [14].

Tecniche basate sulla FISH restituiscono dati con un'altissima risoluzione, fino alla singola molecola, mentre quelle basate su spot hanno una risoluzione più bassa che copre decine di micrometri e include diverse cellule (Figura 2.5). I metodi differiscono anche per il numero di *features* molecolari acquisite: da decine nella FISH a centinaia o migliaia nei metodi basati su sonde specializzate (MERFISH o SeqFISH) fino a decine di migliaia nella trascrittomiche spaziale basata su spot [14]. Ciascun metodo presenta dei limiti e dei punti di forza ed è perciò opportuno utilizzare queste tecnologie in base alle diverse esigenze di analisi e di caratterizzazione del campione studiato.

2.3.1 10x Visium Genomics

Visium Spatial Gene Expression è una piattaforma *10x Genomics* per la profilazione molecolare basata su NGS che permette di mappare l'espressione genica tramite vetrini (o *slide*, Figura 2.6) sui quali è presente una griglia di punti, *spot*, delimitata da una cornice detta *fiducial frame*. Ciascuno spot contiene milioni di oligonucleotidi di cattura con codice a barre spaziale, illustrati in Figura 2.6, che si legano all'mRNA presente nel tessuto.

I campioni su cui è possibile eseguire queste analisi sono di due tipi (Figura 2.7): tessuti congelati poco dopo l'estrazione detti *fresh frozen tissue* e tessuti fissati in formalina e inclusi paraffina o FFPE. Nei tessuti congelati, l'mRNA

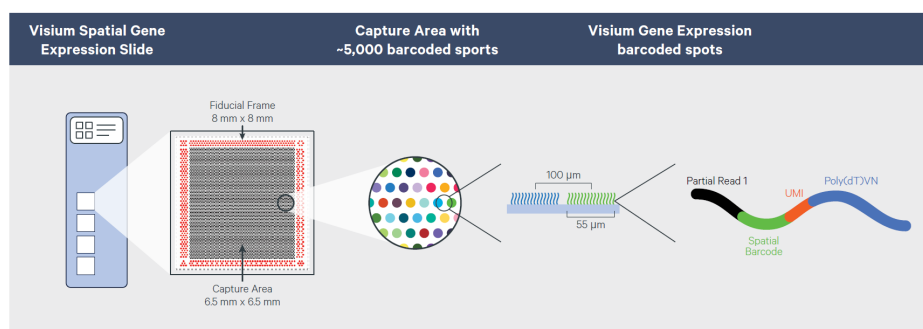


Figura 2.6: Composizione del vetrino 10x Visium. Ciascun vetrino contiene due o quattro aree di cattura con circa 5.000 spot, contenenti milioni di oligonucleotidi di cattura con codice a barre spaziale. L'mRNA tissutale rilasciato si lega a questi oligo, consentendo la cattura delle informazioni sull'espressione genica.

viene rilasciato dalle sezioni di tessuto trattate e si lega agli oligonucleotidi di cattura in una posizione vicina a quella originale. La trascrizione inversa avviene sulla slide con il tessuto ancora in posizione; nei campioni FFPE, delle sonde si legano all'mRNA e il tessuto viene permeabilizzato per rilasciare le coppie ibridate, che si legano agli oligo di cattura sul vetrino. Le coppie di sonde vengono poi estese per generare la libreria. Indipendentemente dal tipo di campione, le librerie di cDNA conservano le informazioni spaziali grazie all'incorporazione dei codici a barre e vengono mappate su un punto specifico del vetrino. I dati di espressione genica sono quindi sovrapposti all'immagine ad alta risoluzione della sezione di tessuto ottenuta al microscopio, consentendo la visualizzazione dell'espressione di uno o più mRNA all'interno della morfologia del campione. Quindi, il workflow per l'analisi con 10x Visium ha step analoghi al bulkRNAseq (Figura 2.3): preparazione del campione, costruzione delle librerie, sequenziamento e analisi dati, con la differenza che in questo caso gli oligonucleotidi sono marcati con barcode spaziali e il sequenziamento avviene direttamente sul vetrino [37].

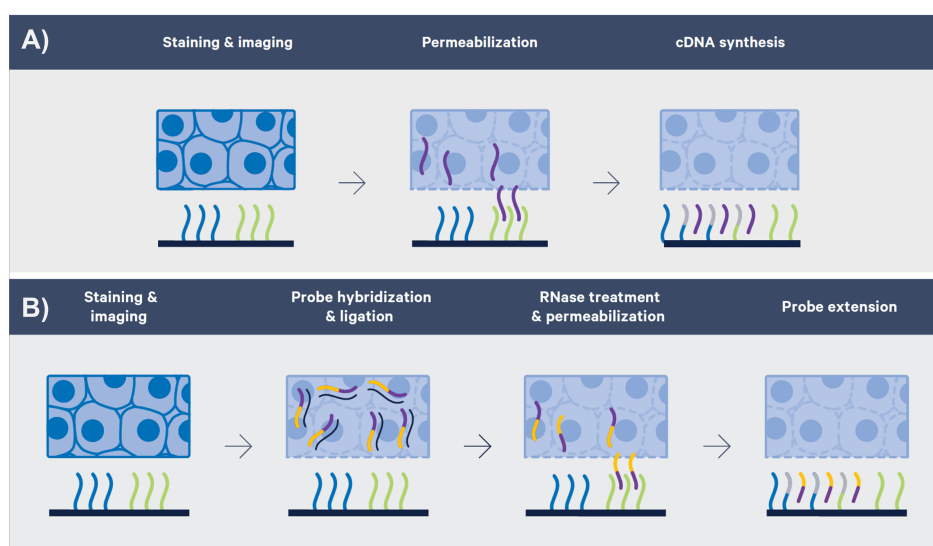


Figura 2.7: Schema della costruzione di librerie con 10x Visium per tessuti congelati (A) e FFPE (B)

2.4 Pipeline di analisi dei dati di TS

I dati di trascrittomica spaziale generati con la piattaforma 10x Visium possono essere analizzati nell'ambiente Rstudio utilizzando diversi pacchetti di *Bioconductor*, una piattaforma *open source* per le analisi dei dati biologici, sviluppata per il linguaggio di programmazione R. Attualmente non è stata ancora definita una pipeline standard per le analisi di TS; tuttavia, trattandosi di dati di RNAseq, è possibile seguire gli stessi passaggi previsti per il scRNAseq [38], assumendo che gli spot possano essere trattati come equivalenti alle cellule.

2.4.1 La classe `SpatialExperiment`

`SpatialExperiment` [39] è la principale struttura per la memorizzazione e la manipolazione dei dati di TS, che si basa sulla classe più generale `SingleCellExperiment` [38] per i dati di scRNAseq, con integrazioni per la memorizzazione di informazioni spaziali, come le coordinate spaziali e i file di immagine. La struttura di un oggetto `SpatialExperiment`, illustrata in Figura 2.8, è la seguente:

- `assays`, slot contenente i conteggi di espressione. Per i dati di TS basati su spot come 10x Visium Genomics, viene utilizzato un singolo saggio denominato `counts`. Per i dati basati sulle molecole come seqFISH, vengono utilizzati due saggi, `counts` e `molecules`;
- `rowData`, contiene informazioni sulle *features*, cioè sui geni;
- `colData`, include informazioni sugli spot o sulle cellule, compresi i metadati non spaziali e spaziali;
- `spatialCoords`, slot per le coordinate spaziali;
- `imgData`, per i dati delle immagini.

2.4.2 Step di pre-elaborazione

Per preparare i dati grezzi di 10x Visium e caricarli in R, sono necessarie alcune fasi di pre-elaborazione che richiedono l'uso di strumenti esterni a R e Bioconductor: `VitoSeg`, `Space Ranger` e `Loupe Browser`, due software per il pre-processamento forniti da 10x Genomics.

- `VitoSeg` [40] è uno strumento utilizzato per l'elaborazione, l'analisi e la visualizzazione interattiva di immagini istologiche ad alta risoluzione in MATLAB. Tra i suoi output, `VitoSeg` fornisce una stima del numero di cellule per spot, che può essere utilizzata come input nelle analisi successive;

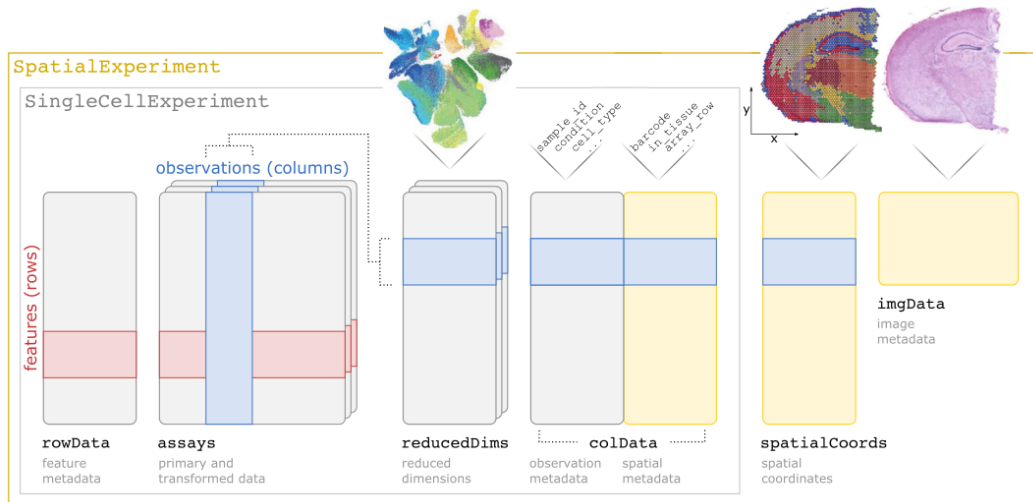


Figura 2.8: Panoramica della struttura della classe `SpatialExperiment` [39]

- *Loupe Browser* è un'applicazione desktop di 10x Genomics che consente di visualizzare i dati di espressione genica senza dover scrivere codice;
- *Space Ranger* è un insieme di pipeline di analisi per l'elaborazione dei dati di sequenza 10x Visium Genomics con immagini al microscopio ad alta risoluzione dei tessuti.

Space Ranger consente di mappare le letture del trascrittoma con l'immagine al microscopio del tessuto da cui sono state ottenute. Include cinque pipeline: *mkfastq*, *count*, *aggr*, *targeted-compare* e *targeted-depth*. Per le analisi dei dati di 10x Visium si utilizza la pipeline *spaceranger count*. Questa richiede diversi input: l'immagine al microscopio, i file FASTQ¹, il numero seriale del vetrino e il trascrittoma di riferimento del campione in esame. Space Ranger esegue l'allineamento delle sequenze, il rilevamento e l'allineamento dei tessuti utilizzando la fiducial frame e il conteggio dei codici a barre. La pipeline *spaceranger count* produce numerosi output, riportati in Figura 2.9. Tra questi, i più importanti sono la matrice geni-spot, i cui dati si trovano nella cartella *raw_feature_bc_matrix*, e la cartella *spatial*, che contiene i file spaziali. Questi vengono combinati con i dati generati da *VistoSeg* in un oggetto R `SpatialExperiment` per le analisi a valle, utilizzando la funzione `SpatialExperiment::read10xVisium()`. Inoltre, Space Ranger restituisce anche la matrice gene-spot filtrata (*filtered_feature_bc_matrix*) ossia priva dei dati

¹formato basato su testo per memorizzare sia una sequenza biologica che i relativi punteggi di qualità. Per brevità, sia la lettera della sequenza che il punteggio di qualità sono codificati con un singolo carattere ASCII.

relativi agli spot non coperti da tessuto.

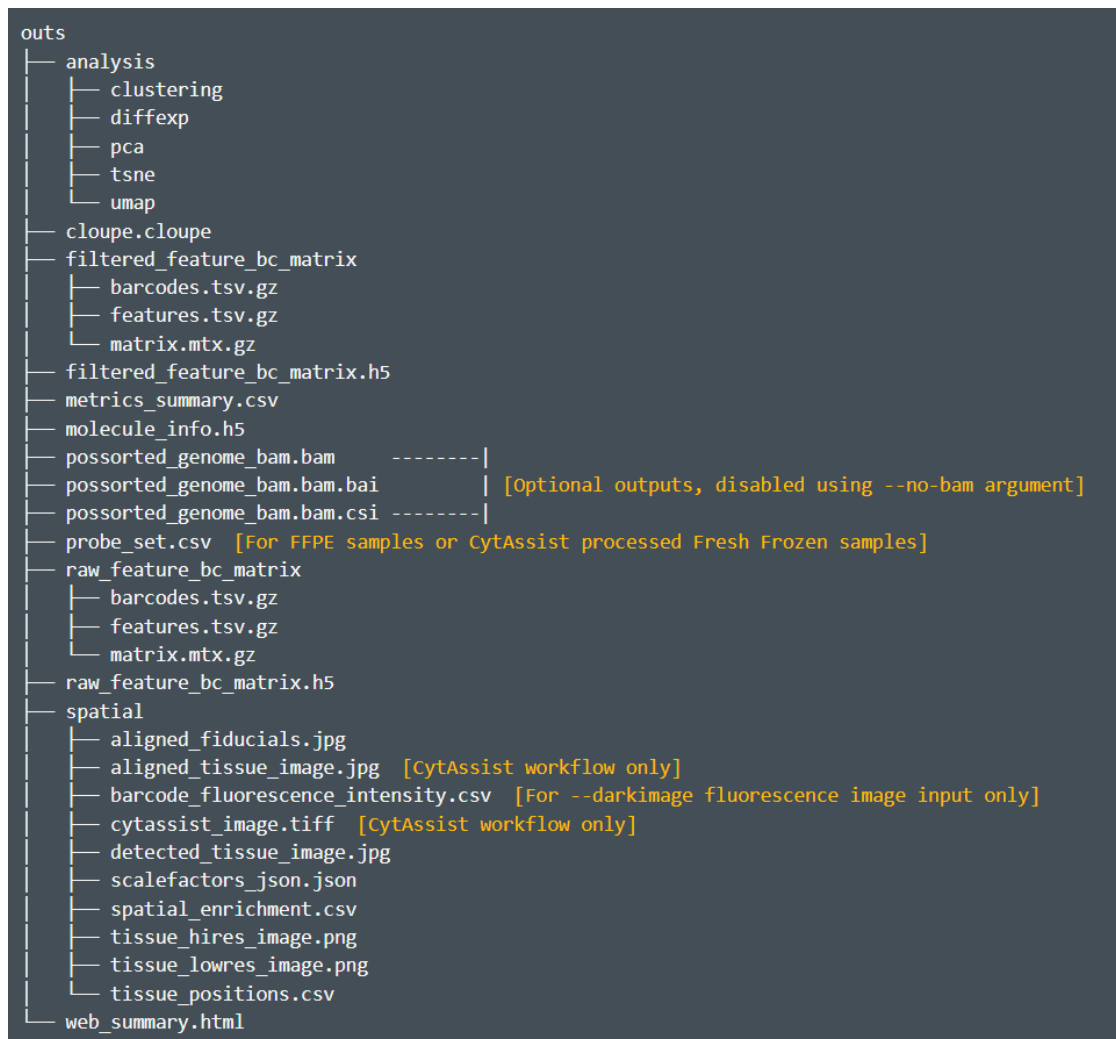


Figura 2.9: Gerarchia degli output ottenuti dalla pipeline spaceranger count

2.4.3 Quality Control e normalizzazione

Le procedure di controllo di qualità (QC) a livello di spot sono fondamentali per garantire l'affidabilità e l'accuratezza dei dati ottenuti dalle analisi di sequenziamento, al fine di evitare problemi durante le analisi a valle. Gli spot di bassa qualità possono verificarsi a causa di problemi durante la preparazione della libreria o di altre procedure sperimentali, come ad esempio la presenza di grandi percentuali di cellule morte o la bassa efficienza di cattura dell'mRNA e ciò può influire negativamente sulla qualità dei dati di sequenziamento, portando a risultati errati o non riproducibili. Ad esempio, gli spot problematici che non vengono rimossi potrebbero apparire come cluster separati, che potrebbero essere identificati erroneamente come tipi di cellule distinte. Per questi motivi, è importante che vengano rimossi prima di ulteriori analisi. Ci sono diverse carat-

teristiche che possono essere utilizzate per identificare gli spot di bassa qualità, tra cui:

- dimensione della libreria (cioè conteggi totali di UMI per spot), valori bassi possono indicare uno scarso tasso di cattura degli mRNA;
- numero di *features* espresse (cioè numero di geni con conteggi UMI non nulli per spot), anche in questo caso valori bassi possono indicare scarsa efficienza nella cattura degli mRNA;
- percentuale di geni mitocondriali, se alta è indice di danno cellulare;
- numero di cellule per spot, se elevato può indicare problemi durante la segmentazione cellulare.

Per la normalizzazione si applicano metodi sviluppati per i dati scRNA-seq, implementati nei pacchetti **scater** [41] e **scran** [42], partendo dal presupposto che questi metodi possono essere applicati ai dati SRT trattando gli spot come equivalenti alle cellule. Il metodo più semplice e indicato per valutare la qualità dei dati è quello basato sulla *library size*, poiché può essere facilmente applicato ai dati TS. Altri metodi popolari per i dati scRNAseq, sono più difficili da giustificare nel contesto dei dati TS basati su spot poiché (i) gli spot possono contenere più tipi di cellule e (ii) i dataset possono contenere più campioni. In alcuni dataset, può essere necessario applicare procedure di controllo della qualità o di filtraggio a livello dei geni, oltre che degli spot: alcuni geni, come quelli mitocondriali, possono essere biologicamente irrilevanti per le analisi successive e quindi devono essere rimossi.

2.4.4 Batch correction

Per aumentare l'affidabilità dei dati, gli esperimenti condotti su un campione vengono replicati più volte nelle stesse condizioni sperimentali, generando le cosiddette *repliche tecniche*. Inoltre, se l'esperimento viene eseguito su campioni biologici diversi, si parla di *repliche biologiche*. Nonostante queste repliche siano realizzate nelle stesse condizioni sperimentali, presentano un certo grado di variabilità che può essere eliminato tramite l'uso di metodi statistici. L'applicazione di tecniche di normalizzazione e pre-elaborazione dei dati tende a ridurre la variabilità tecnica, ma spesso rimangono livelli elevati di eterogeneità sistematica che potrebbe oscurare i fenomeni biologici in studio [43]. Gran parte di questa variabilità sottostante è associata all'elaborazione dei campioni in gruppi (repliche) e pertanto viene comunemente definita effetto *batch*, letteralmente "effetto dei lotti". Esistono due classi generali di metodi di correzione batch:

quelli che utilizzano la modellistica lineare quando i batch sono noti o presunti (ad es. **ComBat** [44]) e quelli che cercano di identificare e controllare i potenziali effetti batch (ad es. analisi di variabili surrogate **SVA** [45]). Poiché il suo impatto può essere significativo, è importante comprendere tale eterogeneità come parte integrante del processo di elaborazione ed esplorazione dei dati genomici.

2.4.5 Riduzione della dimensionalità

La riduzione della dimensionalità è la trasformazione dei dati da uno spazio ad alta dimensionalità in uno spazio a bassa dimensionalità in modo che la rappresentazione a bassa dimensionalità conservi alcune proprietà significative dei dati originali, idealmente vicine alla loro dimensione intrinseca. Lavorare in spazi ad alta dimensionalità può essere scomodo poiché i dati grezzi spesso sono sparsi e l'analisi dei dati è computazionalmente difficile da gestire. I metodi sono comunemente divisi in approcci lineari e non lineari, ma possono anche essere divisi in selezione ed estrazione di *features*. La riduzione della dimensionalità può essere utilizzata per la riduzione del rumore, la visualizzazione dei dati, l'analisi dei cluster, o come passaggio intermedio per facilitare altre analisi [46]. I principali metodi sono:

1. *Principal component analysis* (PCA), tecnica che esegue una mappatura lineare dei dati in modo tale che la varianza dei dati nella rappresentazione a bassa dimensionalità sia massimizzata;
2. *T-distributed Stochastic Neighbor Embedding* (t-SNE), approccio non lineare utile per la visualizzazione di insiemi di dati ad alta dimensionalità. Non è raccomandata per l'uso in analisi come il clustering o la rilevazione degli outlier in quanto non preserva necessariamente bene le densità o le distanze;
3. *Uniform manifold approximation and projection* (UMAP), tecnica non lineare visualmente simile a t-SNE, ma assume che i dati siano distribuiti uniformemente su uno spazio di Riemann localmente connesso e che la metrica di Riemann sia localmente costante. UMAP è utile per la visualizzazione dei dati in due o tre dimensioni, ma può anche essere utilizzato per la riduzione della dimensionalità a fini analitici come il clustering o l'analisi di cluster [47].

2.4.6 Espressione genica differenziale

Un'applicazione tipica dell'RNAseq è l'analisi dell'espressione differenziale (DE), in cui diverse condizioni biologiche vengono confrontate tra loro con l'obiettivo di identificare l'espressione dipendente dalla condizione. Un esempio è

l'analisi di cellule trattate rispetto a cellule non trattate. Un gene viene considerato differenzialmente espresso quando la differenza o il cambiamento osservato nelle letture o nei livelli di espressione tra due condizioni sperimentali è statisticamente significativo. La distribuzione di Poisson e la distribuzione binomiale negativa sono due modelli statistici comunemente utilizzati per l'analisi di DE. La distribuzione di Poisson è caratterizzata dalla sua semplicità e dalla presenza di un solo parametro, ma ha il vincolo della varianza uguale alla media, il che può portare a falsi positivi soprattutto in presenza di repliche biologiche. D'altra parte, la distribuzione binomiale negativa ha due parametri, media e dispersione, il che consente di modellare in modo più accurato la relazione tra media e varianza. Questo modello è spesso utilizzato per gestire la variabilità biologica dei dati [48].

2.4.7 Spatially Variable Genes

L'analisi dei dati ST esplora tipicamente le differenze interne tra le regioni spaziali di un tessuto. Ad esempio, è possibile cercare i geni che presentano modelli di espressione spaziale, chiamati *Spatially Variable Genes* (SVG). Gli approcci esistenti per identificare i geni altamente variabili (HVG), utilizzati per i dati di scRNAseq convenzionali, ignorano le informazioni spaziali e quindi non misurano la variabilità spaziale, infatti gli HVG sono definiti esclusivamente sulla base di caratteristiche molecolari. Gli SVG sono definiti come geni con un modello di espressione altamente correlato a livello spaziale, che varia in base alla distribuzione spaziale di una struttura tissutale di interesse. Per classificare i geni in base all'autocorrelazione spaziale osservata si possono utilizzare misure statistiche standard come la I di Moran o la C di Geary. Nello studio sulla corteccia prefrontale di Maynard et al. è stata eseguita l'analisi della varianza (ANOVA) per testare l'espressione differenziale tra gruppi di cellule, utilizzando annotazioni cellulari predefinite o basandosi sul clustering dei campioni con alcuni metodi che incorporano informazioni spaziali. Tali metodi possono solo rilevare le variazioni dovute alle differenze tra gruppi discreti. SpatialDE [16] è un metodo per identificare gli SVG che si basa sulla regressione tramite processo gaussiano, una classe di modelli utilizzati in geostatistica. In breve, per ogni gene, SpatialDE decompone la variabilità dell'espressione in componenti spaziali e non spaziali, utilizzando due termini di effetto casuale: un termine di varianza spaziale che parametrizza la covarianza dell'espressione dei geni in base alle distanze tra i campioni e un termine di rumore che modella la variabilità

non spaziale. Il rapporto tra la varianza spiegata da queste componenti quantifica la frazione di varianza spaziale (FSV). Gli SVG significativi possono essere identificati confrontando questo modello completo con un modello privo del termine di varianza spaziale. Inoltre, SpatialDE fornisce un metodo di clustering spaziale che identifica insiemi di geni che segnalano distinti pattern di espressione spaziale. Questo consente di effettuare un'analisi automatica dell'espressione istologica (AEH), che collega la struttura del tessuto e la composizione del tipo di cellula utilizzando i pattern di espressione dei geni marker.

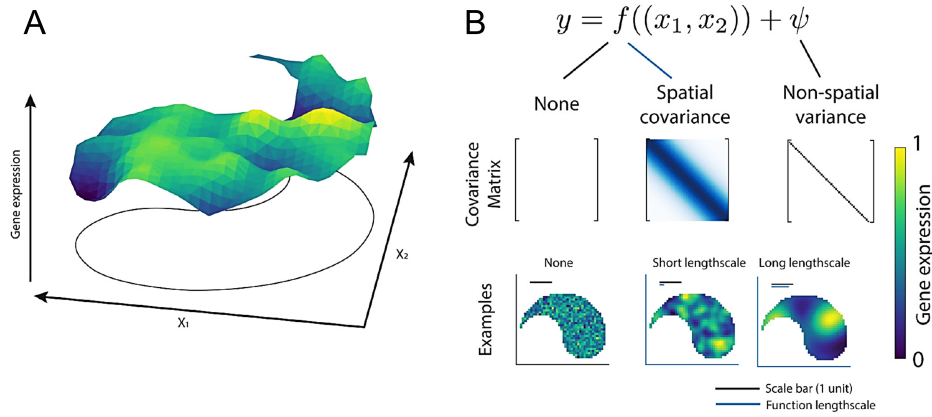


Figura 2.10: Panoramica di SpatialDE per l'identificazione dei geni variabili nello spazio [16]. A) I livelli di espressione sono misurati come funzione delle coordinate spaziali delle cellule o dei campioni. B) Il modello statistico SpatialDE suddivide la variazione dell'espressione in una componente spaziale, caratterizzata dalla covarianza spaziale, e il rumore di osservazione indipendente.

2.5 L'uso della trascrittomico nell'ambito della Drug Discovery

L'uso dei dati di RNAseq nella ricerca di nuovi farmaci è sempre più diffuso. Grazie a queste tecnologie ad high-throughput, è possibile analizzare l'espressione genica di migliaia di geni contemporaneamente, identificare nuovi marcatori biologici e identificare vie metaboliche coinvolte in patologie specifiche. Inoltre, il scRNAseq consente di analizzare il trascrittoma a livello di singola cellula, fornendo informazioni dettagliate sullo sviluppo e sulla differenziazione cellulare, nonché sulla risposta alle terapie. La trascrittomico è un'importante risorsa nella scoperta e sviluppo dei farmaci, poiché consente di identificare, già nelle prime fasi dello sviluppo, gli effetti collaterali dei composti candidati, prevenendo co-

stosi fallimenti in fase avanzata. Inoltre, l'utilizzo di *signature*² di espressione di pochi geni può sostituire un costoso test tardivo per la genotossicità, consentendo di risparmiare tempo e risorse. L'analisi dell'espressione genica può anche rivelare l'eccessiva eterogeneità dei composti candidati cioè la capacità di legarsi a molteplici bersagli molecolari, che può causare effetti collaterali indesiderati [50]. Lo screening computazionale dei farmaci è stato facilitato dalla comparsa di metodi che permettono di mappare e identificare le relazioni tra più sistemi, in particolare la *Connectivity Map* (CMap) e la *Library of Integrated Network-based Cellular Signatures* (LINCS) [51, 52]. CMap e LINCS sono database di perturbazioni farmacologiche completi e su larga scala, contenenti profili trascrittomici di numerose linee cellulari trattate con migliaia di composti chimici [53]. Il concetto di base di CMap consiste nell'utilizzare un database contenente profili di espressione genica specifici per un particolare farmaco e confrontarlo con una *signature* genica specifica della malattia. Per utilizzare CMap, è sufficiente presentare una lista di geni ritenuti rilevanti per la malattia in questione e si riceverà in output un elenco di farmaci che hanno presunta efficacia per la malattia o interi meccanismi d'azione noti. Questo strumento consente di identificare le connessioni tra farmaci, geni e malattie, con l'obiettivo generale di predire potenziali candidati terapeutici. LINCS è un database che comprende circa 5000 perturbatori genetici (come silenziamenti o sovraespressioni di singoli geni) e circa 15000 perturbatori indotti da composti chimici. Grazie alla tecnologia L1000, sviluppata dal team di CMap presso il Broad Institute [52], sono stati profilati e raccolti oltre un milione di profili di espressione genica in modo rapido, flessibile e ad alta efficienza a costi inferiori. La tecnologia L1000 misura solo l'espressione dei cosiddetti geni *landmark*³, che sono 978 in totale, mentre i valori di espressione per tutti i restanti vengono stimati da un modello computazionale che utilizza dati aggiuntivi da *Gene Expression Omnibus* (GEO). La principale sfida nel tracciare l'interazione dei farmaci in vivo attraverso i compartimenti tissutali e i tipi di cellule è la mancanza di strumenti per visualizzare le molecole di farmaco in situ. L'analisi tradizionale bulk del tessuto non

²Gruppo singolo o combinato di geni in una cellula con un modello di espressione genica univocamente caratteristico che si verifica come risultato di un processo biologico alterato o non alterato o di una condizione medica patogena.

³Gene la cui espressione è considerata informativa per caratterizzare il trascrittoma e che viene misurato direttamente nel saggio L1000. I geni *landmark* sono stati selezionati in quanto ampiamente espressi in tutto il sistema e dotati di un buon potere predittivo per ricavare l'espressione di altri geni che non sono direttamente misurati nel saggio.

fornisce informazioni spaziali e cellulari dettagliate. Tuttavia, il rapido sviluppo delle tecniche di TS sta aprendo nuove possibilità nella farmacologia risolta spazialmente, consentendo una valutazione più completa e obiettiva dell'efficacia e della tossicità dei farmaci in vivo. Ad esempio, Clearing Assisted Tissue click CHemistry (CATCH) è una tecnica che permette di visualizzare il legame covalente dei farmaci a risoluzione subcellulare [54]. Ciò consente di identificare dove e quali tipi di cellule all'interno di un organo sono legate quantitativamente da un determinato farmaco, rivelando un coinvolgimento eterogeneo del farmaco all'interno dell'organo.

La trascrittomico rappresenta, dunque, un'importante risorsa nella scoperta, sviluppo e valutazione dei farmaci. In particolare, attraverso le nuove tecnologie di profilazione spaziale è possibile valutare l'impatto dei farmaci su un tessuto a livello cellulare in modo molto preciso. Questo consente di comprendere in modo dettagliato fino a che punto i trattamenti farmacologici agiscono sul tessuto, identificando le aree in cui i farmaci si concentrano e gli effetti che questi producono sulle singole cellule. Ciò apre nuove possibilità nello sviluppo di terapie mirate, che hanno un effetto specifico su determinate cellule o tessuti, migliorando l'efficacia e la sicurezza dei trattamenti e offrendo nuove opportunità per la cura delle patologie.

Capitolo 3

RGDIST: un nuovo metodo per il drug discovery con la trascrittomica spaziale

La trascrittomica spaziale è un metodo innovativo di profilazione molecolare che consente di misurare tutta l'attività genica in un campione di tessuto e di mappare le zone in cui si verificano fenomeni di interesse. Ai dati di spRNA-seq è possibile integrare informazioni spaziali rilevanti, che altre tecnologie non supportano, ed effettuare analisi che sfruttano queste informazioni per caratterizzare meglio l'espressione genica in un contesto morfologico. Ad esempio, in un esperimento in cui i campioni sono trattati con farmaci o molecole con attività biologica, può essere utile sfruttare le coordinate di un sito di iniezione per valutare gli effetti del trattamento in funzione della distanza. Sebbene siano stati proposti diversi metodi per identificare i geni spazialmente variabili, essi non tengono conto della posizione del sito di iniezione. In questo capitolo viene presentato un metodo per valutare gli effetti trascrizionali di un trattamento molecolare basandosi sull'ipotesi che la distanza delle cellule dal sito di iniezione attraverso il tessuto influisca sulla corrispondente risposta trascrizionale. L'obiettivo principale di questa tesi è lo sviluppo di un metodo in grado di identificare i geni responsivi ad un trattamento farmacologico, ovvero quei geni che mostrano una risposta all'azione del farmaco. Questo metodo si basa sull'assunzione che questi geni debbano mostrare gli effetti del trattamento in modo più marcato nelle cellule più vicine al sito di iniezione del farmaco. A tal fine, è stato sviluppato RGDIST (*Responsive Genes based on Distance from Injection Site*), un metodo che sfrutta le coordinate del sito di iniezione per identificare tali geni.

3.1 Dataset

I dati di TS utilizzati per l'analisi sono pubblicamente disponibili sul sito web *Gene Expression Omnibus* (numero di accesso GSE182127). Il dataset è stato ottenuto dopo iniezioni stereotassiche di eme ed emopessina nel cervello di topo. Il presupposto dell'esperimento è che molte patologie cerebrali come

l'ictus, l'ipertensione cronica, l'angiopatia amiloide, le malformazioni vascolari e i tumori cerebrali possono causare emorragie intracraniche [18], in seguito alle quali l'eme viene rilasciato dall'emoglobina liberata dalle cellule e può causare lesioni cerebrali secondarie nell'area dell'emorragia. Lo studio di Buzzi et al. mirava ad analizzare spazialmente i cambiamenti di espressione genica nel cervello di topo simulando un'emorragia cerebrale intraparenchimale tipica, attraverso l'iniezione di eme nel nucleo striato, per valutare l'entità e il tipo di lesioni secondarie provocate. Inoltre, sono stati valutati gli effetti della proteina emopessina come potenziale candidato per revertire gli effetti tossici dell'eme. Il dataset comprende sette campioni di sezioni coronali:

- due controlli, uno non trattato e uno trattato con soluzione salina
- quattro trattati con concentrazioni crescenti di eme (add conc.)
- uno trattato con eme-emopessina.

Per le successive analisi, è stato selezionato il campione trattato con 10 nmol di eme (GSM5519059) poiché si ritiene che possa mostrare gli effetti trascrizionali più pronunciati. Oltre agli output della pipeline *spaceranger count*, il dataset comprende le coordinate del sito di iniezione e i cluster anatomici assegnati manualmente. Dalla successiva analisi sono stati esclusi: (i) i punti non assegnati ad alcun cluster anatomico, poiché questa informazione è necessaria nelle fasi successive dell'analisi; (ii) i geni rilevati in meno di tre punti, poiché sono necessarie almeno tre osservazioni per calcolare le correlazioni. Inoltre, nello studio di Buzzi et al. sono stati identificati gli insiemi di geni significativamente up- e down-regolati in risposta al trattamento con eme, mettendo insieme i risultati di due approcci diversi, che comprendevano l'analisi differenziale spaziale e l'analisi differenziale tra trattato e controllo, insieme all'identificazione del tipo di cellula basata sull'anatomia e sul clustering. L'insieme finale comprende un totale di 62 geni up-regolati e 6 geni down-regolati, che sono stati usati come *ground truth* per le successive analisi di confronto tra metodi. Il dataset è stato caricato con il pacchetto **SpatialExperiment** v1.8.1 nell'ambiente R e sottoposto alle fasi di *quality control* e normalizzazione descritte in precedenza.

3.2 RGDIST

È stata quindi calcolata la distanza euclidea di ogni spot dal sito di iniezione e, poiché è fornito come area, si è scelto il suo punto centrale per il calcolo; una visualizzazione di questa area è riportata in Figura 3.1.

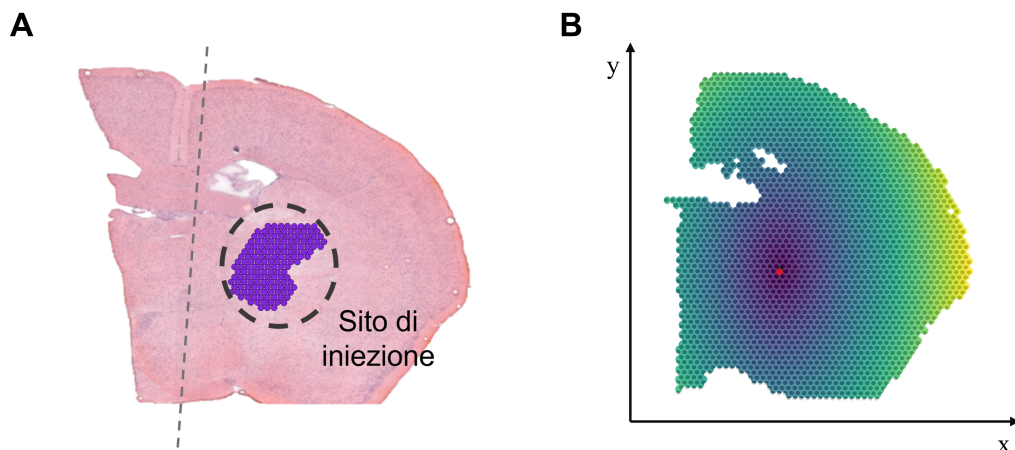


Figura 3.1: A) Area di iniezione. B) Rappresentazione degli spot e del pattern di espressione ricercato da RGDIST.

Il campione in analisi è una sezione coronale del cervello di topo, composta da diverse regioni anatomiche che presentano un certo grado di eterogeneità anche a livello trascrizionale. Per rimuovere questi effetti trascrizionali guidati dal tessuto che non dipendono dal trattamento, è stata eseguita una correzione *batch* (BC) dei dati di espressione utilizzando l'approccio `ComBat_seq` (pacchetto R `sva` v3.35.2 [44]) in cui i cluster anatomici sono stati considerati come effetti batch da correggere. La funzione `ComBat_seq` prende in input la matrice dei conteggi e restituisce una matrice simile, ma con i valori corretti per l'effetto batch richiesto. I conteggi corretti sono stati normalizzati con lo stesso metodo utilizzato per i *counts* cioè per il numero totale di trascritti all'interno di ogni spot. Per valutare qualitativamente l'effetto della correzione sui dati, sono state generate delle UMAP prima e dopo la BC, riportate in Figura 3.2. Infine, per ogni gene è stato calcolato il coefficiente di correlazione di Spearman tra la sua espressione all'interno del tessuto e la distanza degli spot dal sito di iniezione. Le analisi sono state effettuate con il software R v4.2.2, seguendo le pipeline descritte nel capitolo 2.

3.3 Confronto tra metodi

Per dimostrare la validità di RGDIST, è stata eseguita un'analisi dei *rank* classificando i geni in base ai p-value della correlazione di Spearman e tale classificazione è stata poi confrontata con quelle ottenute con altri due metodi. In primo luogo, è stato ideato un metodo *baseline* basato esclusivamente sull'espressione genica, senza informazioni di controllo né spaziali. In questo caso, è stata semplicemente utilizzata l'espressione media di ciascun gene in tutto il campione di

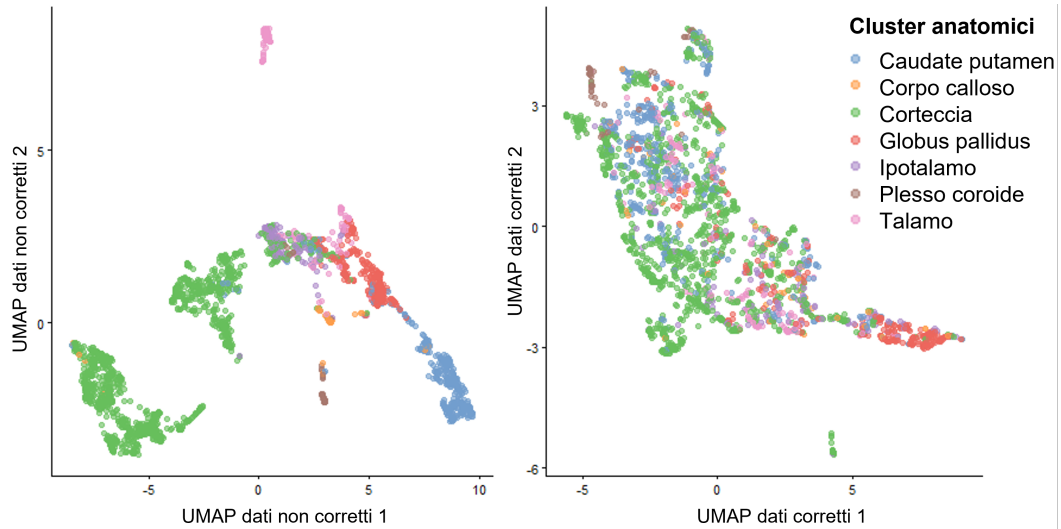


Figura 3.2: UMAP prima e dopo la Batch Correction. I colori rappresentano i cluster anatomici e si nota che dopo la correzione questa informazione è in parte persa; è ancora evidente il cluster del globus pallidus, che corrisponde alla zona di iniezione, in cui gli effetti sono più marcati.

tessuto, normalizzata per l'espressione media di tutti i geni. Come secondo metodo è stato utilizzato uno *state of art* per l'identificazione degli SVG, spatialDE [16], che fa uso solo delle informazioni spaziali del campione trattato. L'analisi SpatialDE è stata applicata seguendo la sua pipeline standard (pacchetto R spatialDE v1.4.3) e successivamente i geni sono stati classificati in base ai loro valori di Log Likelihood Ratio (LLR) in ordine decrescente. Per valutare i tre metodi è stato eseguito un confronto utilizzando come *ground truth* i geni identificati da Buzzi et al. e poi valutato il significato delle discrepanze risultanti. Il boxplot di confronto in Figura 3.3 mostra che sia RGDIST che spatialDE sono accurati nell'identificare geni significativi in risposta all'eme senza l'utilizzo di un campione di controllo. SpatialDE tende a classificare in alto i geni *ground truth* (mediana: 537,5) in base ai pattern spaziali nella loro espressione. D'altra parte, RGDIST li classifica ancora più in alto (mediana: 391), anche se diversi geni sono classificati significativamente più in basso in quanto non mostrano una tendenza nell'espressione in salita o in discesa. L'approccio che utilizza RGDIST e ComBat è risultato essere il più efficace nel trovare i geni che mostrano una risposta al trattamento. Questo perché, sebbene anche l'approccio SpatialDE sia accurato, tende a individuare pattern spaziali in modo generale, compresi quelli che dipendono dal tipo di tessuto, piuttosto che dalla presenza del trattamento. Pertanto, l'utilizzo combinato di RGDIST e ComBat consente di individuare in modo preciso e specifico i geni che mostrano una risposta al trattamento. La Figura 3.3 mostra i grafici di due geni significativi per il metodo spatialDE che presentano un pattern spaziale evidente (Myl4, posizione: 45; Tac1, posizione:

85). Considerando i modelli di espressione su tutto il campione di tessuto e la mancanza di una tendenza rispetto al sito di iniezione, l'ipotesi è che l'espressione di tali geni, ritenuti rilevanti da altri metodi, possa essere influenzata da effetti residui, dipendenti dal tipo cellulare e non dal trattamento con eme.

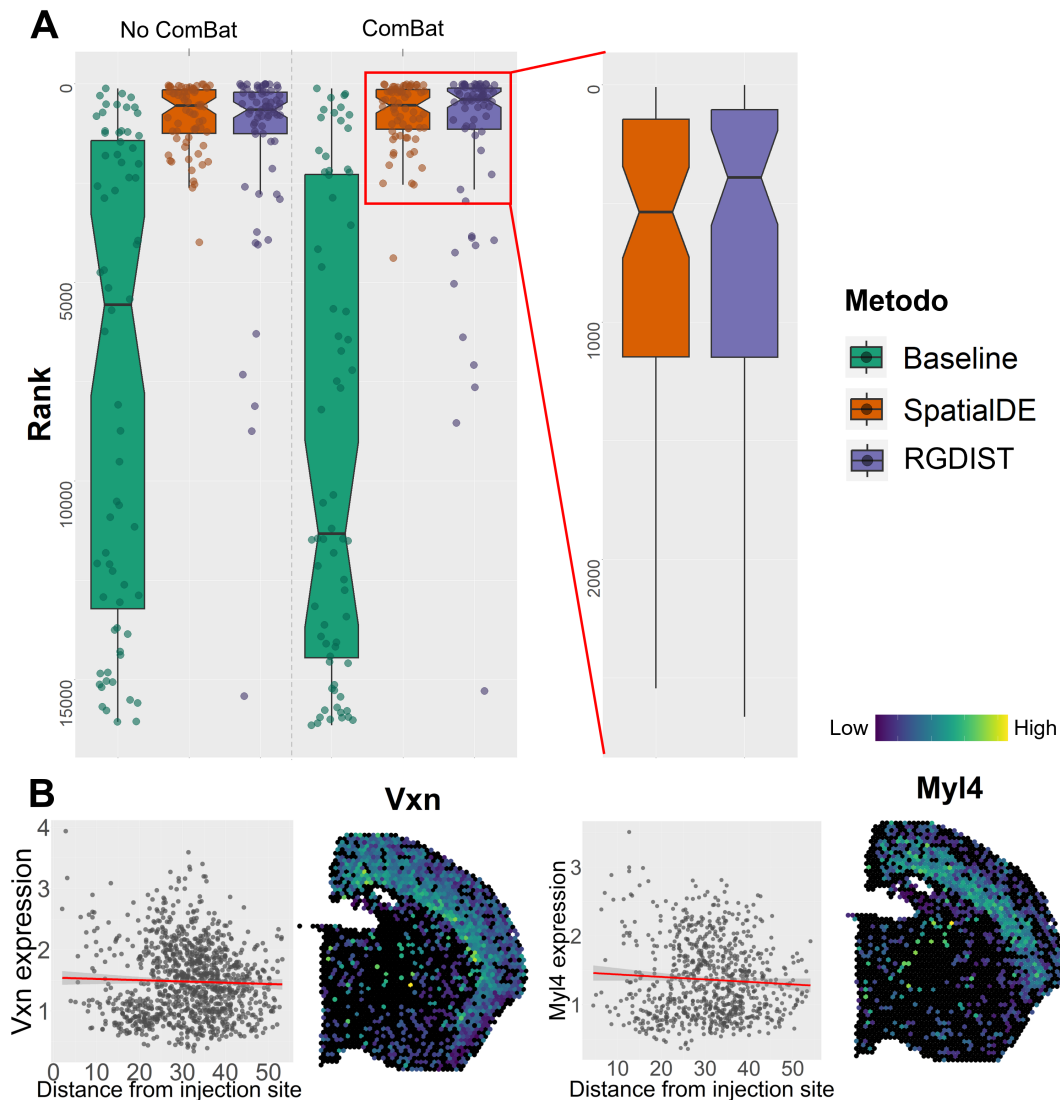


Figura 3.3: A) Prestazioni dei diversi metodi con e senza ComBat nel classificare i geni ground-truth. I metodi più performanti sono ingranditi sulla destra; l'approccio RGDIST+ComBat è il più efficace nel trovare i geni che mostrano una risposta al trattamento. B) Geni significativi per spatialDE, con un marcato pattern spaziale legato al tessuto e probabilmente non al trattamento (confrontare con Figura 3.6 in cui sono riportati i pattern dipendenti dal trattamento).

Il metodo spatialDE ha rilevato 4280 geni che presentano un pattern spaziale, inclusi tutti i geni significativi per RGDIST, sebbene fossero classificati in basso in confronto. Da questi 4 280 e dai risultati del metodo *baseline* sono stati estratti i sottoinsiemi dei primi 56 geni e dei primi 37 geni per confrontarli con i geni significativi per RGDIST, rispettivamente prima e dopo la BC. Nella Figura 3.4 (C) sono riportati i relativi diagrammi di Venn. Tra i geni che sono risultati significativi, due di essi, ovvero Hmox1 e Timp1, sono stati individuati sia da

RGDIST che da spatialDE e fanno entrambi parte dell'insieme ground truth, riportati in Figura 3.4D. I restanti sono individuati esclusivamente da RGDIST e sono discussi in nella sezione successiva. I geni in comune fra i tre metodi sono pochi, questo suggerisce che l'utilizzo di più metodi contemporaneamente può essere utile per estrarre un insieme di geni con differenti caratteristiche nell'espressione, ma significativi.

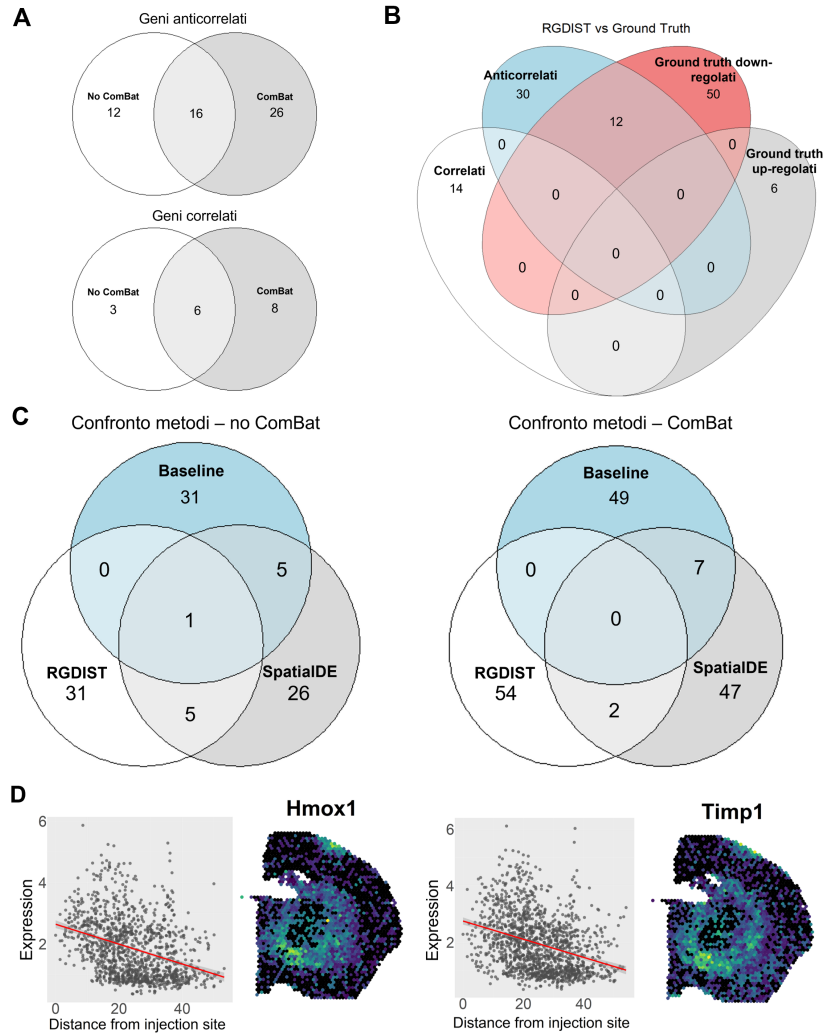


Figura 3.4: A) geni significativi prima e dopo la BC, con la correzione i geni significativi aumentano. B) Geni significativi per RGDIST e l'insieme *ground truth*, 12 dei geni anticorrelati sono in comune. C) Geni significativi per i tre metodi, prima e dopo la BC. D) Geni significativi in comune tra RGDIST e spatialDE.

3.4 Analisi dei risultati

3.4.1 Geni significativi

Per identificare i geni influenzati dal trattamento, è stato calcolato il coefficiente di correlazione per i rank di Spearman tra l'espressione di ciascun gene in tutti i punti e la distanza corrispondente dal sito di iniezione a ciascun punto,

escludendo i conteggi nulli. Le stesse analisi sono state eseguite sia sui dati di espressione non corretti che su quelli corretti per valutare gli effetti della fase di correzione. Per estrarre un insieme di geni che rispondono al trattamento, è stata utilizzata una soglia sul p-value di 0.01 e una soglia sulla correlazione di ± 0.5 per i dati prima della correzione. Lo stesso approccio di selezione è stato applicato ai dati dopo la BC, ma la soglia sulla correlazione è stata fissata a ± 0.35 . Tali soglie sono state scelte sulla base di un'analisi simile a un *volcano plot*¹ in cui l'espressione differenziale è stata sostituita dalla correlazione tra espressione e distanze; il plot relativo ai dati corretti è riportato in Figura 3.5. L'analisi ha rivelato due gruppi di geni significativi: correlati (9 prima della BC; 14 dopo la BC), che mostrano una tendenza crescente nella loro espressione all'aumentare della distanza dal sito di iniezione; e anticorrelati (28 prima della BC; 42 dopo la BC), che mostrano una tendenza decrescente. I 56 geni significativi risultanti dopo la batch correction sono riportati in Tabella 3.1. Tra questi geni significativi sono presenti anche alcuni geni *ground truth*, in particolare 12 dei 42 geni anticorrelati, Figura 3.4, consultabili in Tabella 3.1

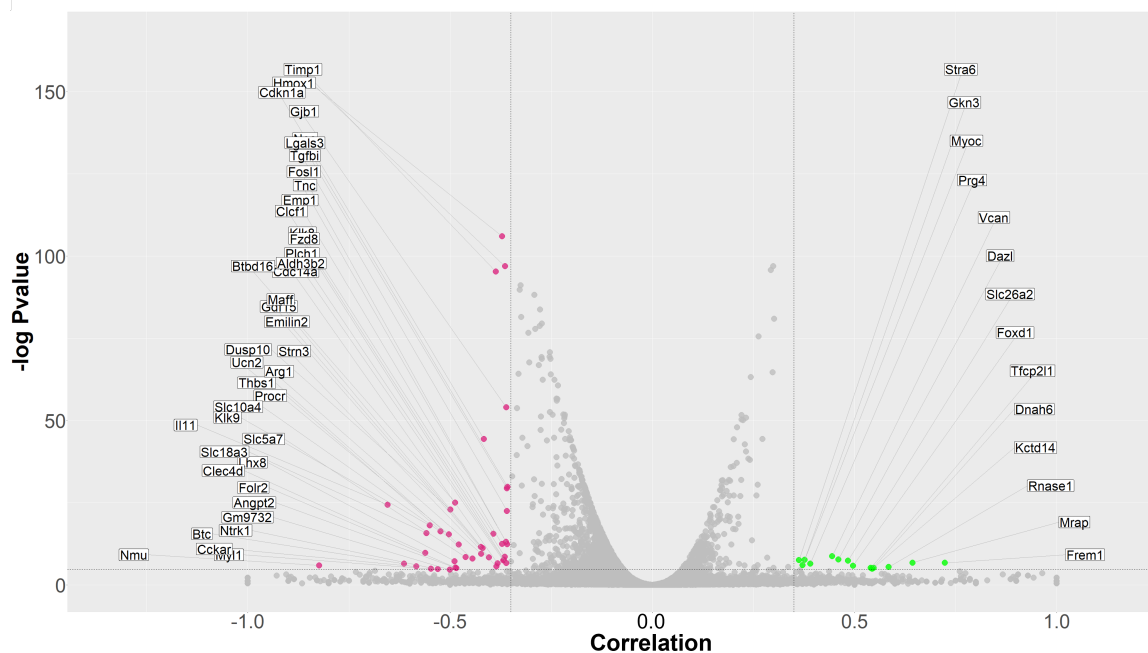


Figura 3.5: La selezione dei geni significativi è stata effettuata impostando una soglia di 0,01 sui valori di p e una soglia di $\pm 0,35$ sui valori di correlazione. I geni correlati sono mostrati in verde (a destra), mentre i geni anticorrelati sono mostrati in rosa (a sinistra).

Tra i geni significativi ce ne sono alcuni di particolare interesse, i cui modelli

¹Grafico a dispersione, o *scatter plot*, che mostra la significatività statistica (p-value) rispetto all'entità del cambiamento nell'espressione (*fold change*), rispettivamente sugli assi y e x. Consente una rapida identificazione visiva dei geni con grandi fold change che sono anche statisticamente significativi.

di espressione sul tessuto e i grafici di espressione in funzione della distanza dal sito di iniezione sono visualizzati in Figura 3.6 (le visualizzazioni dei pattern di espressione spaziale sono state ottenute con il pacchetto R `spatialLIBD` v1.10.1 [55]). Tali geni sono dettagliati di seguito.

Il33

In riferimento a tali geni, *Il33* codifica per IL-33, una citochina nucleare derivata dal tessuto appartenente alla famiglia delle IL-1. È espressa da cellule epiteliali, endoteliali e fibroblasti sia durante l'omeostasi che durante l'infiammazione e agisce come segnale d'allarme per le cellule immunitarie rilasciato in seguito a danni cellulari o tissutali. Oltre all'infiammazione, IL-33 è strettamente correlata allo stress ossidativo. Ricerche condotte su animali hanno dimostrato che l'IL-33 può esercitare un effetto neuroprotettivo in caso di emorragia cerebrale. Tale effetto sembra essere legato all'attivazione selettiva della microglia di tipo M2, che a sua volta può ridurre l'infiammazione, l'apoptosi e l'autofagia [56]. Nel topo, dopo un danno cerebrale, gli oligodendrociti post-mitotici OLIG2+ rilasciano rapidamente IL-33 per promuovere il recupero [57]. Lo studio di Miao et al. suggerisce che alti livelli di IL-33 possono ridurre il danno cerebrale dopo un'ICH.

Gjb1

Gjb1 codifica per la proteina della giunzione gap β -1, nota anche come Connexina 32 (Cx32), espressa dalle cellule di Schwann e dagli oligodendrociti, le cellule mielinizzanti del sistema nervoso periferico e del sistema nervoso centrale, rispettivamente [58]. In seguito ad un ictus, i principali eventi trascrizionali comprendono anche l'attivazione di geni correlati alla sopravvivenza di neuroni e oligodendrociti, al danno neuronale, alla mielinizzazione assonale e alla neurogenesi. I cambiamenti correlati alla mielinizzazione includono la proteina Cx32, che è coinvolta nella formazione di canali di membrana della giunzione gap e nel processo di mielinizzazione assonale [59]. Inoltre, studi hanno evidenziato che i canali della connexina sono coinvolti nei processi di infiammazione, stress ossidativo, rottura della barriera emato-encefalica (BEE) e morte cellulare nell'ICH [60].

Cdkn1a

Il gene *Cdkn1a* codifica per l'Inibitore della chinasi ciclina dipendente 1A.

Uno studio condotto da Fan et al. ha identificato sei geni differenzialmente espressi legati alla ferroptosi (CDKN1A, CXCL2, DDIT4, JUN, SLC7A5 e ZFP36). La ferroptosi è un tipo di morte cellulare regolata dovuta all'accumulo intracellulare di ferro, mediato dalla perossidazione lipidica. In campo neurologico, è stato dimostrato che la ferroptosi si verifica nell'ictus ischemico, nell'emorragia intracerebrale e nell'emorragia subaracnoidea. I risultati dello studio di Fan et al. hanno evidenziato che la coppia CDKN1A/JUN è un biomcatore diagnostico robusto e promettente per identificare i pazienti colpiti da ictus. Inoltre, questi due geni sembrano regolare la ferroptosi durante la progressione dell'ictus [61].

Il11

IL-11, codificata dal gene Il11, è un membro della famiglia di citochine IL-6 che ha vari effetti immunologici su processi immunitari sia innati che adattivi. Studi in vitro hanno dimostrato che IL-11 funziona come citochina anti-infiammatoria modificando la funzione effettrice dei macrofagi; inoltre è coinvolta nella riparazione della barriera epiteliale, che di per sé può modulare le risposte infiammatorie [62].

Fosl1

Fosl1 codifica per l'antigene *fos-like* 1 ed è un membro della famiglia di geni Fos. Questi geni codificano per proteine con struttura a *leucine zipper*, che possono dimerizzare con le proteine della famiglia JUN, formando il complesso di fattori di trascrizione noto come complesso della proteina attivatrice 1 (AP-1) [63]. Il complesso AP-1 è responsabile dell'attivazione dell'infiammazione e i risultati dello studio di Rehnström et al. hanno confermato un'aumentata espressione di geni nei vasi che contribuiscono all'infiammazione, alla disfunzione endoteliale e alla rottura della BEE. Tra questi ci sono Fosl1 e JunB, che fanno parte del complesso AP-1, e sono coinvolti nella trascrizione e nella risposta alle citochine [64]. Fosl1 è uno dei geni significativi in comune con quelli *ground truth*.

Gdf15

Gdf15 codifica per il fattore di crescita e differenziazione 15 (GDF15), un membro della superfamiglia delle proteine del fattore di crescita trasformante- β (TGF- β). GDF15 funziona come marcatore infiammatorio e svolge un ruolo

critico nella regolazione degli adattamenti endoteliali dopo il danno vascolare. I dati supportano l'uso potenziale di GDF15 come biomarcatore nei pazienti con ictus [65].

Procr

Il gene *Procr* codifica per il recettore endoteliale della proteina C (PROCR). È stato dimostrato che PROCR attiva vie di segnalazione intracellulari, portando ad una serie di effetti citoprotettivi. Quando la Proteina C Attivata (APC) si lega alle cellule che esprimono PROCR, scatena molteplici attività citoprotettive benefiche che sopprimono l'apoptosi, l'infiammazione e la rottura della barriera endoteliale [66].

Aldh3b2

Il gene *Aldh3b2* codifica per un membro della famiglia delle Aldeide Deidrogenasi 3, ALDH3B2. Questo enzima è coinvolto nel metabolismo dei prodotti aldeidici generati dalle ROS e mostra una vasta preferenza di substrato, che va dalle aldeidi a catena media a quelle a catena lunga. È stato osservato che ALDH3B2 è localizzato nei droplet lipidici [67]. I neuroni hanno livelli relativamente bassi di trigliceridi, tuttavia, i droplet lipidici sono frequentemente osservati nei neuroni coltivati e il loro numero aumenta in risposta allo stress cellulare, come l'eccitotossicità² o in seguito a un ictus [68].

Gm9732

Particolarmente interessante è il trascritto Gm9732. Il prefisso "Gm-" è un'annotazione tipica del *Mouse Genome Informatics* (MGI), uno dei principali database per il topo da laboratorio, che fornisce dati genetici, genomici e biologici integrati per facilitare lo studio della salute e delle malattie umane [69]; i trascritti che iniziano per "Gm-" sono geni annotati che non hanno ancora un nome canonico e dei quali si sa ben poco. In base alla classificazione dei *BioTypes* di Ensembl, database che fornisce informazioni aggiornate sui principali genomi eucariotici [70], questi geni includono RNA codificanti proteine, lunghi RNA non codificanti, trascritti antisense e pseudogeni. Bisogna prestare attenzione all'abbondanza

²fenomeno di tossicità neuronale conseguente all'esposizione a concentrazioni relativamente alte di acido glutammico, il principale neurotrasmettitore eccitatorio a livello del sistema nervoso centrale.

di questi trascritti, poiché potrebbero essere il risultato di letture mappate in modo errato. Il gene Gm9732 risulta significativo per RGDIST e la sua espressione spaziale sembra essere effettivamente legata all'iniezione di eme; infatti, come si può vedere in Figura 3.7, è espresso nella zona prossima al sito di iniezione e la sua espressione diminuisce in funzione della distanza. Del trascritto Gm9732 non si conoscono ancora le funzioni e l'MGI riporta un *BioType conflict*, infatti è annotato sia come gene codificante sia come *non-coding* RNA [71]. Questo risultato è molto interessante poiché sembra che il prodotto di Gm9732 sia coinvolto nella risposta al trattamento.

3.4.2 Analisi di arricchimento delle pathway

L'analisi di arricchimento genico verifica se un insieme di geni di interesse - ad esempio differenzialmente espressi in un esperimento - abbia un'intersezione statisticamente significativa con insiemi di geni noti biologicamente rilevanti, ad esempio quelli coinvolti nella stessa pathway molecolare. Per investigare i meccanismi cellulari che coinvolgono i geni identificati da RGDIST, è stata condotta un'analisi di arricchimento utilizzando la piattaforma online DAVID [72]. I set di geni arricchiti ottenuti riguardavano una varietà di meccanismi rilevanti, tra cui *risposte a stimoli xenobiotici* ($p \approx 2.9 \times 10^{-4}$), *attività citochinica* ($p \approx 1.3 \times 10^{-3}$), *morte cellulare* ($p \approx 2.5 \times 10^{-3}$), *attività di fattori di crescita* ($p \approx 3.6 \times 10^{-3}$), *regolazione positiva del processo di morte cellulare programmata* ($p \approx 2.1 \times 10^{-2}$), *regolazione positiva del metabolismo delle specie reattive dell'ossigeno* ($p \approx 6.1 \times 10^{-2}$), *risposta a lesioni assonali* ($p \approx 7.3 \times 10^{-2}$). Questi risultati sono coerenti con gli effetti causati dalla presenza di emoglobina libera nei tessuti descritti nel Capitolo 1. Analoghe vie metaboliche emergono dall'analisi dei geni *ground truth* e dei geni significativi identificati da spatialDE.

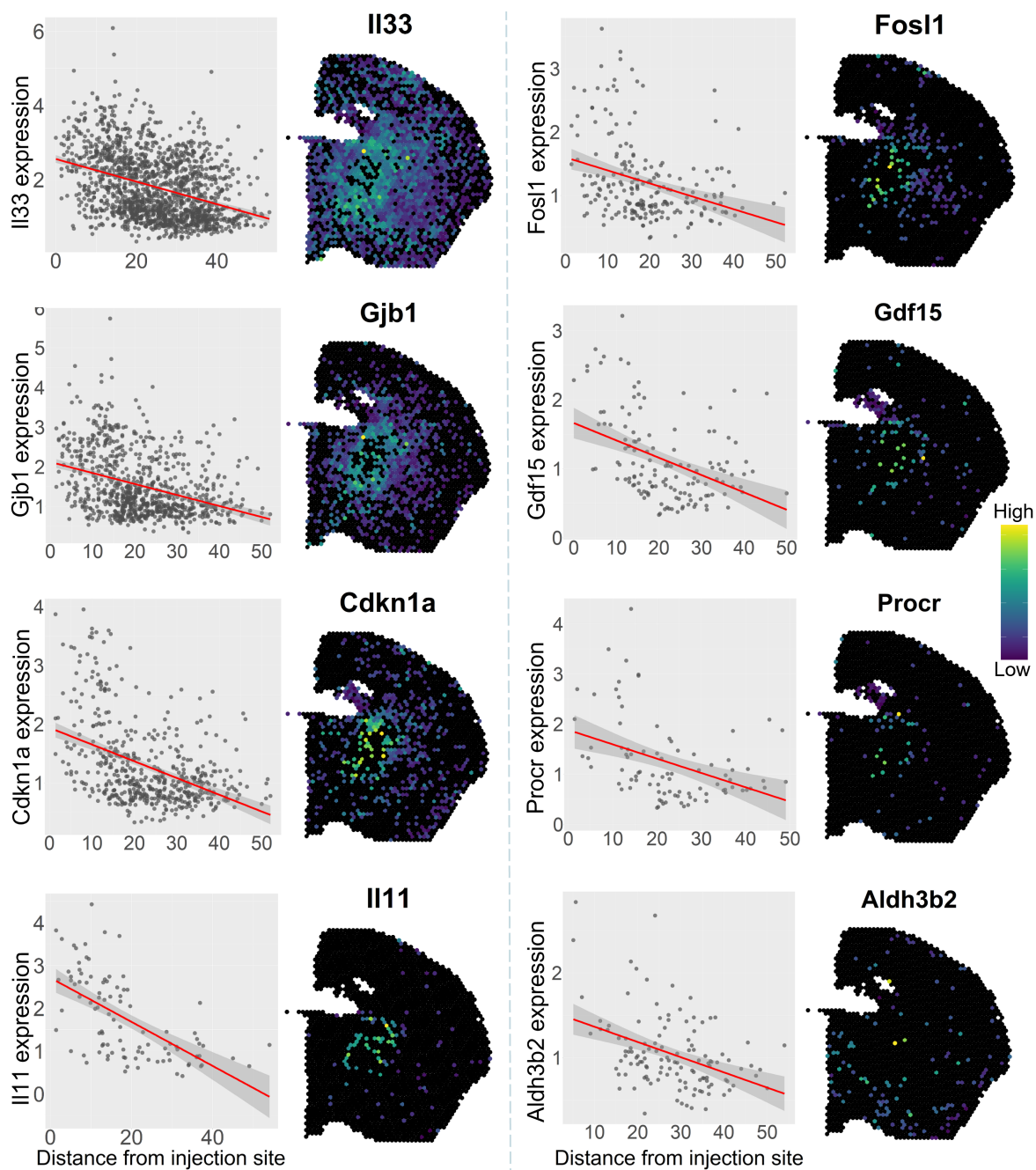


Figura 3.6: Grafici distanza-espressione e visualizzazioni spaziali dell'espressione dei geni selezionati. Nei plot spaziali, è evidente un pattern di espressione legato al trattamento: attorno al sito di iniezione l'espressione è maggiore, ma tende a diminuire allontanandosi da esso (confrontare con Figura 3.3B, in cui sono riportati dei pattern dipendenti dal tessuto e probabilmente non dal trattamento).

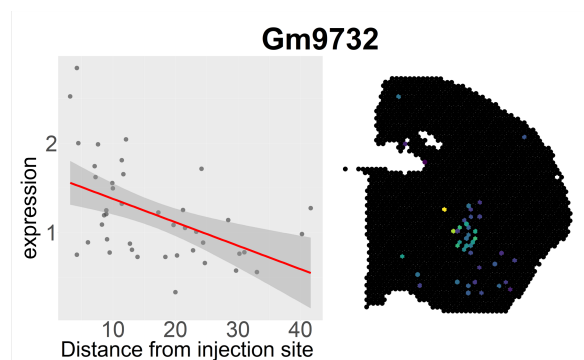


Figura 3.7: A- Visualizzazione del sito di iniezione. B- Grafico distanza-espressione e visualizzazione spaziale dell'espressione del gene Gm9732.

Tabella 3.1: geni significativi per RGDIST classificati in base al p-value della correlazione

Gene	Cor	Pvalue	Rank	Commento del prodotto genico
Geni Anticorrelati				
Il33	-0.37	$9,7 \cdot 10^{-47}$	1	Citochina anti-infiammatoria
Timp1*	-0.36	$8,2 \cdot 10^{-43}$	2	Inibitore della metalloproteinasi-1, ruolo nel volume dell'ematoma[73]
Hmox1*	-0.39	$4,2 \cdot 10^{-42}$	3	Risposta a stress ossidativo, emorragia e traumi [74]
Gjb1	-0.36	$3,4 \cdot 10^{-24}$	4	Proteina del canale di membrana della giunzione gap β -1
Cdkn1a	-0.42	$5,2 \cdot 10^{-20}$	5	Implicato nella ferroptosi
Nes*	-0.36	$1,1 \cdot 10^{-13}$	6	Proteina dei filamenti intermedi espressa nei neuroni e nella glia [75]
Lgals3*	-0.36	$1,7 \cdot 10^{-13}$	7	Coinvolto nelle risposte infiammatorie cerebrali acute [76]
Btbd16*	-0.49	$1,4 \cdot 10^{-11}$	8	Contiene dominio di interazione proteina-proteina BTB [77]
Il11	-0.65	$2,6 \cdot 10^{-11}$	9	Citochina anti-infiammatoria
Dusp10	-0.50	$1,1 \cdot 10^{-10}$	10	Fosfatasi a doppia specificità, indotta da reazioni infiammatorie[78]
Tgfb1*	-0.36	$1,7 \cdot 10^{-10}$	11	Coinvolto nella patogenesi e nell'infiammazione [79]
Klk9*	-0.55	$1,4 \cdot 10^{-8}$	12	Regola pressione sanguigna, permeabilità vasi e plasticità neurale[80]
Slc10a4	-0.52	$8,8 \cdot 10^{-8}$	13	Trasportatore di soluti [81]
Slc18a3	-0.55	$1,5 \cdot 10^{-7}$	14	Trasportatore di soluti [81]
Clcf1*	-0.39	$1,7 \cdot 10^{-7}$	15	Citochina della famiglia IL-6, meccanismi di riparazione neurale[82]
Ucn2	-0.50	$2,1 \cdot 10^{-7}$	16	Proprietà antinfiammatorie che riducono la neurotossicità [83]
Fosl1*	-0.36	$1,8 \cdot 10^{-6}$	17	Fattore trascrizionale con dominio <i>leucine zipper</i>
Emp1*	-0.37	$4,0 \cdot 10^{-6}$	18	Proteina delle giunzioni occludenti della BEE [84]
Tnc*	-0.36	$4,2 \cdot 10^{-6}$	19	Dopo SAH provoca edema cerebrale e rottura della BEE [85]
Thbs1	-0.48	$4,7 \cdot 10^{-6}$	20	Attivazione piastrinica, angiogenesi e infiammazione [86]
Gdf15	-0.42	$9,7 \cdot 10^{-6}$	21	Marker di infiammazione
Maff	-0.42	$1,4 \cdot 10^{-5}$	22	Fattore trascrizionale <i>leucine zipper</i> , stress ossidativo [87]
Clec4d	-0.56	$5,9 \cdot 10^{-5}$	23	Regola la vascolarizzazione cerebrale [88]
Arg1*	-0.42	$7,6 \cdot 10^{-5}$	24	Causa disfunzioni cerebrali e promuove la risposta infiammatoria[89]
Klk8	-0.36	$1,9 \cdot 10^{-4}$	25	Regola pressione sanguigna, permeabilità vasi e plasticità neurale[80]
Slc5a7	-0.46	$2,2 \cdot 10^{-4}$	26	Trasportatore di soluti [81]
Emilin2	-0.40	$2,4 \cdot 10^{-4}$	27	Glicoproteina extracellulare associata a sviluppo vascolare [90]
Procr	-0.44	$3,3 \cdot 10^{-4}$	28	Attiva pathway citoprotettive
Plch1	-0.37	$4,9 \cdot 10^{-4}$	29	Coinvolto nella trasduzione del segnale intracellulare [91]
Aldh3b2	-0.37	$5,9 \cdot 10^{-4}$	30	Metabolismo aldeidi derivanti da ROS
Lhx8	-0.49	$7,3 \cdot 10^{-4}$	31	Fattore trascrizionale, formazione di neuroni colinergici [92]
Fzd8	-0.36	$1,3 \cdot 10^{-3}$	32	Recettore della via Wnt implicato nello sviluppo cerebrale [93]
Cdc14a	-0.38	$1,5 \cdot 10^{-3}$	33	Agisce nella regolazione del danno infiammatorio [94]
Btc	-0.61	$1,6 \cdot 10^{-3}$	34	Membro della famiglia dei fattori di crescita epidermici [95]
Nmu	-0.82	$2,6 \cdot 10^{-3}$	35	Neuropeptide coinvolto nelle malattie infiammatorie[96]
Strn3	-0.39	$3,6 \cdot 10^{-3}$	36	Subunità regolatoria della fosfatasi 2A [97]
Cckar	-0.58	$3,7 \cdot 10^{-3}$	37	Recettore del neuropeptide CCK, sviluppo neocorticale [98]
Folr2	-0.49	$5,1 \cdot 10^{-3}$	38	Recettore dei folati, espresso nel cervello fetale [99]
Angpt2	-0.48	$6,1 \cdot 10^{-3}$	39	Anti-angiogenesi; causa espansione dell'ictus e rottura BEE[100]
My11	-0.55	$7,6 \cdot 10^{-3}$	40	Catena leggera alcalina della miosina [101]
Ntrk1	-0.53	$8,3 \cdot 10^{-3}$	41	Recettore tirosina chinasi neurotrofica-1, ruolo neuroprotettivo [102]
Gm9732	-0.50	$9,6 \cdot 10^{-3}$	42	Funzione non definita
Geni Correlati				
Vcan	0.44	$1,6 \cdot 10^{-4}$	1	Ruolo neuroprotettivo da infiammazione o rottura BEE[103]
Dazl	0.46	$4,4 \cdot 10^{-4}$	2	Proteina che lega RNA espressa in cellule germinali [104]
Myoc	0.38	$4,9 \cdot 10^{-4}$	3	Proteina secreta e glicosilata coinvolta nella crescita dei neuriti[105]
Stra6	0.36	$5,1 \cdot 10^{-4}$	4	Recettore e trasportatore del retinolo espresso sulla BEE [106]
Slc26a2	0.48	$6,0 \cdot 10^{-4}$	5	Trasportatore di soluti [81]
Mrap	0.64	$1,1 \cdot 10^{-3}$	6	Proteina accessoria dei recettori accoppiati a proteine G [107]
Frem1	0.72	$1,2 \cdot 10^{-3}$	7	Proteina della membrana basale, mediatore adesione epiteliale[108]
Prg4	0.39	$1,5 \cdot 10^{-3}$	8	Ruolo nella regolazione dell'infiammazione [109]
Gkn3	0.37	$2,6 \cdot 10^{-3}$	9	Proteina secreta co-espressa con il fattore trefoil-2 nello stomaco[110]
Foxd1	0.50	$2,9 \cdot 10^{-3}$	10	Fattore di trascrizione espresso nelle cellule neuroepiteliali [111]
Rnase1	0.58	$4,3 \cdot 10^{-3}$	11	Stabilizza la BEE, attiva leucociti e microglia dopo danni[112]
Kctd14	0.55	$5,6 \cdot 10^{-3}$	12	Coinvolto in disturbi del neurosviluppo e neuropsichiatrici [113]
Tfcp2l1	0.54	$5,7 \cdot 10^{-3}$	13	Fattore di trascrizione coinvolto in funzioni cerebrali [114]
Dnah6	0.54	$7,1 \cdot 10^{-3}$	14	Costituente delle proteine motrici associate ai microtubuli [115]

*Geni in comune con l'insieme *ground truth*.

Conclusioni

La ricerca di geni che mostrano una tendenza nell'espressione rispetto alla distanza dal sito di iniezione può essere considerata uno strumento efficace nell'analisi dei dati di trascrittomica spaziale. Il confronto tra metodi ha dimostrato che sia spatialDE che RGDIST sono accurati nell'identificare geni in risposta al trattamento, senza l'utilizzo di un campione controllo. In particolare, RGDIST unito a Combat è risultato essere l'approccio più efficace nel trovare tali geni, mentre SpatialDE tende a individuare pattern spaziali in modo generale, compresi quelli che dipendono dal tipo di tessuto, piuttosto che dalla presenza del trattamento.

L'analisi con RGDIST ha rivelato 56 geni significativi, dei quali solo 2 sono stati trovati anche da spatialDE. Questi geni sono biologicamente rilevanti e coerenti con gli effetti attesi dalla presenza di emoglobina ed eme liberi, tra cui il mantenimento della barriera emato-encefalica, attività pro- e anti-infiammatorie, morte cellulare, metabolismo di specie reattive dell'ossigeno e ferroptosi. Questi risultati suggeriscono che l'approccio utilizzato in questo studio può essere utile per identificare i geni che rispondono a un trattamento, anche in assenza di un campione di controllo, basandosi sull'ipotesi che la distanza delle cellule dal sito di iniezione attraverso il tessuto influisca sulla corrispondente risposta trascrizionale.

Pur essendo un metodo accurato nella ricerca di geni responsivi ad un trattamento, RGDIST presenta alcune limitazioni. Infatti, il metodo utilizza una misura lineare e ciò potrebbe non catturare completamente la complessità dei pattern di espressione genica. Inoltre, l'elevata variabilità biologica che caratterizza alcuni tessuti non è del tutto rimossa dalla correzione batch; infatti, è possibile che alcuni geni effettivamente responsivi al trattamento, vengano persi in questa fase e migliorare tale aspetto aiuterebbe ad aumentare l'accuratezza del metodo. Tuttavia, nonostante queste limitazioni, RGDIST rappresenta comunque un metodo di analisi promettente per l'identificazione di geni responsivi ad un trattamento e potrebbe essere integrato con altri approcci di analisi per confermare e ampliare i risultati ottenuti.

La trascrittomica spaziale è un campo in costante evoluzione, con molte nuove tecniche che vengono sviluppate e migliorate continuamente. Tuttavia, nonostante i progressi significativi, ci sono ancora molte opportunità per affinare

le tecniche di raccolta dati, integrare i dati in database completi, esplorare altre specie oltre all'uomo e al topo e creare un flusso di lavoro open-source facile da usare e accessibile a tutti. Ciò permetterà di diffondere le conoscenze e le tecniche a un pubblico più ampio, accelerando così i progressi nella ricerca. Attualmente, la maggior parte delle tecniche di trascrittomica spaziale richiede la preparazione di sezioni di tessuto per l'analisi dell'espressione genica. Per comprendere pienamente l'architettura spaziale dei tessuti biologici e la loro dinamica spazio-temporale, sarà necessario sviluppare strumenti computazionali in grado di allineare e analizzare dati provenienti da più sezioni, sia da individui diversi che da stadi di sviluppo differenti. In questo modo sarà possibile estendere le attuali tecniche di analisi spaziale alle dimensioni temporale e 3D, ottenendo una comprensione più completa delle interazioni tra le cellule e delle loro funzioni all'interno del tessuto [13]. Inoltre, l'analisi spazio-temporale potrebbe essere utile per monitorare i cambiamenti nella distribuzione genica durante lo sviluppo, in risposta a stimoli esterni o nel microambiente tumorale, aiutando a identificare nuovi target terapeutici e biomarcatori per la diagnosi precoce delle malattie.

Di recente, *10x Genomics* ha rilasciato *Xenium In Situ*, un nuovo strumento per l'analisi spaziale con una risoluzione subcellulare che permette la profilazione di 100 o 1.000 target di RNA insieme a proteine multiplexate. L'analisi *in situ* ad alta risoluzione di tessuti complessi rivoluzionerà il modo di comprendere la biologia, fornendo informazioni non ottenibili con altre tecnologie. Inoltre la combinazione di tecnologie complementari consentirà di ottenere una risoluzione più elevata e informazioni biologiche più ricche e dettagliate [116].

In conclusione, metodi simili a RGDIST potrebbero diventare fondamentali in futuro per migliorare la comprensione della risposta genica a diverse condizioni fisiologiche e patologiche. Ad esempio, l'identificazione di geni responsivi ad un trattamento potrebbe essere utile per sviluppare nuovi farmaci o terapie personalizzate. Inoltre, la combinazione di diversi approcci di analisi dei dati, tra cui RGDIST, potrebbe fornire una visione più completa e dettagliata della risposta genica, consentendo di individuare nuovi target terapeutici o biomarcatori di malattia.

Bibliografia

- [1] T. Can, “Introduction to bioinformatics,” vol. 1107, pp. 51–71.
- [2] W. J. S. Diniz and F. Canduri, “REVIEW-ARTICLE bioinformatics: an overview and its applications,” vol. 16, no. 1.
- [3] I. Subramanian, S. Verma, S. Kumar, A. Jere, and K. Anamika, “Multi-omics data integration, interpretation, and its application,” vol. 14, p. 1177932219899051.
- [4] T. A. Brown, “Transcriptomes and proteomes,” in *Genomes. 2nd edition*. Wiley-Liss. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK21121/>
- [5] C. A. Aguilar and H. G. Craighead, “Micro- and nanoscale devices for investigation of epigenetics and chromatin dynamics,” vol. 8, no. 10, pp. 709–718. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4072028/>
- [6] I. Koo, X. Wei, and X. Zhang, “Chapter sixteen - analysis of metabolomic profiling data acquired on GC-MS,” in *Methods in Enzymology*, ser. Cell-wide Metabolic Alterations Associated with Malignancy, L. Galluzzi and G. Kroemer, Eds. Academic Press, vol. 543, pp. 315–324. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128013298000167>
- [7] Y. Hasin, M. Seldin, and A. Lusi, “Multi-omics approaches to disease,” vol. 18, no. 1, p. 83. [Online]. Available: <https://doi.org/10.1186/s13059-017-1215-1>
- [8] M. P. Alfaro, J. L. Sepulveda, and E. Lyon, “Chapter 22 - molecular testing for targeted therapies and pharmacogenomics,” in *Accurate Results in the Clinical Laboratory (Second Edition)*, A. Dasgupta and J. L. Sepulveda, Eds. Elsevier, pp. 349–363. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128137765000224>
- [9] A. R. Thornhill, “Genetic analysis of the embryo: Preimplantation genetic diagnosis,” in *Encyclopedia of Reproduction (Second Edition)*, M. K. Skinner, Ed. Academic Press, pp. 215–221. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128012383648614>
- [10] F. Ji and R. I. Sadreyev, “RNA-seq: Basic bioinformatics analysis,” vol. 124, no. 1, p. e68.
- [11] A. McDermaid, B. Monier, J. Zhao, B. Liu, and Q. Ma, “Interpretation of differential gene expression results of RNA-seq data: review and integration,” vol. 20, no. 6, pp. 2044–2054.
- [12] B. Hwang, J. H. Lee, and D. Bang, “Single-cell RNA sequencing technologies and bioinformatics pipelines,” vol. 50, no. 8, pp. 1–14, number: 8 Publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s12276-018-0071-8>
- [13] L. Moses and L. Pachter, “Museum of spatial transcriptomics,” vol. 19, no. 5, pp. 534–546. [Online]. Available: <https://www.nature.com/articles/s41592-022-01409-2>
- [14] G. Palla, D. S. Fischer, A. Regev, and F. J. Theis, “Spatial components of molecular tissue biology,” vol. 40, no. 3, pp. 308–318. [Online]. Available: <https://www.nature.com/articles/s41587-021-01182-1>
- [15] X. Li and C.-Y. Wang, “From bulk, single-cell to spatial RNA sequencing,” vol. 13, no. 1, pp. 1–6, number: 1 Publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s41368-021-00146-0>
- [16] V. Svensson, S. A. Teichmann, and O. Stegle, “SpatialDE: identification of spatially variable genes,” vol. 15, no. 5, pp. 343–346. [Online]. Available: <http://www.nature.com/articles/nmeth.4636>

- [17] R. M. Buzzi, K. Akeret, N. Schwendinger, J. Klohs, F. Vallelian, M. Hugelshofer, and D. J. Schaer, "Spatial transcriptome analysis defines heme as a hemopexin-targetable inflammatoxin in the brain," vol. 179, pp. 277–287. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0891584921008054>
- [18] D. Rajashekar and J. W. Liang, "Intracerebral hemorrhage," in *StatPearls*. StatPearls Publishing. [Online]. Available: <http://www.ncbi.nlm.nih.gov/books/NBK553103/>
- [19] N. J. Fischbein and C. A. C. Wijman, "Nontraumatic intracranial hemorrhage," vol. 20, no. 4, pp. 469–492.
- [20] D. Kuriakose and Z. Xiao, "Pathophysiology and treatment of stroke: Present status and future perspectives," vol. 21, no. 20, p. 7609. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7589849/>
- [21] R. K. Narayan, A. I. R. Maas, F. Servadei, B. E. Skolnick, M. N. Tillinger, L. F. Marshall, and Traumatic Intracerebral Hemorrhage Study Group, "Progression of traumatic intracerebral hemorrhage: a prospective observational study," vol. 25, no. 6, pp. 629–639.
- [22] A. K. A. Unnithan, J. M. Das, and P. Mehta, "Hemorrhagic stroke," in *StatPearls [Internet]*. StatPearls Publishing. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK559173/>
- [23] J. Aronowski and X. Zhao, "Molecular pathophysiology of cerebral hemorrhage: secondary brain injury," vol. 42, no. 6, pp. 1781–1786.
- [24] F. Schlunk, M. Böhm, G. Boulouis, T. Qin, M. Arbel, I. Tamim, P. Fischer, B. J. Bacskaï, M. P. Frosch, M. Endres, S. M. Greenberg, and C. Ayata, "Secondary bleeding during acute experimental intracerebral hemorrhage," vol. 50, no. 5, pp. 1210–1215. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6478448/>
- [25] T. D. Bornes and K. S. Butcher, "Management of hypertension in the acute phase of stroke," vol. 5, no. 1, pp. 24–39. [Online]. Available: <https://www.eurekaselect.com/article/13528>
- [26] M. Hugelshofer, R. M. Buzzi, C. A. Schaer, H. Richter, K. Akeret, V. Anagnostakou, L. Mahmoudi, R. Vaccani, F. Vallelian, J. W. Deuel, P. W. Kronen, Z. Kulcsar, L. Regli, J. H. Baek, I. S. Pires, A. F. Palmer, M. Dennler, R. Humar, P. W. Buehler, P. R. Kircher, E. Keller, and D. J. Schaer, "Haptoglobin administration into the subarachnoid space prevents hemoglobin-induced cerebral vasospasm," vol. 129, no. 12, pp. 5219–5235, publisher: American Society for Clinical Investigation. [Online]. Available: <https://www.jci.org/articles/view/130630>
- [27] J. E. Meegan, J. A. Bastarache, and L. B. Ware, "Toxic effects of cell-free hemoglobin on the microvascular endothelium: implications for pulmonary and nonpulmonary organ dysfunction," vol. 321, no. 2, pp. L429–L439. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8526348/>
- [28] N. Kelley, D. Jeltama, Y. Duan, and Y. He, "The NLRP3 inflammasome: An overview of mechanisms of activation and regulation," vol. 20, no. 13, p. 3328. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6651423/>
- [29] N. S. Merle, R. Paule, J. Leon, M. Daugan, T. Robe-Rybkin, V. Poillat, C. Torset, V. Frémeaux-Bacchi, J. D. Dimitrov, and L. T. Roumenina, "P-selectin drives complement attack on endothelium during intravascular hemolysis in TLR-4/heme-dependent manner," vol. 116, no. 13, pp. 6280–6285.
- [30] O. I. Butt, P. W. Buehler, and F. D'Agnillo, "Blood-brain barrier disruption and oxidative stress in guinea pig after systemic exposure to modified cell-free hemoglobin," vol. 178, no. 3, pp. 1316–1328.
- [31] X. Zhao, S. Song, G. Sun, R. Strong, J. Zhang, J. C. Grotta, and J. Aronowski, "Neuroprotective role of haptoglobin after intracerebral hemorrhage," vol. 29, no. 50, pp. 15 819–15 827.

- [32] L. Chen, X. Zhang, J. Chen-Roetling, and R. F. Regan, “Increased striatal injury and behavioral deficits after intracerebral hemorrhage in hemopexin knockout mice,” vol. 114, no. 4, pp. 1159–1167.
- [33] T. R. Gregory, “Synergy between sequence and size in large-scale genomics,” vol. 6, no. 9, pp. 699–708, number: 9 Publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/nrg1674>
- [34] I. S. Yang and S. Kim, “Analysis of whole transcriptome sequencing data: Workflow and software,” vol. 13, no. 4, pp. 119–125. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4742321/>
- [35] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szczesniak, D. J. Gaffney, L. L. Elo, X. Zhang, and A. Mortazavi, “A survey of best practices for RNA-seq data analysis,” vol. 17, no. 1, p. 13. [Online]. Available: <https://doi.org/10.1186/s13059-016-0881-8>
- [36] J. F. Hess, T. A. Kohl, M. Kotrová, K. Rönsch, T. Paprotka, V. Mohr, T. Hutzenlaub, M. Brüggemann, R. Zengerle, S. Niemann, and N. Paust, “Library preparation for next generation sequencing: A review of automation strategies,” vol. 41, p. 107537. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0734975020300343>
- [37] P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, and J. Frisén, “Visualization and analysis of gene expression in tissue sections by spatial transcriptomics,” vol. 353, no. 6294, pp. 78–82.
- [38] R. A. Amezcua, A. T. L. Lun, E. Becht, V. J. Carey, L. N. Carpp, L. Geistlinger, F. Marini, K. Rue-Albrecht, D. Risso, C. Soneson, L. Waldron, H. Pagès, M. L. Smith, W. Huber, M. Morgan, R. Gottardo, and S. C. Hicks, “Orchestrating single-cell analysis with bioconductor,” vol. 17, no. 2, pp. 137–145, number: 2 Publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s41592-019-0654-x>
- [39] D. Righelli, L. M. Weber, H. L. Crowell, B. Pardo, L. Collado-Torres, S. Ghazanfar, A. T. L. Lun, S. C. Hicks, and D. Risso, “SpatialExperiment: infrastructure for spatially-resolved transcriptomics data in r using bioconductor,” vol. 38, no. 11, pp. 3128–3131. [Online]. Available: <https://academic.oup.com/bioinformatics/article/38/11/3128/6575443>
- [40] M. Tippi, H. R. Divecha, J. L. Catallini II, S. H. Kwon, L. M. Weber, A. Spangler, A. E. Jaffe, S. C. Hicks, K. Martinowich, L. Collado-Torres, S. C. Page, and K. R. Maynard, “Vistoseg: processing utilities for high-resolution visium/visium-if images for spatial transcriptomics data,” *bioRxiv*, 2022. [Online]. Available: TODO
- [41] D. J. McCarthy, K. R. Campbell, A. T. L. Lun, and Q. F. Wills, “Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in r,” vol. 33, no. 8, pp. 1179–1186. [Online]. Available: <https://doi.org/10.1093/bioinformatics/btw777>
- [42] A. T. L. Lun, D. J. McCarthy, and J. C. Marioni, “A step-by-step workflow for low-level analysis of single-cell RNA-seq data with bioconductor,” type: article. [Online]. Available: <https://f1000research.com/articles/5-2122>
- [43] A. E. Jaffe, T. Hyde, J. Kleinman, D. R. Weinberg, J. G. Chenoweth, R. D. McKay, J. T. Leek, and C. Colantuoni, “Practical impacts of genomic data "cleaning" on biological discovery using surrogate variable analysis,” vol. 16, p. 372.
- [44] Y. Zhang, G. Parmigiani, and W. E. Johnson, “ComBat-seq: batch effect adjustment for RNA-seq count data,” vol. 2, no. 3, p. lqaa078. [Online]. Available: <https://academic.oup.com/nargab/article/doi/10.1093/nargab/lqaa078/5909519>

- [45] J. T. Leek, W. E. Johnson, H. S. Parker, E. J. Fertig, A. E. Jaffe, Y. Zhang, J. D. Storey, and L. C. Torres, *sva: Surrogate Variable Analysis*, 2022, r package version 3.46.0.
- [46] L. Van Der Maaten, E. Postma, and J. Van den Herik, “Dimensionality reduction: a comparative review,” *J Mach Learn Res*, vol. 10, pp. 66–71, 2009.
- [47] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell, “Dimensionality reduction for visualizing single-cell data using UMAP.”
- [48] A. Anjum, S. Jaggi, E. Varghese, S. Lall, A. Bhowmik, and A. Rai, “Identification of differentially expressed genes in RNA-seq data of arabidopsis thaliana: A compound distribution approach,” vol. 23, no. 4, pp. 239–247.
- [49] K. R. Maynard, L. Collado-Torres, L. M. Weber, C. Uyttingco, B. K. Barry, S. R. Williams, J. L. Catallini, M. N. Tran, Z. Besich, M. Tippi, J. Chew, Y. Yin, J. E. Kleinman, T. M. Hyde, N. Rao, S. C. Hicks, K. Martinowich, and A. E. Jaffe, “Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex,” vol. 24, no. 3, pp. 425–436. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8095368/>
- [50] B. Verbist, G. Klambauer, L. Vervoort, W. Talloen, Z. Shkedy, O. Thas, A. Bender, H. W. H. Göhlmann, and S. Hochreiter, “Using transcriptomics to guide lead optimization in drug discovery projects: Lessons learned from the QSTAR project,” vol. 20, no. 5, pp. 505–513. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1359644614004851>
- [51] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub, “The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease,” vol. 313, no. 5795, pp. 1929–1935, publisher: American Association for the Advancement of Science. [Online]. Available: <https://www.science.org/doi/10.1126/science.1132939>
- [52] D. Vidović, A. Koletić, and S. C. Schürer, “Large-scale integration of small molecule-induced genome-wide transcriptional responses, kinome-wide binding affinities and cell-growth inhibition profiles reveal global trends characterizing systems-level drug action,” vol. 5, p. 342.
- [53] A. Musa, L. S. Ghorraie, S.-D. Zhang, G. Glazko, O. Yli-Harja, M. Dehmer, B. Haibe-Kains, and F. Emmert-Streib, “A review of connectivity map and computational approaches in pharmacogenomics,” vol. 19, no. 3, pp. 506–523. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5952941/>
- [54] Z. Pang and L. Ye, “Illuminating spatial pharmacology with in situ drug imaging,” vol. 12, no. 8, p. e1024. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9420420/>
- [55] B. Pardo, A. Spangler, L. M. Weber, S. C. Page, S. C. Hicks, A. E. Jaffe, K. Martinowich, K. R. Maynard, and L. Collado-Torres, “spatialLIBD: an r/bioconductor package to visualize spatially-resolved transcriptomics data,” vol. 23, no. 1, p. 434. [Online]. Available: <https://doi.org/10.1186/s12864-022-08601-w>
- [56] Y. Miao, Z.-x. Zhang, X. Feng, and W.-m. Sun, “IL-33 as a novel serum prognostic marker of intracerebral hemorrhage,” vol. 2021, p. 5597790. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8019392/>
- [57] C. Cayrol and J.-P. Girard, “Interleukin-33 (IL-33): A nuclear cytokine from the IL-1 family,” vol. 281, no. 1, pp. 154–168. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1111/imr.12619>
- [58] S. Scherer, S. Deschenes, Y. Xu, J. Grinspan, K. Fischbeck, and D. Paul, “Connexin32 is a myelin-related protein in the PNS and CNS,” vol. 15, no. 12, pp. 8281–8294. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6577923/>

- [59] A.-M. Buga, M. Sascau, C. Pisoschi, J. G. Herndon, C. Kessler, and A. Popa-Wagner, "The genomic response of the ipsilateral and contralateral cortex to stroke in aged rats," vol. 12, no. 6, pp. 2731–2753, [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1582-4934.2008.00252.x](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1582-4934.2008.00252.x). [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1582-4934.2008.00252.x>
- [60] Y. Zhang, S. Khan, Y. Liu, R. Siddique, R. Zhang, V. W. Yong, and M. Xue, "Gap junctions and hemichannels composed of connexins and pannexins mediate the secondary brain injury following intracerebral hemorrhage," vol. 11, no. 1, p. 27. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8772966/>
- [61] J. Fan, M. Chen, S. Cao, Q. Yao, X. Zhang, S. Du, H. Qu, Y. Cheng, S. Ma, M. Zhang, Y. Huang, N. Zhang, K. Shi, and S. Zhan, "Identification of a ferroptosis-related gene pair biomarker with immune infiltration landscapes in ischemic stroke: a bioinformatics-based comprehensive study," vol. 23, p. 59. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8761271/>
- [62] K. Y. Fung, C. Louis, R. D. Metcalfe, C. C. Kosasih, I. P. Wicks, M. D. Griffin, and T. L. Putoczki, "Emerging roles for IL-11 in inflammatory diseases," vol. 149, p. 155750. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1043466621003392>
- [63] L. Li, W. Zhang, S. Zhao, and M. Sun, "FOS-like antigen 1 is a prognostic biomarker in hepatocellular carcinoma," vol. 25, no. 6, pp. 369–376. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6941458/>
- [64] M. Rehnström, S. D. Frederiksen, S. Ansar, and L. Edvinsson, "Transcriptome profiling revealed early vascular smooth muscle cell gene activation following focal ischemic stroke in female rats – comparisons with males," vol. 21, no. 1, p. 883. [Online]. Available: <https://doi.org/10.1186/s12864-020-07295-2>
- [65] L. Rochette, M. Zeller, Y. Cottin, and C. Vergely, "Insights into mechanisms of GDF15 and receptor GFRAL: Therapeutic targets," vol. 31, no. 12, pp. 939–951. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1043276020302010>
- [66] X. V. Yang, Y. Banerjee, J. A. Fernández, H. Deguchi, X. Xu, L. O. Mosnier, R. T. Urbanus, P. G. de Groot, T. C. White-Adams, O. J. T. McCarty, and J. H. Griffin, "Activated protein c ligation of ApoER2 (LRP8) causes dab1-dependent signaling in u937 cells," vol. 106, no. 1, pp. 274–279.
- [67] T. Kitamura, S. Takagi, T. Naganuma, and A. Kihara, "Mouse aldehyde dehydrogenase ALDH3b2 is localized to lipid droplets via two c-terminal tryptophan residues and lipid modification," vol. 465, no. 1, pp. 79–87.
- [68] I. Ralhan, C.-L. Chang, J. Lippincott-Schwartz, and M. S. Ioannou, "Lipid droplets in the nervous system," vol. 220, no. 7, p. e202102136. [Online]. Available: <https://doi.org/10.1083/jcb.202102136>
- [69] M. Ringwald, J. E. Richardson, R. M. Baldarelli, J. A. Blake, J. A. Kadin, C. Smith, and C. J. Bult, "Mouse genome informatics (MGI): latest news from MGD and GXD," vol. 33, no. 1, pp. 4–18.
- [70] F. Cunningham, J. E. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, O. Austine-Orimoloye, A. G. Azov, I. Barnes, R. Bennett, A. Berry, J. Bhai, A. Bignell, K. Billis, S. Boddu, L. Brooks, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C. G. Giron, T. Genez, J. G. Martinez, C. Guizarro-Clarke, A. Gymer, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. C. Marugán, S. Mohanan, A. Mushtaq, M. Naven, D. N. Ogeh, A. Parker, A. Parton, M. Perry, I. Piližota, I. Prosovetskaia, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, J. G. Pérez-Silva, W. Stark, E. Steed, K. Sutinen, R. Sukumaran, D. Sumathipala, M.-M. Suner, M. Szpak, A. Thormann,

- F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T. A. Walsh, B. Walts, N. Willhoft, A. Winterbottom, E. Wass, M. Chakiachvili, B. Flint, A. Frankish, S. Giorgetti, L. Haggerty, S. E. Hunt, G. R. Hsley, J. E. Loveland, F. J. Martin, B. Moore, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, J. Tate, D. Thybert, S. J. Trevanion, S. Dyer, P. W. Harrison, K. L. Howe, A. D. Yates, D. R. Zerbino, and P. Flicek, “Ensembl 2022,” vol. 50, pp. D988–D995.
- [71] Gm9732 MGI mouse gene detail - MGI:3796514 - predicted gene 9732. [Online]. Available: <https://www.informatics.jax.org/marker/MGI:3796514>
- [72] B. T. Sherman, M. Hao, J. Qiu, X. Jiao, M. W. Baseler, H. C. Lane, T. Imamichi, and W. Chang, “DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update),” vol. 50, pp. W216–W221.
- [73] M. Navarro-Oviedo, R. Muñoz-Arrondo, B. Zandio, J. Marta-Enguita, A. Bonaterra-Pastra, J. A. Rodríguez, C. Roncal, J. A. Páramo, E. Toledo, J. Montaner, M. Hernández-Guillamon, and J. Orbe, “Circulating TIMP-1 is associated with hematoma volume in patients with spontaneous intracranial hemorrhage,” vol. 10, no. 1, p. 10329, number: 1 Publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s41598-020-67250-9>
- [74] Q.-Q. Li, L.-J. Li, X.-Y. Wang, Y.-Y. Sun, and J. Wu, “Research progress in understanding the relationship between heme oxygenase-1 and intracerebral hemorrhage,” vol. 9, p. 682. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6109777/>
- [75] T. Nakamura, G. Xi, Y. Hua, J. T. Hoff, and R. F. Keep, “Nestin expression after experimental intracerebral hemorrhage,” vol. 981, no. 1, pp. 108–117. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0006899303029913>
- [76] X.-J. Yan, G.-F. Yu, Y.-Q. Jie, X.-F. Fan, Q. Huang, and W.-M. Dai, “Role of galectin-3 in plasma as a predictive biomarker of outcome after acute intracerebral hemorrhage,” vol. 368, pp. 121–127.
- [77] P. J. Stogios, G. S. Downs, J. J. Jauhal, S. K. Nandra, and G. G. Privé, “Sequence and structural analysis of BTB domain proteins,” vol. 6, no. 10, p. R82. [Online]. Available: <https://doi.org/10.1186/gb-2005-6-10-r82>
- [78] M. Jiménez-Martínez, K. Stamatakis, and M. Fresno, “The dual-specificity phosphatase 10 (DUSP10): Its role in cancer, inflammation, and immunity,” vol. 20, no. 7, p. 1626. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6480380/>
- [79] A. Corona and G. C. Blobe, “The role of the extracellular matrix protein TGFBI in cancer,” vol. 84, p. 110028. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0898656821001170>
- [80] S. Weinsheimer, K. A. Goddard, A. R. Parrado, Q. Lu, M. Sinha, E. R. Lebedeva, A. Ronkainen, M. Niemelä, E. K. Khusnutdinova, R. I. Khusainova, K. Helin, J. E. Jääskeläinen, V. P. Sakovich, S. Land, H. Kuivaniemi, and G. Tromp, “Association of kallikrein gene polymorphisms with intracranial aneurysms,” vol. 38, no. 10, pp. 2670–2676, publisher: American Heart Association. [Online]. Available: <https://www.ahajournals.org/doi/10.1161/STROKEAHA.107.486225>
- [81] X. Liu, “SLC family transporters,” vol. 1141, pp. 101–202.
- [82] L. Crisponi, I. Buers, and F. Rutsch, “CRLF1 and CLCF1 in development, health and disease,” vol. 23, no. 2, p. 992. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8780587/>
- [83] K. W. Choy, A. P.-Y. Tsai, P. B.-C. Lin, M.-Y. Wu, C. Lee, A. Alias, C.-Y. Pang, and H.-K. Liew, “The role of urocortins in intracerebral hemorrhage,” vol. 10, no. 1, p. 96. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7022917/>

- [84] T. Bangsow, E. Baumann, C. Bangsow, M. H. Jaeger, B. Pelzer, P. Gruhn, S. Wolf, H. von Melchner, and D. B. Stanimirovic, "The epithelial membrane protein 1 is a novel tight junction protein of the blood—brain barrier," vol. 28, no. 6, pp. 1249–1260, publisher: SAGE Publications Ltd STM. [Online]. Available: <https://doi.org/10.1038/jcbfm.2008.19>
- [85] M. Fujimoto, M. Shiba, F. Kawakita, L. Liu, N. Shimojo, K. Imanaka-Yoshida, T. Yoshida, and H. Suzuki, "Deficiency of tenascin-c and attenuation of blood-brain barrier disruption following experimental subarachnoid hemorrhage in mice," vol. 124, no. 6, pp. 1693–1702.
- [86] C. Chen, X. Chen, S. Yang, Q. Li, Z. Ren, L. Wang, Y. Jiang, X. Gu, F. Liu, J. Mu, L. Liu, Y. Wang, J. Li, Y. Yu, J. Zhang, and C. Shen, "Association of THBS1 genetic variants and mRNA expression with the risks of ischemic stroke and long-term death after stroke," vol. 14, p. 1006473. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9545898/>
- [87] X. Wang, Y. Zhang, X. Wan, C. Guo, J. Cui, J. Sun, and L. Li, "Responsive expression of MafF to -amyloid-induced oxidative stress," vol. 2020, p. 8861358. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7787795/>
- [88] T. V. Arumugam, S. Manzanero, M. Furtado, P. J. Biggins, Y.-H. Hsieh, M. Gelderblom, K. P. MacDonald, E. Salimova, Y.-I. Li, O. Korn, D. Dewar, I. M. Macrae, R. B. Ashman, S.-C. Tang, N. A. Rosenthal, M. J. Ruitenber, T. Magnus, and C. A. Wells, "An atypical role for the myeloid receptor mincle in central nervous system injury," vol. 37, no. 6, pp. 2098–2111. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5444551/>
- [89] R. Bi, Z. Fang, M. You, Q. He, and B. Hu, "Microglia phenotype and intracerebral hemorrhage: A balance of yin and yang," vol. 15, p. 765205. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8549831/>
- [90] Q. Sa and J. L. Hoover-Plow, "EMILIN2 (elastin microfibril interface located protein), potential modifier of thrombosis," vol. 9, p. 9. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3113922/>
- [91] J.-I. Hwang, Y.-S. Oh, K.-J. Shin, H. Kim, S. H. Ryu, and P.-G. Suh, "Molecular cloning and characterization of a novel phospholipase c, PLC-eta," vol. 389, pp. 181–186.
- [92] Y. Zhao, O. Marín, E. Hermes, A. Powell, N. Flames, M. Palkovits, J. L. R. Rubenstein, and H. Westphal, "The LIM-homeobox gene *lhx8* is required for the development of many cholinergic neurons in the mouse forebrain," vol. 100, no. 15, pp. 9005–9010. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC166428/>
- [93] J. L. Boyd, S. L. Skove, J. P. Rouanet, L.-J. Pilaz, T. Bepler, R. Gordân, G. A. Wray, and D. L. Silver, "Human-chimpanzee differences in a FZD8 enhancer alter cell cycle dynamics in the developing neocortex," vol. 25, no. 6, pp. 772–779. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4366288/>
- [94] L. Zuo, J. Xie, Y. Liu, S. Leng, Z. Zhang, and F. Yan, "Down-regulation of circular RNA CDC14a peripherally ameliorates brain injury in acute phase of ischemic stroke," vol. 18, p. 283. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8653620/>
- [95] M. R. Schneider, M. Dahlhoff, N. Herbach, I. Renner-Mueller, C. Dalke, O. Puk, J. Graw, R. Wanke, and E. Wolf, "Betacellulin overexpression in transgenic mice causes disproportionate growth, pulmonary hemorrhage syndrome, and complex eye pathology," vol. 146, no. 12, pp. 5237–5246. [Online]. Available: <https://doi.org/10.1210/en.2005-0418>
- [96] S. Budhiraja and A. Chugh, "Neuromedin u: physiology, pharmacology and therapeutic potential," vol. 23, no. 2, pp. 149–157.

- [97] Y. Tang, G. Fang, F. Guo, H. Zhang, X. Chen, L. An, M. Chen, L. Zhou, W. Wang, T. Ye, L. Zhou, P. Nie, H. Yu, M. Lin, Y. Zhao, X. Lin, Z. Yuan, S. Jiao, and Z. Zhou, “Selective inhibition of STRN3-containing PP2a phosphatase restores hippo tumor-suppressor activity in gastric cancer,” vol. 38, no. 1, pp. 115–128.e9. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1535610820302695>
- [98] S. Nishimura, K. Bilgüvar, K. Ishigame, N. Sestan, M. Günel, and A. Louvi, “Functional synergy between cholecystokinin receptors CCKAR and CCKBR in mammalian brain development,” vol. 10, no. 4, p. e0124295.
- [99] S. H. Kirsch, W. Herrmann, and R. Obeid, “Genetic defects in folate and cobalamin pathways affecting the brain,” vol. 51, no. 1, pp. 139–155, publisher: De Gruyter. [Online]. Available: <https://www.degruyter.com/document/doi/10.1515/cclm-2012-0673/html>
- [100] S. Gurnik, K. Devraj, J. Macas, M. Yamaji, J. Starke, A. Scholz, K. Sommer, M. Di Tacchio, R. Vutukuri, H. Beck, M. Mittelbronn, C. Foerch, W. Pfeilschifter, S. Liebner, K. G. Peters, K. H. Plate, and Y. Reiss, “Angiopoietin-2-induced blood–brain barrier compromise and increased stroke size are rescued by VE-PTP-dependent restoration of tie2 signaling,” vol. 131, pp. 753–773. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4835530/>
- [101] J. England and S. Loughna, “Heavy and light roles: myosin in the morphogenesis of the heart,” vol. 70, no. 7, pp. 1221–1239. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3602621/>
- [102] J. Chen, J. Qin, Q. Su, Z. Liu, and J. Yang, “Treadmill rehabilitation treatment enhanced BDNF-TrkB but not NGF-TrkA signaling in a mouse intracerebral hemorrhage model,” vol. 529, no. 1, pp. 28–32. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0304394012012396>
- [103] K. Rhodes and J. Fawcett, “Chondroitin sulphate proteoglycans: preventing plasticity or protecting the CNS?” vol. 204, no. 1, pp. 33–48. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1571240/>
- [104] L. L. Zagore, T. J. Sweet, M. M. Hannigan, S. M. Weyn-Vanhentenryck, R. Jobava, M. Hatzoglou, C. Zhang, and D. D. Licatalosi, “DAZL regulates germ cell survival through a network of PolyA-proximal mRNA interactions,” vol. 25, no. 5, pp. 1225–1240.e6. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6878787/>
- [105] B. Zhou, X. Lin, Z. Li, Y. Yao, J. Yang, and Y. Zhu, “Structure–function–pathogenicity analysis of c-terminal myocilin missense variants based on experiments and 3d models,” vol. 13, p. 1019208. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9577182/>
- [106] J. Amengual, N. Zhang, M. Kemmerer, T. Maeda, K. Palczewski, and J. Von Lintig, “STRA6 is critical for cellular vitamin a uptake and homeostasis,” vol. 23, no. 20, pp. 5402–5417. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168826/>
- [107] T. R. Webb and A. J. L. Clark, “Minireview: The melanocortin 2 receptor accessory proteins,” vol. 24, no. 3, pp. 475–484. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5419097/>
- [108] J. R. Timmer, T. W. Mak, K. Manova, K. V. Anderson, and L. Niswander, “Tissue morphogenesis and vascular stability require the frem2 protein, product of the mouse myelencephalic blebs gene,” vol. 102, no. 33, pp. 11 746–11 750, publisher: Proceedings of the National Academy of Sciences. [Online]. Available: <https://www.pnas.org/doi/10.1073/pnas.0505404102>
- [109] R. J. Krawetz, S. Abubacker, C. Leonard, A. O. Masson, S. Shah, N. Narendran, P. Tailor, S. C. Regmi, E. Labit, N. Ninkovic, J. M. Corpuz, K. Ito, T. M. Underhill, P. T. Salo, T. A. Schmidt, and J. A. Biernaskie, “Proteoglycan 4 (PRG4) treatment enhances wound closure and tissue regeneration,” vol. 7, no. 1, pp. 1–18, number: 1 Publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s41536-022-00228-5>

- [110] T. R. Menheniott, A. J. Peterson, L. O'Connor, K. S. Lee, A. Kalantzis, I. Kondova, R. E. Bontrop, K. M. Bell, and A. S. Giraud, "A novel gastrokine, gkn3, marks gastric atrophy and shows evidence of adaptive gene loss in humans," vol. 138, no. 5, pp. 1823–1835.
- [111] E. A. Newman, D. W. Kim, J. Wan, J. Wang, J. Qian, and S. Blackshaw, "Foxd1 is required for terminal differentiation of anterior hypothalamic neuronal subtypes," vol. 439, no. 2, pp. 102–111. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5964039/>
- [112] T. J. Krämer, P. Hübener, B. Pöttker, C. Gölz, A. Neulen, T. Pantel, H. Goetz, K. Ritter, M. K. E. Schäfer, and S. C. Thal, "Ribonuclease-1 treatment after traumatic brain injury preserves blood–brain barrier integrity and delays secondary brain damage in mice," vol. 12, p. 5731. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8986812/>
- [113] X. Teng, A. Aouacheria, L. Lionnard, K. A. Metz, L. Soane, A. Kamiya, and J. M. Hardwick, "KCTD: A new gene family involved in neurodevelopmental and neuropsychiatric disorders," vol. 25, no. 7, pp. 887–902. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6566181/>
- [114] A. Taracha, G. Kotarba, and T. Wilanowski, "Neglected functions of TFCP2/TFCP2l1/UBP1 transcription factors may offer valuable insights into their mechanisms of action," vol. 19, no. 10, p. 2852. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6213935/>
- [115] L. B. Areal, L. P. Pereira, F. M. Ribeiro, I. G. Olmo, M. R. Muniz, M. do Carmo Rodrigues, P. F. Costa, C. Martins-Silva, S. S. G. Ferguson, D. A. M. Guimarães, and R. G. W. Pires, "Role of dynein axonemal heavy chain 6 gene expression as a possible biomarker for huntington's disease: a translational study," vol. 63, no. 3, pp. 342–348.
- [116] A. Janesick, R. Shelansky, A. Gottscho, F. Wagner, M. Rouault, G. Beliakoff, M. F. d. Oliveira, A. Kohlway, J. Abousoud, C. Morrison, T. Y. Drennon, S. Mohabbat, S. Williams, x. D. Teams, and S. Taylor, "High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue," pages: 2022.10.06.510405 Section: New Results. [Online]. Available: <https://www.biorxiv.org/content/10.1101/2022.10.06.510405v1>

Ringraziamenti

Con queste righe desidero esprimere la mia sincera gratitudine al professor Napolitano per l'importante contributo che ha dato alla mia formazione accademica con la sua guida e il suo sostegno. La ringrazio di cuore per aver dedicato il suo tempo, la sua esperienza e la sua competenza per farmi crescere sia come studente che come persona. La sua pazienza e disponibilità sono state fondamentali per il successo di questo lavoro e per la mia crescita professionale e personale. Spero di poter continuare ad imparare da lei in futuro.