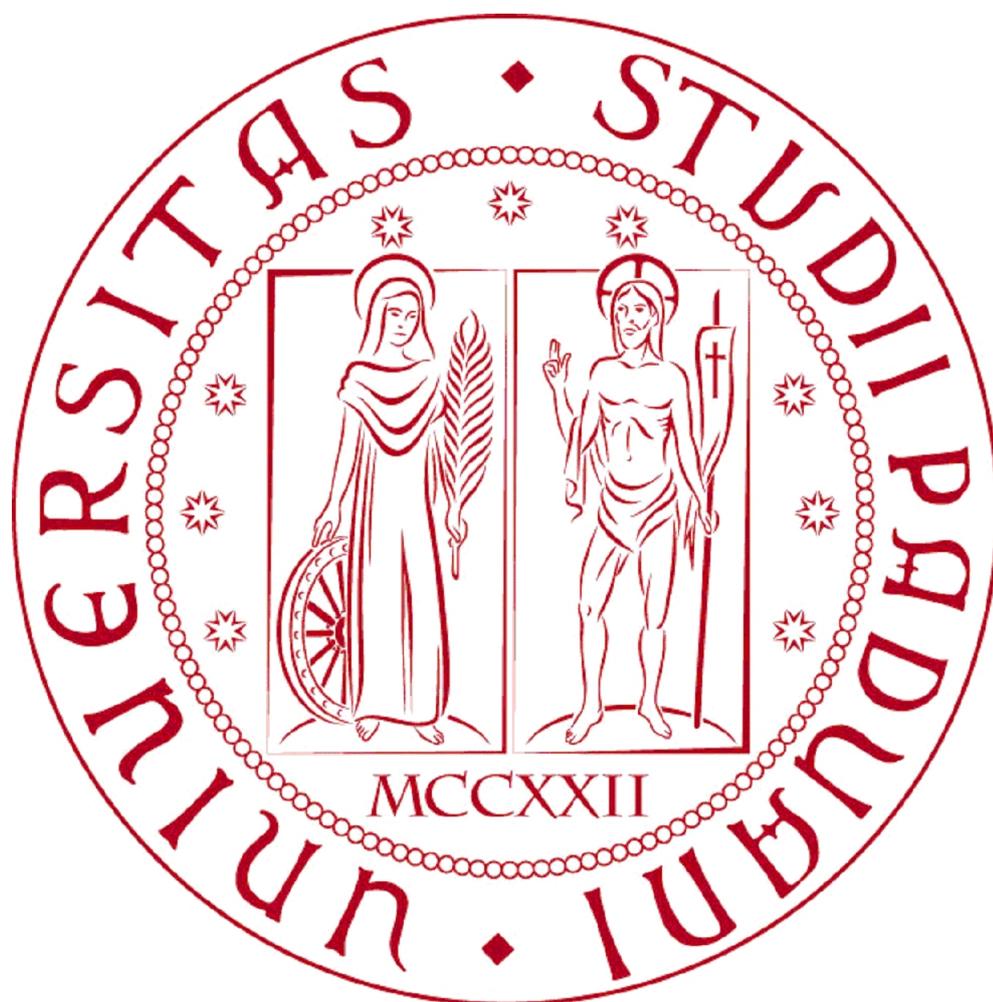


UNIVERSITÀ DEGLI STUDI DI PADOVA  
FACOLTÀ DI INGEGNERIA





UNIVERSITÀ DEGLI STUDI DI PADOVA  
FACOLTÀ DI INGEGNERIA

—  
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

—  
TESI DI LAUREA MAGISTRALE IN BIOINGEGNERIA

ANALISI MULTIMODALE DEL  
SISTEMA NEURORECETTORIALE:  
INTEGRAZIONE TRA IMAGING PET  
E LIVELLI DI ESPRESSIONE  
GENICA

RELATORE: CHIAR.MA PROF.SSA ALESSANDRA BERTOLDO

CORRELATORE: DOTT.GAIA RIZZO, DOTT.MATTIA VERONESE

LAUREANDO: ANDREA STEVAN

ANNO ACCADEMICO 2012-2013



*Alla mia famiglia.*



*“A cento anni ho perso un po’ la vista, molto l’udito. Alle conferenze non vedo le proiezioni e non sento bene. Ma penso più adesso di quando avevo vent’anni.*

*Il corpo faccia quello che vuole. Io non sono il corpo: io sono la mente.”*

RITA LEVI MONTALCINI



# Indice

<b>Sommario</b>	<b>XI</b>
<b>Introduzione</b>	<b>XIII</b>
<b>1 I neuro-recettori</b>	<b>1</b>
1.1 I recettori degli oppioidi . . . . .	2
1.2 I recettori della serotonina . . . . .	4
<b>2 Imaging funzionale: la PET</b>	<b>7</b>
2.1 Principi di funzionamento . . . . .	7
2.2 Quantificazione delle immagini PET . . . . .	9
2.2.1 Tracciante $[^{11}\text{C}]\text{DPN}$ . . . . .	11
2.2.2 Tracciante $[^{11}\text{C}]\text{CUMI}$ . . . . .	13
2.2.3 Tracciante $[^{11}\text{C}]\text{WAY}$ . . . . .	15
<b>3 Il DNA e la misura dell'espressione genica</b>	<b>17</b>
3.1 I Microarray: principi di funzionamento . . . . .	19
3.1.1 Microarray a due canali - Agilent . . . . .	23
3.1.2 Microarray ad un canale - Affymetrix . . . . .	26
<b>4 Dataset analizzati</b>	<b>31</b>
4.1 Dati PET . . . . .	31
4.1.1 Dataset DPN . . . . .	31
4.1.2 Dataset CUMI . . . . .	32
4.1.3 Dataset WAY . . . . .	32
4.2 Dati di espressione genica . . . . .	34

4.3	Atlanti anatomici MNI usati . . . . .	36
<b>5</b>	<b>Pre-processing e metodi</b>	<b>37</b>
5.1	Sincronizzazione degli atlanti . . . . .	37
5.2	Correlazione e precisione delle stime . . . . .	42
<b>6</b>	<b>Risultati e discussione</b>	<b>45</b>
6.1	Analisi dei dati genetici: variabilità e consistenza . . . . .	45
6.2	Analisi dei dati PET . . . . .	52
6.3	Correlazione per i recettori oppioidi . . . . .	54
6.4	Correlazione per i recettori serotoninergici . . . . .	57
	<b>Conclusioni</b>	<b>61</b>
	<b>Appendice A</b>	<b>63</b>
	<b>Appendice B</b>	<b>73</b>
	<b>Bibliografia</b>	<b>75</b>

# Sommario

Oggi, grazie allo sviluppo nelle tecnologie microarray, immensi dataset con i profili di espressione genica (post-mortem) del cervello umano sono disponibili. Questa grande mole di dati genetici può essere usata per comprendere meglio il funzionamento in-vivo del cervello. Inoltre, la *Positron Emission Tomography* mette a disposizione una vasta gamma di traccianti utilizzabili per monitorare molte funzioni biologiche. Lo scopo di questa tesi è di integrare dati PET con dati di espressione genica per investigare sulla possibile correlazione tra i livelli di espressione per un determinato gene e l'effettiva densità di distribuzione dei neuro-recettori per i quali tale gene codifica. In questo studio sono stati valutati due sistemi neuro-recettoriali: quello per gli oppioidi e quello per la serotonina. In una prima fase i dati genetici sono stati rimappati in uno spazio stereotassico comune ai dati PET; successivamente sono state condotte analisi indipendenti sui dataset e analisi di correlazione tra i soggetti dei dataset PET e di espressione. Usando il coefficiente di correlazione di Spearman sono state proposte delle mappe che riassumono il livello di correlazione tra i soggetti. Per la prima volta è stato fatto un confronto completo ed esaustivo sulla relazione tra i livelli di espressione genica e i dati PET, nel cervello umano. Dei sistemi neuro-recettoriali studiati, l'analisi dei recettori oppioidi ha rilevato discreti livelli di correlazione, mentre è stato confermato un forte legame tra l'espressione dei recettori serotoninergici e le loro stime PET. E' stata infine realizzata un'interfaccia grafica che implementa il codice sviluppato e rende possibile applicare le analisi condotte con qualunque dataset disponibile.



# Introduzione

Al giorno d'oggi, grazie all'enorme sviluppo nelle tecniche di sequenziamento del DNA, una grande mole di dati genetici è disponibile. Migliaia di geni sono stati sequenziati in vari distretti anatomici, permettendo di indagare sul funzionamento di vari aspetti dell'organismo umano come ad esempio il sistema neuro-recettoriale, o sull'origine genetica di malattie di grande importanza sociale come il morbo di Parkinson o la corea di Huntington. Una nuova metodologia per determinare se un gene è responsabile per qualche malattia o come contribuisce nel funzionamento della catena regolatoria di un apparato od un sistema del corpo umano è quella di integrare informazioni provenienti da diverse sorgenti informative con i dati genetici. In questo frangente, grazie alla presenza di traccianti, le tecniche di imaging PET (Positron Emission Tomography) si prestano molto bene nel monitorare l'attività genetica [1]. Grazie a questo connubio è possibile valutare ad esempio l'efficacia di terapie geniche nella cura del cancro [2] e la variazione di attività neuro-recettoriale in base all'espressione genica in patologie psichiatriche come il disturbo bipolare [3], permettendo di individuare la corretta terapia, genica o non, da utilizzare per intervenire efficacemente.

Con questi presupposti, e grazie alla presenza di nuovi atlanti [4] dell'espressione genica del cervello umano, è ora possibile studiare anche il sistema neuro-recettoriale dal punto di vista genetico, integrando i dati di espressione con le informazioni ricavate da tecniche PET. Vengono da porsi quindi alcune domande: *Per il sistema neuro-recettoriale esiste un legame tra i livelli di espressione per un determinato gene e la distribuzione spaziale dei neuro-recettori per i quali tale gene codifica?*, e poi: *Se il suddetto legame esiste, quanto forte è tale correlazione?*.

Questa tesi, in primo luogo si pone l'obbiettivo di confrontare i dati genetici di diversi sistemi neuro-recettoriali con i dati PET ricavati usando traccianti specifici per tali sistemi; e in secondo luogo di determinare il livello di correlazione, in regioni di interesse comuni, tra l'espressione genica e le misure PET dei neuro-recettori considerati. Infatti i metodi applicati per la PET rappresentano metodologie convalidate per misure di legame specifico nei sistemi neuro-recettoriali.

Il primo capitolo riassume cosa sono e come funzionano i neuro-recettori, focalizzandosi poi nello specifico, sui recettori degli oppioidi e della serotonina. Il secondo capitolo descrive la PET come metodica di imaging funzionale, spiegando i metodi di quantificazione usati per i dataset analizzati. Il terzo capitolo parla del DNA e dell'espressione genica, spiegando cosa sono e come funzionano le due tecnologie Microarray disponibili e quali vantaggi e svantaggi possiedono. Il quarto capitolo descrive i dataset analizzati, mentre nel quinto capitolo vengono presentati il pre-processing fatto sui dati e le successive analisi statistiche condotte. Nel sesto capitolo vengono esposti e discussi i risultati. Infine, vengono fatte delle brevi considerazioni finali sui risultati ottenuti.

# Capitolo 1

## I neuro-recettori

In questo capitolo si spiega brevemente cosa sono e come funzionano i neuro-recettori in generale, entrando poi nello specifico per i recettori per i quali è stato condotto lo studio, cioè o recettori degli oppioidi e della serotonina.

I recettori sono delle particolari proteine situate nelle membrane cellulari; in particolare i neuro-recettori sono proteine trans-membrana, cioè possiedono un dominio intracellulare, uno transmembrana e uno extracellulare. Ai recettori si legano reversibilmente specifiche molecole chiamate ligando, che una volta accoppiate causano una modifica nella conformazione del recettore, la quale provoca una risposta cellulare.

Sono situati in tutte le membrane pre e post-sinaptiche dei neuroni con sinapsi chimiche. In questo tipo di sinapsi il potenziale d'azione che scorre lungo l'assone del primo neurone, arriva alla membrana pre-sinaptica e scatena il rilascio del mediatore nello spazio sinaptico. Nella membrana post-sinaptica il ligando mediatore (neuro-trasmittitore) si lega a suoi specifici recettori (appartenenti ad una delle due classi sopra descritte) e scatena la risposta cellulare nel secondo neurone.

In base al tipo di risposta i recettori trans-membrana possono essere classificati in due famiglie ben distinte:

- i recettori *ionotropici*, che rendono possibile un flusso di ioni attraverso la membrana: quando il rispettivo ligando si lega, avviene una variazione nella conformazione strutturale del recettore che apre il canale attraverso il quale

gli ioni fluiranno. Alcuni esempi importanti di questa classe sono i recettori nicotinici e del GABA;

- i recettori *metabotropici*, che causano una serie di reazioni intracellulari a cascata quando il loro ligando si lega. Tali reazioni sono veicolate da un secondo messaggero, caratteristico di questo tipo di recettori, e determinano quella che viene definita trasduzione di segnale nelle cellule. Questo fenomeno è alla base della modifica dell'espressione genica, infatti il secondo messaggero può indurre la sintesi di proteine regolatorie che modificano l'attività delle RNA polimerasi e quindi l'espressione genica. A questa classe appartengono recettori come quelli serotoninergici, dopaminergici e oppioidi.

Le molecole che interagiscono con i recettori possono essere distinte in due categorie, in base al tipo di risposta che provocano:

- *agonisti*, molecole che si legano in maniera specifica al recettore e mimano il comportamento del ligando endogeno, andando in competizione con quest'ultimo. La risposta indotta nel recettore da queste molecole può essere minore, uguale o superiore rispetto alla risposta causata dal ligando endogeno.
- *antagonisti*, molecole che si legano specificamente con un recettore senza però attivarlo, e quindi senza provocare la normale risposta biologica che provocherebbe il ligando endogeno. Gli antagonisti si legano al recettore, bloccando il sito di legame ed impedendo quindi la normale attività degli agonisti endogeni e non.

## 1.1 I recettori degli oppioidi

Sono neuro-recettori della classe dei recettori metabotropici, e sono sensibili ai peptidi oppioidi (es. endorfine). Sono recettori esclusivamente inibitori, come si può vedere in Figura 1.1 una volta attivati riducono la produzione di AMP ciclico, abbassano il flusso di ioni  $\text{Ca}^{2+}$  ed aumentano la fuoriuscita di ioni  $\text{K}^+$ ,

così facendo si riduce l'eccitabilità della cellula (conseguentemente il rilascio di neuro-trasmittitori) e quindi l'attività nervosa. Lo scopo principale di riduzione è l'inibizione del dolore in diversi livelli ed aree cerebrali.

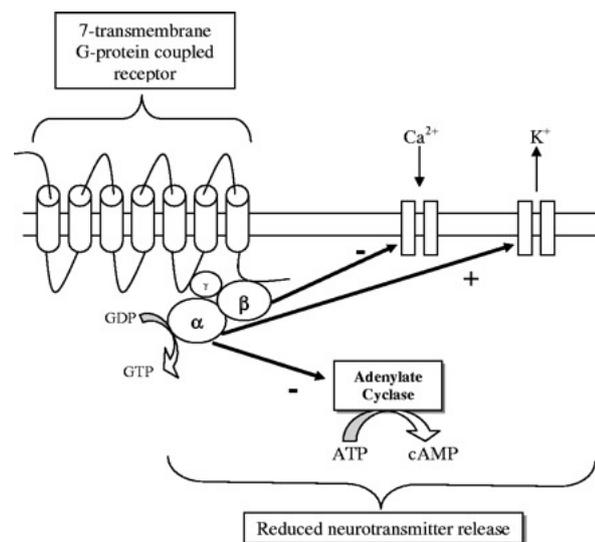


Figura 1.1: Schema delle reazioni causate dall'attivazione dei recettori degli oppioidi da parte di un agonista [5]

Si possono trovare distribuiti in tutto il cervello, nel midollo spinale e nel tratto digestivo. Ci sono tre tipi di recettore per gli oppioidi [6] [7] :

- $\delta$  o *DOP*: si trovano principalmente nella corteccia cerebrale, nell'amigdala, nei bulbi olfattivi, nel nucleus accumbens e nei neuroni sensoriali periferici. L'attivazione di questi recettori aumenta il flusso di  $K^+$ , produce effetti analgesici (in misura inferiore rispetto ai recettori  $\mu$ ), antidepressivi e convulsivi. Agonisti per questi recettori generano depressione respiratoria e riducono la mobilità gastrointestinale.
- $\kappa$  o *KOP*: sono situati prevalentemente nell'ipotalamo, nel claustrum, nella sostanza grigia periacqueduttale, nel midollo spinale e nei neuroni sensoriali periferici. Diminuiscono il flusso di ioni  $Ca^{2+}$ , hanno funzione analgesica, anticonvulsiva e hanno effetti sedativi. Il principale vantaggio degli agonisti

per questi recettori è che non causano depressione respiratoria, vengono inoltre utilizzati in ambito terapeutico per la cura nella dipendenza da droghe [8].

- $\mu$  o *MOP*: sono i più diffusi, si possono trovare nella corteccia cerebrale, nel talamo, nella sostanza grigia periacqueduttale, nel corpo striato, nel midollo rostrale, nel midollo spinale, nei neuroni sensoriali periferici e nel tratto intestinale. Hanno una forte funzione analgesica e provocano depressione respiratoria e miosi, ma al contrario di  $\kappa$  generano euforia; hanno inoltre effetti sul sistema cardiovascolare, nella termoregolazione, nella secrezione degli ormoni e nel sistema immunitario. In ambito medico il loro agonista più noto ed usato è la morfina.

## 1.2 I recettori della serotonina

Appartengono anch'essi alla classe dei neuro-recettori metabotropici (ad eccezione di  $5-HT_3$  che è ionotropico), si trovano sulla membrana di molte cellule, non solo quelle nervose. Mediano sia la neuro-trasmissione eccitatoria che inibitoria e il loro ligando naturale è la serotonina. La loro attività regola il rilascio di molti neurotrasmettitori come il GABA, la dopamina, l'epinefrina e l'acetilcolina e ormoni come l'ossitocina, la prolattina e il cortisolo. Questi recettori contribuiscono in moltissimi aspetti come l'appetito e la termoregolazione. I recettori serotoninergici [9] sono sette, e vengono denominati con la dicitura  $5-HT_{\#}$  ( $\#$  è un numero da 1 a 7). Tranne  $5-HT_1$  e  $5-HT_5$  che sono inibitori, gli altri cinque hanno tutti funzione eccitatoria.

Il  $5-HT_{1A}$  è il neuro-recettore più diffuso in tutto il corpo, tra i recettori della serotonina. Si può trovare in alte concentrazioni nella corteccia cerebrale, nell'amigdala, nell'ippocampo e nei nuclei di Raphe, e in bassi livelli anche nel talamo. Regolano ad esempio l'aggressività, l'ansia, l'appetito, la frequenza cardiaca e la pressione sanguigna, in particolare per queste ultime due la gestione avviene a diversi livelli, inducendo la vasodilatazione periferica e stimolando il nervo vago. La vasodilatazione periferica favorisce la dissipazione di calore abbassando così la temperatura corporea. Questi recettori sono anche bersaglio di alcune farmaci

che funzionano da parziali agonisti, e sono usati nella cura della schizofrenia e del morbo di Parkinson. Invece si è visto che degli antagonisti per tali recettori facilitano l'apprendimento e la memoria nei roditori, portando quindi allo studio nuovi trattamenti per il morbo di Alzheimer. Infine, le varianti geniche della sequenza di nucleotidi che codifica per questi neuro-recettori sono tuttora soggetto di studi per il trattamento di disturbi psichiatrici di varia natura.



# Capitolo 2

## Imaging funzionale: la PET

In ambito diagnostico e sperimentale la PET (*Positron Emission Tomography*) è considerata una tecnica di imaging funzionale, che permette di produrre immagini delle funzioni biologiche del corpo umano.

### 2.1 Principi di funzionamento

La PET [10] è una tecnica molto usata in medicina nucleare, sfrutta molecole biologiche normalmente presenti nell'organismo opportunamente modificate per emettere radiazioni. Queste molecole radioattive vengono definite *traccianti* perchè possono essere usate per seguire e studiare l'attività del metabolismo cellulare in ogni parte dell'organismo. La modifica effettuata in queste molecole consiste nella sostituzione di uno degli atomi che la compongono con un suo isotopo (stesso numero atomico  $A$ , ma diverso numero di massa  $Z$ ) radioattivo, definito radionuclide. Questi isotopi sono prodotti mediante l'uso di un ciclotrone (acceleratore di particelle) poco prima del loro utilizzo. L'uso dell'acceleratore di particelle è dovuto al fatto che la modifica a queste molecole perde la sua efficacia in breve tempo (da alcuni minuti a circa due ore), i traccianti cioè hanno una breve emivita (o tempo di dimezzamento, cioè tempo richiesto perché la metà dei nuclei radioattivi instabili decada in nuclei stabili) e quindi per essere utili devono essere usate subito dopo la loro creazione. Dato il tipo di funzionamento questa tecnica può essere usata, ovviamente, solo su organismi viventi. L'emis-

sione radioattiva viene misurata dall'esterno e permette di ricostruire l'esatta distribuzione spazio-temporale del tracciante nel tessuto, e quindi ricostruire un volume 3D in cui i valori di ogni voxel possono variare nel tempo.

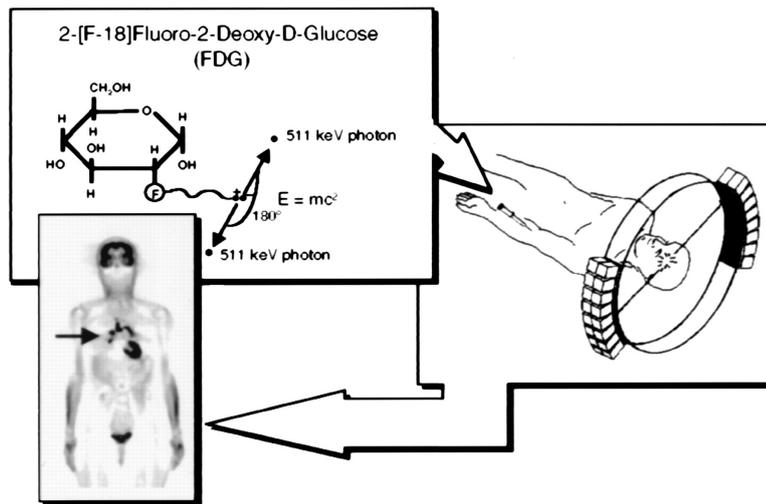


Figura 2.1: Principio di funzionamento della PET. Il tracciante usato nell'esempio è il  $^{18}\text{F}$ FDG, un analogo del glucosio. [11]

La radiazione emessa dai traccianti PET è dovuta al decadimento  $\beta^+$  in cui gli isotopi di uno degli elementi che compone la molecola di tracciante decadono emettendo positroni  $\beta^+$  (da qui il nome *Positron Emission Tomography*). I positroni emessi ionizzano il tessuto che attraversano, ma annichilano praticamente subito (si parla di tempi nell'ordine di  $10^{-9}\text{s}$ ) con il primo elettrone che trovano nel loro percorso. L'annichilazione porta all'emissione di due fotoni  $\gamma$ , come si può vedere nel primo riquadro della Figura 2.1, alla specifica energia di 511 KeV, che si muovono nella stessa direzione ma con verso opposto rispetto al punto di annichilazione. Una serie di rilevatori posti su anelli coassiali intorno al soggetto si incarica di "catturare" questa coppia di fotoni. Grazie al fatto che i due fotoni si muovono nella stessa traiettoria, ma in direzioni opposte, misurando la differenza di tempo che intercorre tra la rilevazione dei singoli fotoni di ogni coppia si può risalire alle coordinate spaziali in cui è avvenuto il singolo evento di annichilazione. Il numero di collisioni tra positroni ed elettroni, e quindi il numero di coppie di fotoni lette, è proporzionale alla quantità di tracciante presente in quella zona.

Per discriminare tra le vere coppie provenienti da uno stesso evento e altri fotoni rilevati si usa un circuito detto *circuito di coincidenza*.

Per i traccianti vengono usati isotopi di elementi che costituiscono la materia biologica, come l'ossigeno, il carbonio, l'azoto o il fluoro. Un tipico tracciante, molto usato in neurologia ed oncologia come analogo del glucosio, è il *fluorodeossiglucosio* [ $^{18}\text{F}$ ]FDG in cui l'isotopo radioattivo è il fluoro  $^{18}\text{F}$ . Oppure, ad esempio, negli studi sul sistema neuro-recettoriale, possono essere usati farmaci specifici ideati per legarsi ai recettori voluti e modificati con un isotopo radioattivo in modo da fungere da tracciante per generare mappe parametriche dei neuro-recettori, come nel caso del tracciante [ $^{11}\text{C}$ ]DPN applicato per gli studi dei recettori degli oppioidi [12].

## 2.2 Quantificazione delle immagini PET

Quando nell'imaging funzionale si parla di "quantificazione" si intende una descrizione numerica, e quindi un'analisi puramente oggettiva, di quanto funziona un determinato organo o zona anatomica usando la metodica di indagine più appropriata a descrivere tale fenomeno.

La quantificazione di immagini PET richiede l'applicazione di modelli matematici adatti al tipo di tracciante usato e che permettano il calcolo di parametri fisiologici di interesse, producendo come risultato mappe parametriche. In base al modello usato il numero di parametri stimati varia, e le mappe parametriche prodotte sono tante quanti sono tali parametri.

Per procedere alla quantificazione sono richieste la misura dell'attività tissutale, o TAC (*Time Activity Curve*), che si ricava dalle immagini PET, e un campionamento invasivo da cui si ricava la concentrazione arteriale del tracciante ( $C_p$ ). Dopo aver ripulito la misura plasmatica dalla presenza di metaboliti (prima della correzione, parte del tracciante nel plasma è legato a scarti metabolici) si applica il modello matematico adatto a quantificare il tracciante in esame, stimando poi i parametri del modello.

Indipendentemente dal metodo usato, la quantificazione di immagini PET può essere quasi sempre fatta su due diversi livelli dal punto di vista dell'analisi delle immagini:

**A livello di ROI** Le immagini vengono valutate considerando varie gruppi di pixel costituenti le varie zone anatomiche di interesse, chiamate regioni di interesse o ROI (*Region Of Interest*). Le TAC hanno un buon rapporto segnale/rumore e, dato il ridotto numero di regioni da analizzare, è possibile usare metodi computazionalmente onerosi come WNLLS (minimi quadrati non lineari pesati) per stimare i parametri. Di contro si ha una perdita nella risoluzione spaziale delle immagini perchè si stimano parametri per delle macro-regioni invece che per i singoli pixel.

**A livello di Pixel** Le immagini vengono valutate pixel per pixel, le curve di attività tissutale presentano un rapporto segnale/rumore più basso di quello a livello ROI, ma la risoluzione spaziale resta inalterata, uguale a quella dell'immagine originale. Tuttavia, valutando un pixel alla volta, la mole di dati da elaborare è molto elevata e quindi metodi come WNLLS non possono essere utilizzati.

Nel corso degli anni sono stati sviluppati vari metodi e modelli matematici per ricavare le informazioni fisiologiche da immagini PET in base al tipo di tracciante usato e ai parametri di interesse. I metodi applicati per i traccianti di questa tesi sono descritti nei paragrafi successivi.

### 2.2.1 Tracciante [ $^{11}\text{C}$ ]DPN

Questo tracciante, anche definito [ $^{11}\text{C}$ ]-*Diprenorfina*, è un radio-ligando antagonista reversibile per recettori degli oppioidi, non selettivo per i tre sottotipi. Presenta alta affinità per i recettori  $\mu, \kappa, \delta$ , agendo da antagonista per i sottotipi  $\mu$  e  $\delta$ , e da parziale agonista per i recettori  $\kappa$ . Il metodo usato [12] per quantificare questo tracciante consiste prima in un'analisi cinetica con modello mono-compartimentale a livello di ROI e poi in una stima MAP (*Maximum A Posteriori*) applicata a livello di pixel, che usa le stime trovate precedentemente a livello ROI come *prior*. Per definire le ROI su cui effettuare l'analisi è stata effettuata una segmentazione del volume cerebrale usando il *Maximum Probability Hammersmith Brain Atlas* [13]

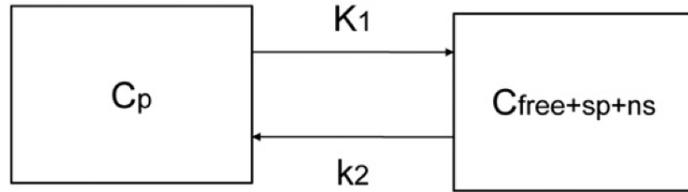


Figura 2.2: Modello mono-compartimentale (anche chiamato 1T)

Il modello compartimentale usato è quello rappresentato in Figura 2.2 e descritto dalle seguenti equazioni:

$$\dot{C}_T(t) = K_1 C_p(t) - k_2 C_T(t) \quad C_T(0) = 0 \quad (2.1)$$

$$C_{VOI}(t) = (1 - V_b) C_T(t) + V_b C_b(t) \quad (2.2)$$

dove  $C_p$  rappresenta la concentrazione di tracciante nel compartimento plasmatico e  $C_T$  la concentrazione nel compartimento tissutale (incorpora la frazione di recettori liberi, di recettori non specifici e di recettori specifici legati).  $C_b$  è la concentrazione di [ $^{11}\text{C}$ ]DPN nel sangue e  $V_b$  è la frazione di volume ematico nella ROI.  $K_1$  e  $k_2$  sono i coefficienti di trasferimento, che permettono di determinare il volume di distribuzione  $V_d$ , concettualmente definito come il rapporto, in stato

stazionario, tra la concentrazione di tracciante nel tessuto e la concentrazione di tracciante nel plasma ripulito dai metaboliti. Nel caso mono-compartimentale:

$$V_T = \frac{K_1}{k_2} \quad [mL/g] \quad (2.3)$$

Si sono stimati i  $k_i$  risolvendo il modello a livello di ROI usando i minimi quadrati non lineari (compito computazionalmente poco oneroso perché fatto su un numero limitato di regioni). Successivamente, si è linearizzato il modello ottenendo:

$$C_T(t) = K_1 \int_0^t C_P(\tau) d\tau - k_2 \int_0^t C_T(\tau) d\tau \quad (2.4)$$

Considerando la componente ematica nulla a livello di pixel, è ora possibile usare uno stimatore MAP, dove si può usare come prior le stime dei  $k_i$  determinate a livello di ROI. La funzione costo dello stimatore è:

$$\mathbf{J}(\mathbf{p}) = [\mathbf{C}_T - \mathbf{G}(\mathbf{p})]^T \Sigma_V^{-1} [\mathbf{C}_T - \mathbf{G}(\mathbf{p})] + (\mathbf{p} - \mu_p)^T \Sigma_p^{-1} (\mathbf{p} - \mu_p) \quad (2.5)$$

dove, con  $n$  numero di misure tissutali,  $\mathbf{C}_T$  è il vettore  $n \times 1$  delle misure affette da rumore,  $\mathbf{G}(\mathbf{p})$  è l'uscita  $n \times 1$  predetta del modello,  $\Sigma_V$  è la matrice  $n \times n$  di covarianza dell'errore delle misure,  $\mathbf{p}$  è il vettore  $2 \times 1$  dei parametri da stimare e  $\mu_p$  e  $\Sigma_p$  sono rispettivamente il vettore  $2 \times 1$  media e la matrice  $2 \times 2$  di covarianza del prior. Dalla (2.5), i parametri da stimare si trovano risolvendo l'equazione:

$$p_{MAP} = (G^T \Sigma_V^{-1} G + \Sigma_p^{-1})^{-1} (G^T \Sigma_V^{-1} C_T + \Sigma_p^{-1} \mu_p) \quad (2.6)$$

con:

$$G = \begin{bmatrix} \int_0^{t_1} C_p(\tau) d\tau & - \int_0^{t_1} C_T(\tau) d\tau \\ \vdots & \vdots \\ \int_0^{t_n} C_p(\tau) d\tau & - \int_0^{t_n} C_T(\tau) d\tau \end{bmatrix} \quad (2.7)$$

Essendo che la stima MAP qui usata è lineare il costo computazionale è molto basso anche a livello di pixel. Infine il  $V_d$  si è calcolato con (2.3) usando i nuovi parametri  $k_i$  stimati con MAP.

### 2.2.2 Tracciante [ $^{11}\text{C}$ ]CUMI

Il tracciante [ $^{11}\text{C}$ ]CUMI-101 è un radio-ligando altamente selettivo per i recettori  $5\text{-HT}_{1A}$  che agisce da parziale agonista per quest'ultimi. La metodica utilizzata per la quantificazione è il *metodo grafico di Logan* [14], un metodo che si applica ai modelli reversibili (modelli in cui il tracciante non si lega in maniera irreversibile ad un recettore), che non permette di stimare tutti i parametri, ma solo il macro-parametro  $V_d$  ( $\frac{ml_{plasma}}{gr_{tessuto}}$ ). Per applicare questo metodo si fanno le seguenti ipotesi sul modello:

1. non ci sono compartimenti irreversibili
2. il tracciante lascia i compartimenti reversibili attraverso il plasma
3. la cinetica del tracciante è descritta da equazioni differenziali lineari, del primo ordine e a coefficienti costanti
4. se il tracciante viene metabolizzato dall'organismo, i prodotti metabolici sono misurabili
5. la concentrazione iniziale di tracciante nel tessuto è nulla
6. *ipotesi aggiuntiva*: si suppone che esista un istante  $t^*$  tale che per  $t > t^*$  (vedi Figura 2.3) tutti i compartimenti siano in equilibrio, cioè che il rapporto tra la concentrazione  $C_i$  nel compartimento  $i$ -esimo e la concentrazione plasmatica  $C_p$  sia costante

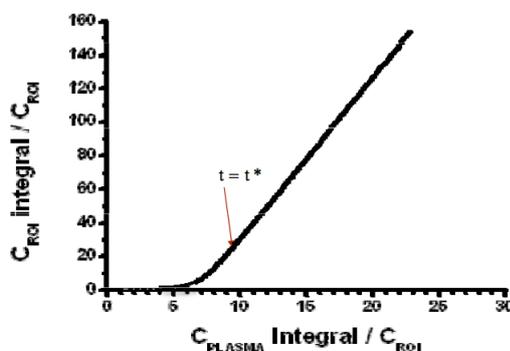


Figura 2.3: Andamento dei rapporti di concentrazione nel metodo di Logan

Allora per  $t > t^*$  l'andamento è lineare e può essere rappresentato come una retta, in cui  $x$  e  $y$  sono definite come:

$$x = \frac{\int_0^t C_p(\tau) d\tau}{C_{misurato}(t)} \quad y = \frac{\int_0^t C_{misurato}(\tau) d\tau}{C_{misurato}(t)} \quad (2.8)$$

La pendenza di questa retta allora coincide con  $V_d$ , il volume di distribuzione. Si procede quindi considerando tutti i valori di  $x$  e  $y$  dopo l'istante  $t^*$  e si usano i minimi quadrati lineari per stimare pendenza ( $V_d$ ) e intercetta.

Nel caso degli studi del sistema neuro-recettoriale si può usare l'informazione data dalle mappe del volume di distribuzione del tracciante per generare mappe parametriche che descrivono il *potenziale di legame* (BP) dei recettori per i quali il tracciante usato si lega. Definiti i parametri:

- $B_{MAX}$  massimo numero di recettori affini presenti nel pixel
- $1/K_d$  affinità del tracciante per i recettori (costante)

Concettualmente il BP può essere definito come:

$$BP = \frac{B_{MAX}}{K_d} \quad (2.9)$$

Essendo l'affinità del tracciante per i neuro-recettori costante per il tracciante usato, variazioni del potenziale di legame sono proporzionali a variazioni del numero di recettori in quel pixel. Quindi il BP è un buon parametro per discriminare se in una determinata regione anatomica c'è un alto numero di recettori affini al tracciante studiato.

Se è nota una zona anatomica in cui non ci sono recettori specifici per il sistema studiato, questa può essere usata come regione di riferimento (*ref*). Definendo  $V_d^{ref}$  come il volume di distribuzione della regione di riferimento, il potenziale di legame può essere calcolato come:

$$BP = \frac{V_d - V_d^{ref}}{V_d^{ref}} \quad (2.10)$$

### 2.2.3 Tracciante [ $^{11}\text{C}$ ]WAY

E' un tracciante che agisce da antagonista selettivo per i recettori  $5\text{-HT}_{1A}$ . Per la quantificazione di questo tracciante si è proceduto con una tecnica chiamata *Spectral Analysis* [15], un metodo alternativo all'uso dei modelli compartimentali. Si interpreta il modello come un sistema lineare SISO (*Single Input Single Output*) in cui la risposta impulsiva è del tipo:

$$h(t) = \sum_{i=1}^M \alpha_i \cdot e^{-\beta_i t} \quad \alpha_i, \beta_i \geq 0 \quad (2.11)$$

La concentrazione di tracciante nel voxel diventa allora la convoluzione dell'ingresso del sistema ( $C_p(t)$ ) con la risposta impulsiva:

$$C_{\text{voxel}}(t) = C_p(t) \otimes h(t) = \sum_{i=1}^M \alpha_i \cdot \int_0^t C_p(\tau) e^{-\beta_i(t-\tau)} d\tau \quad \beta_i \geq 0 \quad (2.12)$$

Il parametro  $M$  corrisponde al numero di compartimenti con cui si vuole modellizzare il sistema, e gli  $\alpha_i$ ,  $\beta_i$  ( $i = 1, \dots, M$ ) da stimare sono i parametri dell' $i$ -esimo compartimento, e si assumono essere a valori reali e non negativi.

Una volta scelto il valore per  $M$  (solitamente viene scelto un valore alto, tra 100 e 200), i valori dei  $\beta_i$  sono predeterminati e fissati per coprire un'opportuna gamma spettrale. Infine gli  $\alpha_i$  possono essere stimati dalla curva dell'attività tissutale (TAC) con NNLS (*nonnegative least squares*).

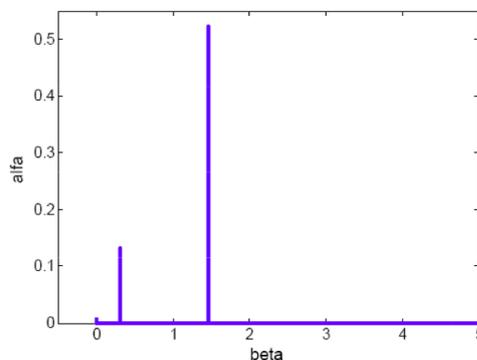


Figura 2.4: Esempio di grafico dei parametri nella Spectral Analysis

Il nome di questa tecnica deriva dalla rappresentazione che si può adottare per visualizzare i parametri, infatti interpretando i  $\beta_i$  e gli  $\alpha_i$  come ascisse e ordinate si avrà un grafico che visivamente assomiglia ad uno spettro in frequenza (Figura 2.4).

Degli  $M$  valori di  $\alpha_i$  che si stimeranno, molti saranno nulli e solo alcuni saranno diversi da zero. Questo fatto può essere utilizzato per capire il numero di compartimenti necessari per modellizzare il sistema [16], infatti per ogni  $\alpha_i \neq 0$  si ha un compartimento. Inoltre se  $\alpha_1 \neq 0$  allora è presente un compartimento irreversibile, mentre per  $i = 2, \dots, M$  per ogni  $\alpha_i \neq 0$  si avrà un compartimento reversibile. Infine, il volume di distribuzione si può poi calcolare come rapporto tra gli  $\alpha_i$  ed i  $\beta_i$ :

$$V_d = \sum_{i=2}^M \frac{\alpha_i}{\beta_i} \quad (2.13)$$

## Capitolo 3

# Il DNA e la misura dell'espressione genica

Il DNA (*DeoxyriboNucleic Acid*) è una catena polimerica di origine organica formata da unità monomeriche chiamate nucleotidi. Com'è noto i nucleotidi hanno tutti la stessa conformazione chimica, a meno delle basi azotate che li caratterizza, tali basi sono quattro: adenina, citosina, guanina e timina. Normalmente il DNA si trova come strutture ad elica a doppio filamento, questo è possibile grazie alla complementarietà delle basi (le quattro basi sono complementari due a due), infatti in una stessa doppia elica un filamento è complementare dell'altro a cui è legato. L'informazione genetica è codificata nel DNA come sequenza dei quattro nucleotidi, infatti l'ordine in cui sono disposti lungo la catena determina la codifica genetica e l'informazione che caratterizza i geni. Nel DNA sono quindi immagazzinate tutte le informazioni necessarie per la vita e il comportamento della cellula, dalla produzione delle proteine alle regolazioni delle attività cellulari.

Un'altra molecola di importanza fondamentale per la vita è l'RNA, anch'esso una catena polimerica formata da nucleotidi come il DNA, ma si differenzia da quest'ultimo per diversi aspetti: lo zucchero da cui è costituito è il ribosio e non il desossiribosio, il nucleotide della timina è sostituito dall'uracile e si trova normalmente come singolo filamento. Se il DNA immagazzina le informazioni, l'RNA è responsabile della loro codifica e decodifica, regolazione ed espressione.

L'*espressione genica* è la trasformazione delle informazioni codificate nei geni in mRNA (RNA messaggero) e poi in proteine, o in RNA che ha il compito di esprimere una qualche funzione. L'*informazione genica* è uguale in tutte le cellule somatiche di un organismo, essa determina la struttura ed il funzionamento di tutte le proteine e di tutte le molecole producibili dalle cellule, mentre l'*espressione genica* determina il comportamento della singola cellula, rendendo possibile il differenziamento cellulare, cioè la funzione specifica di gruppi di cellule e quindi dei tessuti. Tale espressione può variare nelle cellule di uno stesso tessuto in base ai bisogni e al momento. In un certo senso si può dire quindi che l'*espressione genica* interpreta le informazioni contenute nei geni e nel DNA.

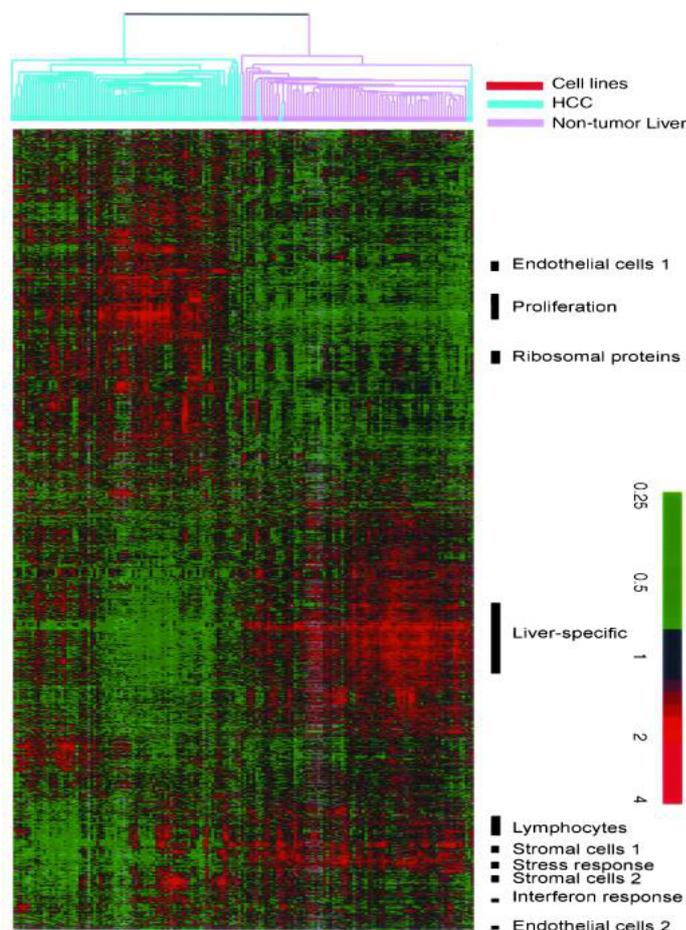


Figura 3.1: Variazione del profilo di espressione di 3180 geni in 156 campioni di tessuto epatico, di cui 74 sani e 82 con *HCC* (carcinoma epato-cellulare). Ogni riga rappresenta un gene e ogni colonna un tessuto o un campione cellulare. [17]

Conoscere l'espressione di un determinato gene ha grande importanza sia dal punto di vista puramente scientifico perché permette di comprendere al livello più intimo possibile i meccanismi che regolano le cellule, sia dal punto di vista medico perché tale conoscenze possono essere usate per sviluppare cure geniche mirate e comprendere meglio l'origine e la natura di moltissime malattie (Figura 3.1 a pag. 18). Per indagare se un gene è espresso o meno in un cellula si cercano quindi gli RNA e le proteine per i quali tale gene codifica. Il “quanto è espresso” dipende dalla concentrazione di questi elementi nella cellula. Per misurare tale concentrazione le biotecnologie hanno prodotto diversi strumenti: i più utilizzati in assoluto sono i *Microarray*.

### 3.1 I Microarray: principi di funzionamento

I *Gene Chip* o *Microarray* sono nati a metà degli anni '90 [18] e sono lo strumento di eccellenza usato per analizzare l'espressione genica in un campione biologico. Esistono molti tipi di array, che si differenziano per tecniche costruttive, efficienza nell'individuare sequenze e costi.

Gli array sono formati da una piccolo strato di materiale solito come vetro, plastica o silicio su cui sono immobilizzate migliaia di piccole sonde, o *probes*, diverse fino a formare un array non più grande di 1-2  $cm^2$ .

I *Microarray* possono essere utilizzati solo dopo il sequenziamento del genoma dell'organismo in esame, cioè quando si conosce l'esatta sequenza di nucleotidi di cui è composto il gene di cui si vuole monitorare l'espressione, infatti tutte le sonde sono costruite da una sequenza nota di nucleotidi, come nota è l'esatta posizione di ogni probe nel substrato di supporto. Le sonde vengono disegnate in modo tale da presentare una sequenza complementare al pattern che si vuole monitorare cosicché, una volta deposto il campione sul *Microarray*, l'RNA target (opportunamente trattato e marcato) vada a legarsi con la probe che presenta la sua sequenza complementare; questo comportamento è definito *ibridazione inversa*.

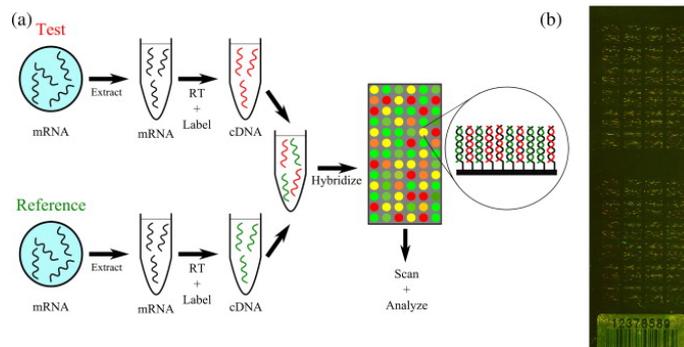


Figura 3.2: Set-up sperimentale per un *Microarray a due canali* [19] : (a) processo di estrazione e preparazione del campione; (b) immagine di un *Microarray a cDNA* di un *salmone Atlantico*.

Il procedimento di preparazione ed utilizzo di un *Microarray* descritto in Figura 3.2 si può riassumere in cinque fasi:

**Prima fase** Si prepara un *Microarray* già confezionato con un set di probe standard o lo si costruisce disegnando e immobilizzando le sonde volute sul substrato; aziende del settore hanno sviluppato diversi metodi di immobilizzazione, ognuno di questi metodi caratterizza il tipo di chip usato. Le tecniche usate sono principalmente due: sintesi *in-situ* di oligonucleotidi e deposizione mediante stampa a getto (*ink-jet*) di sonde pre-sintetizzate, sempre di oligonucleotidi o di frammenti di *cDNA*.

**Seconda fase** Si prepara il campione biologico seguendo un procedimento è abbastanza semplice:

- estrazione e purificazione dell'RNA contenuto nel campione
- retro-trascrizione in *cDNA* (*complementaryDNA*: molecole di DNA a doppio filamento ricavate utilizzando l'enzima trascrittasi inversa su un frammento di *mRNA*) ed etichettatura con marcatori fluorescenti, o generazione di catene di oligonucleotidi
- deposizione sul *Microarray* per l'ibridazione e lavaggio

**Terza fase** Si generano le immagini eccitando con un laser i fluorofori usati per marcare le sonde: l'intensità luminosa determina quanto una probe ha ibridato, e quindi quanto il gene associato è espresso. Se per marcare le probe si è usato un solo marcatore si parla di *Microarray a canale singolo*, se si sono usati due marcatori diversi (come in Figura 3.2 a pag. 20) allora si parla di *Microarray a due canali*. Nel primo caso gli array indicano il livello di ibridazione relativo della probe rispetto al target, nel secondo caso si misura l'intensità relativa di due fluorofori diversi, uno per il *test* e uno per il *reference*, per misurare ad esempio l'*up-regolazione* e la *down-regolazione* di un gene.

**Quarta fase** Si analizzano le immagini per “pulire” i dati dal rumore e per rimuovere tutte le probe che hanno avuto problemi durante l'ibridazione.

**Quinta fase** Si passa dai dati grezzi delle immagini alla quantificazione vera e propria dell'espressione applicando diverse tecniche: dalla correzione dello sfondo alla normalizzazione dei dati. Il risultato è la mappa del profilo di espressione del campione esaminato. I valori di questa mappa sono normalmente espressi in  $\log_2$ : definiti  $S_{Cy5}$  e  $S_{Cy3}$  i segnali di intensità dei due fluorofori dopo la sottrazione dello sfondo si considera il logaritmo in base due del loro rapporto, si parla cioè di *log-ratio LR*:

$$LR = \log_2(S_{Cy5}) - \log_2(S_{Cy3}) = \log_2\left(\frac{S_{Cy5}}{S_{Cy3}}\right) \quad (3.1)$$

questo permette di avere simmetria tra i casi in cui il campione è più espresso nel *test* che nel *reference* e i casi in cui succede il contrario. Oltre al  $\log_2$ , i dati possono essere espressi come *z-score*, cioè come valori a media ( $\mu$ ) zero e deviazione standard ( $\sigma$ ) unitaria, con la formula:

$$z\text{-score} = \frac{\text{ratio} - \mu}{\sigma} \quad (3.2)$$

Data l'elevata densità dei *Microarray* attuali è necessario usare marcatori fluorescenti per poter leggere correttamente ogni spot. In sostanza quindi, andando a “leggere“ quante e quali sonde hanno ibridizzato si può risalire al livello di espressione di ogni gene. Con questo approccio, grazie alle moderne tecnologie, si possono immobilizzare sul substrato migliaia di sonde diverse rendendo possibile analizzare in parallelo decine di migliaia di sequenze in un solo array, virtualmente anche l'intero genoma di un organismo. A tal proposito, il profilo mostrato nell'esempio in Figura 3.1 a pag. 18 è stato ricavato con *Microarray a due canali a cDNA* contenenti circa 23 000 filamenti, rappresentanti circa 17 400 geni umani per il monitoraggio del cancro nei tessuti epatici.

### 3.1.1 Microarray a due canali - Agilent

I *Microarray a due canali* sono un tipo di array molto diffuso ed utilizzano nei laboratori di ricerca ormai da molti anni. Un'industria molto famosa nella produzione dei *Microarray* è l'*Agilent*, azienda che produce array sia con strategie ad un canale che a due canali. Entrambi i tipi di array sono prodotti con una tecnica chiamata *ink-jet*; questa tecnologia è stata creata prendendo spunto dal metodo di funzionamento delle comuni stampanti a getto di inchiostro (da cui il nome), infatti sfrutta la propulsione piezoelettrica per "stampare" le probe sul supporto. Il procedimento di costruzione è molto semplice:

1. un robot preleva le probe (di oligonucleotidi o *cDNA* pre-sintetizzati) per monitorare un singolo gene da una serie di pozzetti
2. attraverso un sistema di pennini mobili il robot "stampa" con un microgetto le probe sugli spot dell'array (ogni spot ha una dimensione nell'ordine di alcuni  $\mu\text{m}$ )
3. i pennini vengono lavati e si ricomincia dal punto 1. con un nuovo insieme di probe

Questa tecnica permette di immobilizzare per ogni spot circa  $10^4 - 10^6$  catene identiche sul supporto, per una densità di circa  $10\,000\text{ cDNA}/\text{cm}^2$ . Le probe che costituiscono questo genere di array possono essere composte da oligonucleotidi (corte molecole di DNA o RNA a singolo filamento fabbricate da sintetizzatori automatici) o da frammenti di *cDNA*. Usare oligonucleotidi pre-sintetizzati nei pozzetti permette di avere catene lunghe fino a 60-mer (mer: unità di misura che deriva dal greco *meros* che significa *parte*), mentre le catene di *cDNA* possono avere lunghezze variabili, ma prevedono un ulteriore trattamento dopo l'immobilizzazione sul supporto per denaturare il doppio filamento in un filamento singolo, prima della fase di ibridazione del campione biologico. Grazie alle lunghezze raggiungibili con questo tipo di *Microarray* si possono disegnare probe molto specifiche per individuare i geni bersaglio dell'analisi. La durata della fase di ibridazione però può durare fino ad 8-10 ore.

In Figura 3.3 si può vedere un *Microarray a due canali Agilent* in cui è immediato notare la presenza di spot di tre colori diversi: rosso, giallo e verde. Questo è

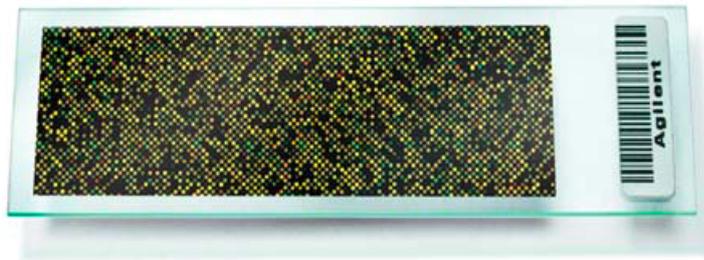


Figura 3.3: *Agilent SurePrint G3 Exon Microarray*, contiene le probe per più di 20 000 geni e 150 000 esoni umani. Array simili sono disponibili anche per animali comunemente usati nei laboratori, come il topo.

dovuto al fatto che comunemente nei *Microarray a due canali* i due fluorofori usati sono la Cianina 3 (*Cy3*) e la Cianina 5 (*Cy5*) che emettono radiazione luminosa con lunghezze d'onda nel campo del verde e del rosso rispettivamente. L'utilizzo di due fluorofori diversi permette di analizzare il campione soggetto dell'analisi (*test*) ed il campione di riferimento (*reference*) direttamente sullo stesso array. Le catene del campione *test* sono marcate con *Cy5* mentre quelle del *reference* con *Cy3*. Le soluzioni contenenti l'RNA del *test* e del *reference* vengono mescolate e deposte sul *Microarray*; se nei due campioni ci sono trascritti per uno stesso gene si avrà sull'array il fenomeno definito *ibridazione competitiva*, nel quale le due catene marcate con i due fluorofori diversi competeranno per legarsi alla stessa sonda. Se l'RNA che codifica per un certo gene è presente in maggior quantità nel campione *test* (quindi il gene è più espresso) rispetto al *reference* allora lo spot che contiene l'insieme di probe per quel gene ibriderà con più catene del primo che del secondo ed emetterà luce con una lunghezza d'onda verso il rosso. Se il suddetto RNA è più espresso nel *reference* allora lo spot emetterà radiazione luminosa con una lunghezza d'onda nel campo del verde. Gli spot gialli sono dovuti ad una combinazione di *Cy3* e *Cy5*, che avviene quando si ha una parità nella fase di ibridazione, cioè quando il gene monitorato da quello spot è espresso ugualmente sia nel *test* che nel *reference*.

La strategia a due canali presenta un vantaggio legato al costo, infatti si possono confrontare due campioni biologici (*test* e *reference*) sullo stesso *Microarray* in maniera competitiva invece di doverne usare due come si vedrà nella sezione

3.1.2, nel caso dei *Microarray ad un canale* e dei chip *Affymetrix*. Oltre per i costi, questo può essere un vantaggio anche dal punto di vista dell'esperimento perché permette di avere un confronto diretto tra *test* e *reference* in una situazione sperimentale perfettamente identica, di contro diventa meno importante se si volessero confrontare i dati tra più esperimenti con campioni diversi.

Invece, un problema legato ai *Microarray a due canali* è rappresentato dai fluorofori stessi usati, infatti *Cy3* e *Cy5* hanno un comportamento diverso: non presentano la stessa efficienza nell'emissione della radiazione luminosa e sono costituiti da molecole di dimensione diversa (*Cy5* è più ingombrante di *Cy3*) e quindi spazialmente, nella catena di nucleotidi, occupano un volume diverso, andando ad incidere nell'efficienza della fase di ibridazione. Attualmente si ovvia al problema dell'efficienza di emissione effettuando una correzione via software dei dati rilevati, mentre per il problema della diversa dimensione si usa la tecnica denominata *dye-swap* che consiste nel ripetere l'esperimento scambiando i fluorofori per *test* e *reference* andando poi a confrontare i risultati con quelli del primo esperimento.

### 3.1.2 Microarray ad un canale - Affymetrix

I chip *Affymetrix* sono un tipo di *Microarray ad un canale* molto diffuso ed utilizzato. Sono disponibili molti modelli per un grande varietà di organismi viventi, oltre a mammiferi anche invertebrati, piante e pesci. Questi array sono creati immobilizzando sulla superficie catene di oligonucleotidi sintetici invece di *cDNA*. La più importante caratteristica che li differenzia rispetto agli altri *Microarray* in commercio è che gli oligonucleotidi vengono sintetizzati su un supporto di silicio utilizzando tecnologie fotolitografiche [20] simili a quelle usate nell'industria dei processori per computer (da qui il nome *Gene Chip*).



Figura 3.4: *Affymetrix GeneChip Gene 1.0 ST Array*, chip di questo tipo sono disponibili per molti organismi: dal *C. elegans* al pollo, dal riso allo *Zebrafish*.

L'azienda detiene il brevetto di questa tecnica che permette di costruire chip ad altissima densità, riuscendo ad immobilizzare sul supporto fino a 500 000 *oligonucleotidi/cm<sup>2</sup>*, sintetizzando per ogni spot (gruppo di sonde progettato per uno specifico trascritto) sull'array circa  $10^9$  sequenze identiche di oligonucleotidi. Usare la fotolitografia permette la sintesi *in-situ* delle catene di nucleotidi sintetici che andranno a costituire le varie probe.

La costruzione delle sonde avviene con ripetizione in successione dei seguenti passi:

1. deposizione del primo strato di nucleotidi, la cui testa è protetta da una molecola fotosensibile
2. posizionamento di una maschera forata per permettere di irradiare selettivamente le zone di interesse
3. irraggiamento con una sorgente di ultravioletti, i nucleotidi non protetti dalla maschera perdono la molecola fotosensibile
4. deposizione del successivo strato di nucleotidi, i quali si legheranno solo a quelli dello strato precedente rimasti senza molecola protettiva
5. si ripete dal punto 2. variando la posizione dei buchi nella maschera ed i nucleotidi depositi per arrivare ad avere le probe con le sequenze desiderate

Sviluppando maschere adeguate e prestando attenzione alla sequenza in cui si aggiungono i nucleotidi si possono sintetizzare in contemporanea le probe per molte centinaia di geni, producendo chip in grado di monitorare interi genomi. Questa densità è tale da permettere anche di evitare fenomeni di saturazione e rendere l'analisi molto sensibile, infatti l'alto numero di probe posizionabili permette di immobilizzare, per ogni spot, sonde che differiscono anche solo di poche basi, cosicché se una non ibridizza lo farà un'altra simile.

Altra differenza rispetto ad altre tecniche è che questa tecnologia genera catene di circa 20-25 oligonucleotidi di lunghezza, una quantità molto inferiore in confronto ad altre tecnologie, tipo quella *ink-jet*, usate nei *Microarray a due canali*. Questa ridotta lunghezza delle probe rende i chip meno specifici rispetto a quelli con sonde più lunghe, e per questo il chip prevede di immobilizzare due tipi di probe per ogni gene da monitorare: per quelle del primo tipo vengono ideate 10-20 sequenze diverse di probe e ogni sonda di questo set è definita *perfect-match* (PM), il secondo tipo di probe contiene sonde identiche ad ogni *perfect-match* tranne per l'oligonucleotide centrale della catena, cambiato in modo da non legarsi in maniera specifica con il gene monitorato. Queste ultime catene vengono definite sequenza *mismatch* (MM), e permettono di monitorare le cross-ibridazioni e avere un controllo negativo sugli appaiamenti. L'insieme dei *perfect-match* e dei *mismatch* viene detto *probe-set* di un gene.

Al contrario dei *Microarray a due canali*, per misurare l'espressione di un *test* rispetto ad un *reference* l'ibridazione non può essere fatta sullo stesso array nel caso degli *Affymetrix* e quindi per ogni analisi bisogna preparare due *Microarray*, uno per i campioni e uno per il controllo, e successivamente fare un rapporto tra le intensità. Nel caso del *test* (per il *reference* è uguale), l'RNA target viene estratto, retro-trascritto in *cDNA* con l'enzima *trascrittasi inversa* e ritrascritto in RNA complementare legato a molecole di *biotina*, infine le catene vengono frammentate in filamenti più corti per poter legare con le probe corte dei *Gene Chip* e il tutto viene messo a contatto con l'array in un ambiente sterile ad umidità e temperatura controllati per alcune ore. Ad ibridazione avvenuta il *Microarray* viene lavato per eliminare i frammenti che non hanno legato, e viene aggiunto il marcatore fluorescente che lega la biotina (l'aggiunta del fluoroforo viene fatta in questa fase perché così la sua presenza non interferisce nella fase di ibridazione). Successivamente avviene tutto come nei *Microarray* normali, si eccitano gli spot con il laser e uno scanner misura l'intensità per creare le immagini da analizzare. Le matrici risultanti saranno "mono-cromatiche", infatti in questi *Microarray* le misure di espressione sono assolute e l'intensità di ogni spot dipenderà da come hanno ibridato i PM e gli MM per il gene a cui lo spot si riferisce. Se questo gene è molto espresso i suoi trascritti si legheranno in maniera più forte alle probe PM rispetto alla media delle probe MM. Infine, bisogna prestare attenzione al fatto che la misura di intensità è sì assoluta rispetto all'array, ma non esprime il livello assoluto di espressione di un gene; il valore misurato è solo un livello di espressione che si può comparare con un altro campione o con lo stesso campione in condizioni sperimentali diverse.

I chip *Affymetrix*, e più in generale i *Microarray ad un canale*, presentano alcuni vantaggi, infatti nel caso di un campione "falsato" questo non rischierà di compromettere i dati provenienti dagli altri campioni proprio perché ogni campione è analizzato separatamente. Cioè, ad esempio, se il *test* è "falsato" si può ripetere l'analisi sul quest'ultimo senza dover rivalutare anche il *reference*. Altro vantaggio è che usando un solo "colore" non ci sono problemi per quanto riguarda l'ingombro delle molecole di fluorofori che nel caso dei *Microarray a due canali* presentano due dimensioni diverse, pesando quindi in modo diverso nel processo di ibridazione.

In questi array si usa solo la biotina e quindi il comportamento è uguale per tutti i campioni. Se la dimensione ridotta delle probe, e quindi la bassa specificità dell'array, non viene considerata come un punto a sfavore, l'unico ovvio svantaggio per la strategia ad un canale è che questa raddoppia esattamente i costi dell'esperimento.

Infine una tabella comparativa tra i *Microarray* Agilent e Affymetrix:

<i>Caratteristiche</i>	<i>Agilent</i>	<i>Affymetrix</i>
Strategia ad un canale	V	V
Strategia a due canali	V	X
Basso costo	V	X
Alta densità	X	V
Confronto campioni semplice	V	V
Flessibilità nell'esperimento	X	V
Bassa specificità probe	X	V
Probe lunghe	V	X
Probe corte	V	V



# Capitolo 4

## Dataset analizzati

### 4.1 Dati PET

I dataset PET su cui è stato effettuato lo studio sono tre: uno per i recettori degli oppioidi e due per i recettori della serotonina, messi a disposizione dall'Istituto di Psichiatria del King's College, Londra, UK. Tutti i dataset sono stati corretti per valori negativi o non fisiologici.

<i>Tracciante</i>	<i>Recettori</i>	<i>Azione</i>	<i>Soggetti</i>	<i>ROIs atlante</i>
$[^{11}\text{C}]\text{DPN}$	$\delta, \kappa, \mu$	AT( $\delta, \mu$ ), AG ( $\kappa$ )	10	83
$[^{11}\text{C}]\text{CUMI}$	$5\text{-HT}_{1A}$	AG	13	73
$[^{11}\text{C}]\text{WAY}$	$5\text{-HT}_{1A}$	AT	15	45

Tabella 4.1: Riassunto dataset PET (AT = antagonista, AG = agonista)

Di seguito vengono presentate nel dettaglio le informazioni disponibili per i dataset PET valutati in questa tesi, e riassunte nella Tabella 4.1.

#### 4.1.1 Dataset DPN

I dati provengono dallo studio [12] e sono stati elaborati con la tecnica descritta nella Sezione 2.2.1 del Capitolo 2. Erano disponibili le mappe parametriche di 5 soggetti sani sottoposti a due diversi esami PET (test/retest). La scansione è

stata effettuata con lo scanner PET “Siemens/CTI ECAT EXACT3D“ in seguito all’iniezione di circa 185 MBq di tracciante  $[^{11}\text{C}]\text{DPN}$ .

Il volume cerebrale di ciascun soggetto è rappresentato da una matrice  $128 \times 128 \times 95$ . Le 10 matrici 3D sono stati resi disponibili sia come mappe del volume di distribuzione ( $V_d$ ) del tracciante, sia come mappe del potenziale di legame (BP) calcolato applicando la formula (2.10) esposta nel Capitolo 2. Inoltre per ogni soggetto era disponibile un ulteriore volume maschera, risultato della segmentazione di immagini MR pesate T1 del volume cerebrale rispetto ad un atlante anatomico di riferimento [13] (si veda la Sezione 4.3). Usando l’atlante in questione l’intero volume cerebrale è stato suddiviso in 83 regioni anatomiche distinte.

#### 4.1.2 Dataset CUMI

I dati provengono da uno studio [21] sul monitoraggio della sensibilità a variazioni endogene nei recettori  $5\text{-HT}$ , usando un farmaco inibitore, del tracciante  $[^{11}\text{C}]\text{CUMI-101}$ , un radio-ligando per i recettori  $5\text{-HT}_{1A}$ . Sono stati elaborati usando il *metodo grafico di Logan* [14] descritto nella Sezione 2.2.2 del Capitolo 2. Allo studio hanno partecipato 13 soggetti sani. Dopo l’immissione di circa 375 MBq di tracciante, le scansioni PET sono state effettuate con lo scanner “GE Discovery RX PET/CT“.

Dopo aver stimato il volume di distribuzione successivamente è stato calcolato e valutato il potenziale di legame. Il dataset è caratterizzato da una matrice  $79 \times 95 \times 68$  contenente i dati del potenziale di legame di ogni soggetto, per un totale di 13 volumi. L’atlante di riferimento della maschera ed usato per segmentare le immagini MR pesate T1 è lo stesso del dataset DPN. In questo caso le regioni anatomiche valutate sono state 73.

#### 4.1.3 Dataset WAY

I dati provengono da uno studio [22] sulla relazione della densità dei recettori  $5\text{-HT}_{1A}$  e  $5\text{-HTT}$  (trasportatori della serotonina, proteine incaricate di recuperare la serotonina dallo spazio intersinaptico, riportandola nei bottoni pre-sinaptici)

in alcune regioni cerebrali. Il tracciante usato per monitorare i recettori  $5-HT_{1A}$  è  $[^{11}C]WAY-100635$ . I dati sono stati elaborati con la *Spectral Analysis*, una tecnica descritta nella Sezione 2.2.3 del Capitolo 2. Allo studio hanno partecipato 42 soggetti sani e per questa tesi sono state disponibili le mappe parametriche di 15 soggetti. Sono stati inoculati 300 MBq di tracciante e la scansione PET è stata effettuata con lo scanner “Siemens/CTI ECAT EXACT3D“.

In questo caso erano disponibili le mappe parametriche del volume di distribuzione. I 15 volumi cerebrali sono caratterizzati da matrici  $128 \times 128 \times 95$  contenenti i valori del  $V_d$ . In questo caso i dati erano già disponibili in forma di cluster, uno per ognuna delle 45 regioni di interesse in cui sono stati segmentati i volumi cerebrali.

## 4.2 Dati di espressione genica

Il dataset con i dati di espressione genica del cervello umano proviene dall'Allen Institute [23]. Questa sorgente pubblica online contiene i dati genetici e neuro-anatomici dell'intero encefalo murino, e da poco anche di quello umano [4]. Al momento dello sviluppo di questa tesi, sul database, sono presenti i dati di espressione genica di quattro donatori (in seguito chiamati donatore 1,2,3,4), ma probabilmente in futuro ne verranno aggiunti altri per avere un insieme di donatori più vario e completo, per quanto riguarda età ed etnie, di tutti i trascritti a livello cerebrale.

In Tabella 4.2 si possono vedere i dettagli di ogni donatore analizzato in questo studio. Anche se i donatori sono quattro, come si può notare dall'ultima colonna, sono stati mappati entrambi gli emisferi cerebrali per solo i donatori 1 e 2, per i donatori 3 e 4 sono disponibili i dati di espressione del solo emisfero sinistro (quindi circa metà dei campioni rispetto agli altri due donatori). Per i primi due è stato analizzato il profilo genetico di circa 900 aree anatomiche cerebrali diverse (equamente divise tra gli emisferi), rendendo disponibile un atlante estremamente dettagliato e completo, per un totale di circa 62 000 probe, cioè una copertura del 93% di tutti i geni conosciuti (per ogni gene sono state così rese disponibili almeno due probe). I quattro donatori analizzati erano maschi in un'età compresa tra i 24 ed i 57 anni, la cui anamnesi non riportava nessuna malattia neuropsichiatrica o neuro-psicologica. Sono stati inoltre effettuati uno screening tossicologico completo per diverse droghe e un'analisi postmortem per malattie come Epatite B-C e HIV per escludere patologie che potessero alterare i profili genetici.

<i>Donatore</i>	<i>Età</i>	<i>Sesso</i>	<i>Etnia</i>	<i>Emisferi disponibili</i>
1	24	M	Afroamericano	ES/ED
2	39	M	Afroamericano	ES/ED
3	57	M	Caucasico	ES
4	31	M	Caucasico	ES

Tabella 4.2: Dataset di espressione (ES = emisfero sinistro, ED = emisfero destro)

Le varie aree anatomiche sono state classificate e dissezionate con strumentazione laser da neuro-anatomopatologi esperti, nel giro di pochi giorni dal decesso.

L'isolamento dei campioni di RNA e la successiva trasformazione in *cRNA* ed analisi sono stati affidati ad un'azienda specializzata nelle analisi genetiche con *Microarray*. Per generare i dati sono stati usati array "Agilent 8x60K" progettati su misura dall'azienda e dall'Allen Institute. Questi *Microarray custom* includevano tutte le sequenze esistenti dell'array "4x44K Agilent Whole Human Genome probe set" più le sequenze di altre 16 000 probe, per un totale di almeno due probe per ogni gene monitorato. Dal database dell'Allen Institute sono stati scaricati i dati di espressione riguardanti i 4 geni di interesse di seguito riassunti:

- per i recettori per gli oppioidi abbiamo un dataset per ognuno dei 3 sottotipi:
  - il recettore  $\delta$  presenta 2 probe
  - i recettori  $\kappa, \mu$  presentano 3 probe ciascuno
- i recettori della serotonina  $5-HT_{1A}$  presentano 3 probe

I dati genetici sono stati poi normalizzati prima rispetto al cervello di appartenenza e poi rispetto agli altri cervelli, per permettere il confronto tra più donatori. Infine, il profilo di espressione di ogni donatore è stato reso pubblico sia come valori normalizzati in  $\log_2$  sia come *z-score*, raggruppandoli in tre diversi livelli di dettaglio dal punto di vista anatomico: il più semplice (*coarse*) presenta i dati di espressione in maniera grossolana, raggruppati a livello di macro-regioni cerebrali, quello intermedio (*structures*) li presenta a livello di strutture (le macro-regioni suddivise in varie parti), mentre il più basso (*samples*) presenta i dati a livello del singolo campione istologico analizzato.

Per questa tesi il dataset è stato considerato a livello del singolo campione. Come si vedrà nei prossimi capitoli, questo permette di avere un controllo completo sui dati, avendo chiara la numerosità dei campioni per ogni regione e permettendo una sorta di filtraggio in fase di pre-processing. I dati sono stati scaricati in formato *.csv*, suddivisi in 3 file: uno per i dati di espressione di ogni probe (*Expression.csv*), uno per le informazioni sui donatori e i riferimenti anatomici dei dati di espressione (*Columns.csv*) e uno con i dati relativi al gene e alle probe (*Probes.csv*).

### 4.3 Atlanti anatomici MNI usati

Per questo studio, fondamentalmente è stato usato un unico atlante anatomico di riferimento: il *Maximum Probability Hammersmith Brain Atlas* [13] (in seguito chiamato “atlante di Hammers”). Questo atlante MNI è stato creato sfruttando le informazioni delle scansioni MRI di 30 soggetti adulti e suddivide il cervello in 83 regioni anatomiche distinte. La definizione di queste regioni è stata effettuata in due momenti diversi: nel 2003 è stato creato l’atlante comprensivo delle prime 49 regioni [13], successivamente nel 2008 è stato sviluppato un atlante più ricco, con l’aggiunta delle regioni dalla 50 all’83 [24].

Una caratteristica fondamentale dell’atlante di Hammers, che ne ha semplificato il suo utilizzo, è che le regioni dell’atlante sono state numerate in modo che gli ID dispari corrispondano a regioni dell’emisfero destro, mentre gli ID pari corrispondono a regioni dell’emisfero sinistro del cervello. L’unica eccezione in questa numerazione sono il tronco encefalico (dispari), il corpo calloso (pari) e il terzo ventricolo (dispari), che sono zone centrali, e non presentano quindi una posizione predominante in uno dei due emisferi.

In questa tesi sono state usate 3 versioni dell’atlante di Hammers:

- 1) per i dati DPN l’atlante completo con 83 regioni
- 2) per i dati CUMI l’atlante nella versione con 73 regioni
- 3) per i dati WAY l’atlante nella versione con 45 regioni

Per quanto riguarda i dati di espressione, l’atlante di riferimento (in seguito chiamato “atlante di Allen”) non era disponibile al download come i dati stessi, erano invece disponibili le coordinate spaziali dei campioni rispetto all’atlante di Allen e le matrici di trasformazione per passare dalle coordinate XYZ dell’atlante di Allen a coordinate MNI standard.

# Capitolo 5

## Pre-processing e metodi

### 5.1 Sincronizzazione degli atlanti

Il primo passo è stato quello di rendere confrontabili i dati genetici e PET, mappati rispetto a due atlanti diversi. Infatti i dati di espressione erano mappati in spazi volumetrici diversi e rispetto all'atlante (non disponibile) dell'Allen Institute, mentre i dati PET erano tutti coregistrati rispetto a diverse versioni dello stesso atlante di Hammers. Si è quindi scelto di rimappare i dati di espressione dall'atlante di Allen all'atlante di Hammers, avendo così tutti i dati in uno spazio stereotassico comune.

Tutti i dati di espressione erano accompagnati dalle corrispondenti coordinate XYZ dell'atlante su cui erano mappati; inoltre l'Allen Institute metteva a disposizione anche delle API (*Application Programming Interface*) che rendevano disponibili le matrici di trasformazione affine, da applicare ad ogni donatore, per passare dalle coordinate XYZ ( $\vec{c}_{XYZ}$ ) a coordinate standard in spazio MNI ( $\vec{c}_{MNI}$ ) (le coordinate MNI, da *Montreal Neurological Institute* a Montreal in Canada, sono coordinate standardizzate per applicazioni neurologiche). Questo primo passaggio è schematizzato al punto (A) in Figura 5.1 a pag. 38 e descritto formalmente dalla seguente equazione:

$$\begin{bmatrix} A & \vec{b} \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \vec{c}_{XYZ} \\ 1 \end{bmatrix} = \begin{bmatrix} \vec{c}_{MNI} \\ 1 \end{bmatrix} \quad (5.1)$$

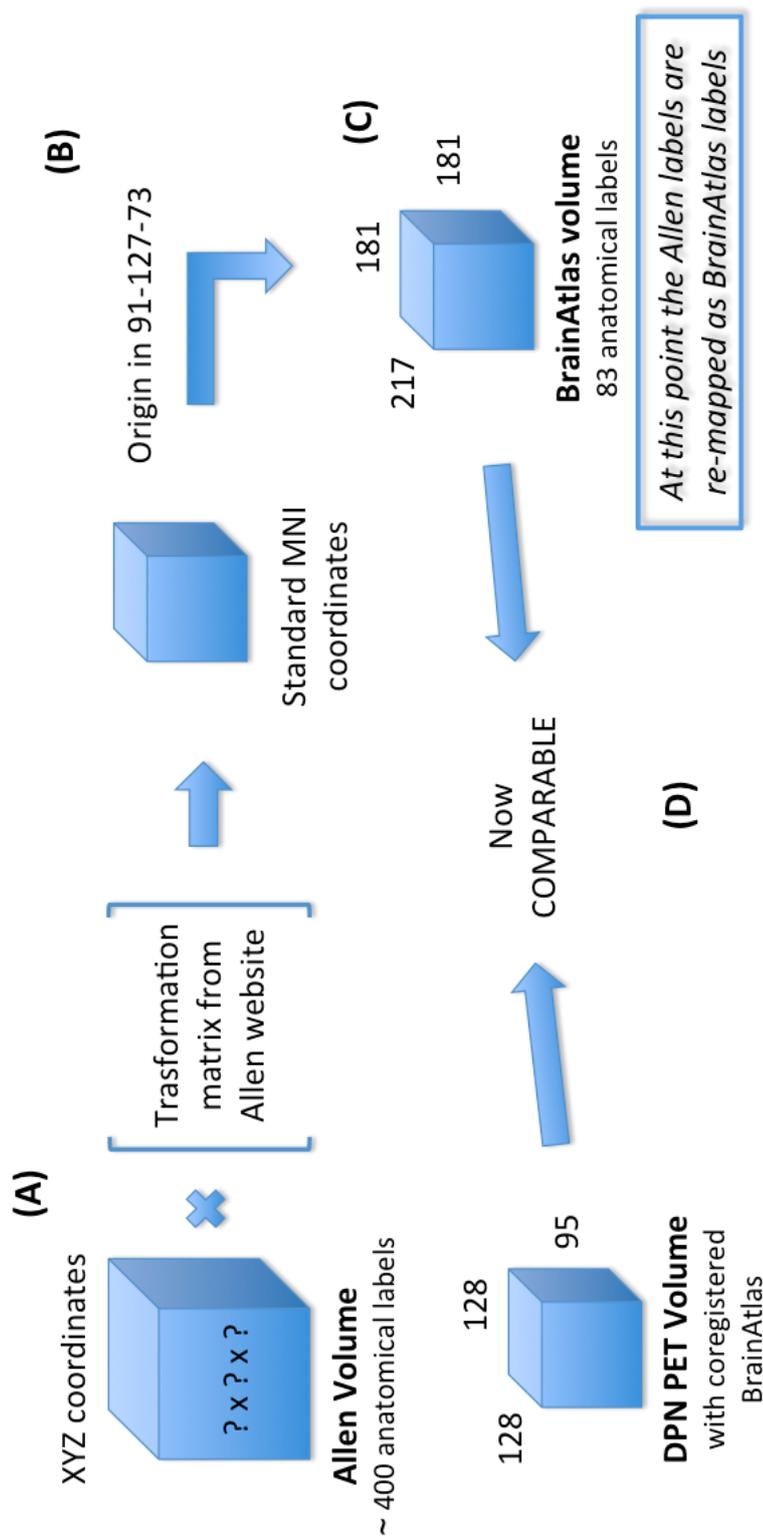


Figura 5.1: Sincronizzazione degli atlanti. In questo esempio si considera la sincronizzazione dei dati sugli oppioidi, l'atlante di Hammers è nella versione a 83 regioni.

dove  $A$  è la matrice di rotazione e  $\vec{b}$  è il vettore della traslazione. L'ultima riga della prima matrice e gli "1" nei due vettori sono dei valori fittizi richiesti dall'algebra lineare per poter rappresentare la composizione (rotazione e traslazione) della trasformazione affine come un'unica matrice. Successivamente, al punto (B), per passare dallo spazio MNI alle coordinate XYZ ( $\vec{d}$ ) dell'atlante di Hammers è stato sufficiente spostare l'origine degli assi sommando l'origine  $\vec{o} = [91, 127, 73]$ . Formalmente:

$$\vec{c}_{MNI} + \vec{o} = \vec{d} \quad (5.2)$$

Dopo l'applicazione dell'equazione (5.1) e (5.2), le coordinate sono riferite all'atlante voluto ed si può quindi passare alla fase di "pairing" (accoppiamento), visibile al punto (C) della Figura 5.1. In questa fase tutti i campioni di espressione sono stati rimappati, riassegnando l'etichetta anatomica data da Allen con l'etichetta dell'area di Hammers in cui il campione è stato posizionato. In pratica, prendendo sempre come esempio i dataset per gli oppioidi, se più campioni di Allen con etichette diverse cadono in una delle 83 regioni dell'atlante di Hammers, questi vengono riclassificati come tutti appartenenti a quella regione. Un esempio è visibile di seguito, in Figura 5.2.

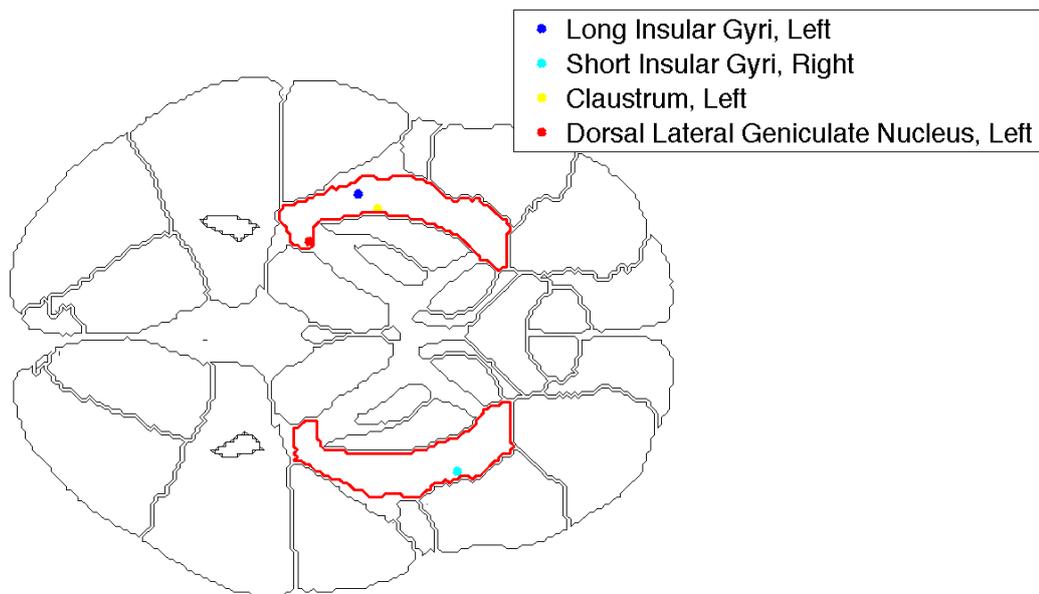


Figura 5.2: Campioni di espressione (legenda) del recettore  $\kappa$  rimappati come Insula (ROIs rosse) rispetto all'atlante di Hammers a 83 regioni (sog.1, slice 74)

Effettuando il *pairing* su tutti i campioni si passa dalle circa 400 etichette di Allen alle 83 (o 73 o 45) etichette di Hammers, si effettua quindi una sorta di sotto-campionamento. Questo passaggio non impoverisce in nessun caso le analisi successive dato che comunque in seguito si andranno a valutare le medie e le mediane dei dati di espressione individuati in ognuna delle regioni. Si ricorda infatti che è di interesse la relazione tra espressione genica e la densità recettoriale in una regione anatomica nella sua interezza.

La fase di *pairing* presenta comunque un inconveniente, infatti può succedere che alcuni campioni di espressione non vengano mappati in nessuna area dell'atlante di Hammers, portando quindi ad un accoppiamento nullo, identificato nelle liste degli accoppiamenti come “*NO MATCH*“. In Tabella 5.1 si possono vedere le percentuali di campioni che non trovano un match corretto. Queste percentuali di mismatch sono dovute essenzialmente alla diversa segmentazione delle aree fatta nell'atlante di Allen rispetto alle varie versioni dell'atlante di Hammers. Un esempio di mismatch è visibile in Figura 5.3, si può infatti notare quello che succede nella slice 74 del donatore 1 per un campione di espressione del recettore  $\kappa$  nel caso dell'atlante a 83 regioni: il campione cade in un zona vuota in cui non è identificata nessuna delle 83 regioni cerebrali possibili.

<i>Donatore</i>	<i>NoMatch DPN</i>	<i>NoMatch CUMI</i>	<i>NoMatch WAY</i>
1	11.6 %	11.0 %	4.3 %
2	13.2 %	12.2 %	2.9 %
3	19.3 %	17.9 %	6.3 %
4	21.2 %	17.6 %	6.4 %
Media $\pm$ SD	(16.3 $\pm$ 4.6)%	(14.7 $\pm$ 3.6)%	(5.0 $\pm$ 1.7)%

Tabella 5.1: Percentuali di “*NO MATCH*“ per ogni donatore rispetto ai 3 atlanti

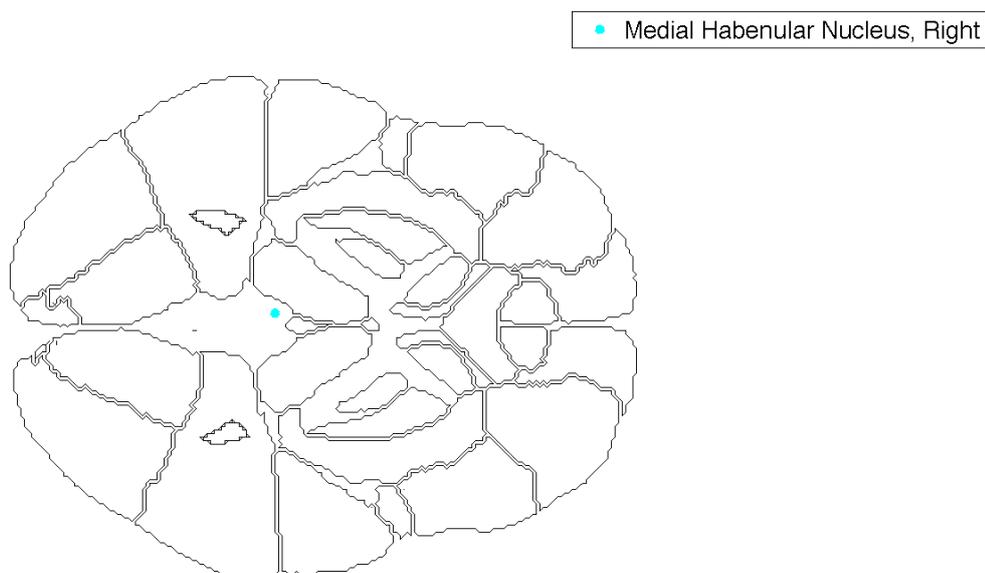


Figura 5.3: Esempio di campioni “NO MATCH“ nella stessa fetta della Figura 5.2

Un volta giunti al punto (D) della Figura 5.1 a pag. 38 i dati di espressione sono confrontabili con quelli PET, infatti entrambi i dataset presentano i volumi cerebrali allineati rispetto allo stesso atlante di riferimento, e suddivisi nello stesso numero di regioni anatomiche, con tutti i campioni di espressione rimappati con gli ID assegnati nell’atlante di Hammers.

## 5.2 Correlazione e precisione delle stime

Dopo una prima analisi visiva a campione sui dati di espressione si vede che è presente una bassa numerosità ed un'alta variabilità nel numero di campioni genetici per ogni ROI. Queste due cose rendono difficile supportare l'ipotesi di gaussianità della distribuzione di valori all'interno delle ROI.

Per ridurre l'incidenza della variabilità nel numero di campioni ed avere una base statistica minima, è stata applicata una soglia sul numero di campioni minimo in ogni regione. La soglia scelta è stata di 5 campioni per ROI, cosicché tutte le regioni con meno di 5 campioni non fossero considerate nelle successive analisi. Osservando che non tutti i donatori presentano la stessa numerosità per regioni corrispondenti, e ricordando anche che per i donatori 3 e 4 non sono presenti dati sull'emisfero destro, si è allora deciso di analizzare solo i valori di espressione delle ROI comuni a tutti e quattro i donatori. Anche le analisi di correlazione dei dataset PET sono state effettuate valutando solo le ROI usate per condurre l'analisi dei dati di espressione.

Possedendo i dati genetici un alta variabilità intrinseca, si è scelto di adottare la mediana, invece della media, per rappresentare il livello di espressione genica media nelle regioni considerate, e di usare il coefficiente di correlazione per ranghi di Spearman ( $\rho$ ) invece del coefficiente di correlazione lineare di Pearson.

La correlazione di Spearman è un test non parametrico, in cui la distribuzione dei valori non deve essere gaussiana. Può essere usato anche per campioni di dati con piccola numerosità (tra 7 e 20) in cui l'unica ipotesi è che questi siano in scala ordinale (cioè i valori possono essere ordinati e si può stabilire se un valore è maggiore, minore od uguale ad un altro). Date due variabili  $X_i$  e  $Y_i$ , il coefficiente si calcola con la formula:

$$\rho = 1 - \frac{6 \sum_i D_i^2}{n(n^2 - 1)} \quad (5.3)$$

con  $D_i = r_i - s_i$  la differenza dei ranghi ( $r_i$  e  $s_i$  ranghi dell' $i$ -esima osservazione delle due variabili), e  $n$  il numero di osservazioni.  $\rho$  assumerà un valore tra -1 e +1 in base al grado di relazione tra le due variabili. In tutte le analisi di correlazione della tesi è stato valutato il quadrato del coefficiente di correlazione, cioè  $\rho^2$ .

Non avendo fatto assunzioni sulla distribuzione di probabilità dei valori di espressione nelle ROI, si è adottato il metodo bootstrap per valutare la precisione delle stime di correlazione. Il bootstrap è una tecnica di ricampionamento che permette di approssimare la distribuzione di una statistica, permettendo di calcolare la media e la varianza approssimate di uno stimatore. Essenzialmente l'algoritmo ricampiona le due variabili analizzate un numero  $n$  di volte ed applica su queste nuove variabili lo stimatore analizzato, producendo  $n$  stime. Nel nostro caso lo stimatore in questione è la correlazione di Spearman e ricampionando 100 volte i dati si ricavano con il bootstrap 100 valori di correlazione. Calcolata la deviazione standard dei 100 valori di correlazione, la precisione delle stime si è valutata poi come  $SD \cdot 100$ .



# Capitolo 6

## Risultati e discussione

### 6.1 Analisi dei dati genetici: variabilità e consistenza

Dopo la fase di pre-processing in cui si sono stati sincronizzati gli atlanti, come descritto nella Sezione 5.1 del Capitolo 5, portando tutti i dati in uno spazio comune, si sono estratti i dati di espressione suddivisi in base alle regioni dell'atlante di riferimento e si è proceduto con una prima analisi individuale dei dati di espressione, per valutare il range di variabilità e la consistenza dei dataset.

Come si può vedere nella seconda colonna delle Tabelle 6.1, 6.2 e 6.3 per ogni donatore, in media è presente una bassa numerosità ed un'alta variabilità nel numero di campioni genetici per ogni ROI. Queste caratteristiche, comuni a tutti i dataset, sono dovute al campionamento non uniforme effettuato da Allen. In base alle zone in cui si concentrano i campioni di espressione si è visto che è stata data priorità ad alcune regioni anatomiche specifiche, concentrando il campionamento nella zona centrale del volume cerebrale.

Nell'ultima colonna delle Tabelle 6.1, 6.2 e 6.3 si può vedere il numero di ROI eliminate applicando la soglia scelta (minimo 5 campioni per regione) come descritto nella Sezione 5.2 del Capitolo 5, in media sono state rimosse circa 40 regioni per donatore in tutti i casi analizzati.

Come si può notare, le tabelle sono tre anche se i dati di espressione provengono da due dataset (oppioidi e serotoninergici). Questo è dovuto al fatto che per il dataset sui recettori della serotonina erano presenti due traccianti PET diversi ( $[^{11}\text{C}]\text{CUMI}$  e  $[^{11}\text{C}]\text{WAY}$ ) coregistrati su due versioni diverse dell'atlante di Hammers: l'atlante a 73 regioni (Tabella 6.2) e l'atlante a 45 regioni (6.3). Data la diversa segmentazione dei due atlanti, le due tabelle mostrano valori abbastanza diversi. L'atlante a 45 regioni presenta i valori migliori dal punto di vista delle numerosità e di conseguenza dal punto di vista delle ROI rimosse. Il fattore numerosità è dovuto al fatto che avere meno regioni a parità di volume corrisponde ad una segmentazione più ampia e quindi più campioni per ogni ROI; di conseguenza anche il numero di regioni rimosse applicando la soglia è minore in proporzione.

Le successive analisi di correlazioni sono state effettuate su un numero di ROI variabile tra 15 e 18 in base al dataset, avendo così valori comuni e confrontabili tra tutti i donatori.

<i>Donatore</i>	<i># camp. per ROI media±SD</i>	<i># ROI vuote</i>	<i># ROI rimosse</i>
1	12±19	12/83	35/83
2	10±21	9/83	38/83
3	7±8	42/83	62/83
4	10±13	41/83	58/83

Tabella 6.1: Informazioni sui campioni per ogni donatore nel dataset degli oppioidi  
*# camp. per ROI* = campioni per ROI, *# ROI vuote* = ROI per le quali non ci sono campioni genetici, *# ROI rimosse* = ROI eliminate dopo l'applicazione della soglia

<i>Donatore</i>	<i># camp per ROI media±SD</i>	<i># ROI vuote</i>	<i># ROI rimosse</i>
1	13±18	6/73	28/73
2	12±21	7/73	29/73
3	8±9	37/73	53/73
4	12±14	38/73	46/73

Tabella 6.2: Informazioni sui campioni per ogni donatore nel dataset del  $5-HT_{1A}$  (atlante a 73 regioni)

<i>Donatore</i>	<i># camp per ROI media±SD</i>	<i># ROI vuote</i>	<i># ROI rimosse</i>
1	22±30	4/45	13/45
2	21±33	4/45	15/45
3	14±17	21/45	29/45
4	21±22	21/45	25/45

Tabella 6.3: Informazioni sui campioni per ogni donatore nel dataset del  $5-HT_{1A}$  (atlante a 45 regioni)

Dopo la valutazione dei dati di espressione per ogni donatore si può passare all'analisi di correlazione tra i donatori dei dataset di espressione. Nelle Tabelle da 6.4 a 6.8 l'ultima colonna rappresenta la correlazione media e la standard deviation percentuale tra i donatori per ogni probe e, come si può vedere dai valori tra parentesi, le correlazioni presentano un'alta variabilità inter-soggetto in molti casi. Queste tabelle possono essere valutate da due punti di vista:

- dal punto di vista dei donatori, cioè valutando la variabilità tra i donatori per una stessa probe
- dal punto di vista delle probe, cioè valutando la correlazione media tra le varie probe

Ricordando innanzi tutto che si stanno valutando dati di espressione genica, che come già detto presentano una variabilità intrinseca, rispetto ai donatori la variabilità delle correlazione per una stessa probe è molto alta (sopra il 50%) in molti casi. Questo chiaramente è dovuto al fatto che una data probe non ha lo stesso comportamento per tutti i donatori. Di contro, le probe con bassa variabilità presentano anche un'alta correlazione media. Ricordando quanto descritto nella Sezione 4.2 del Capitolo 4, l'Allen Institute ha disegnato almeno due probe diverse per ogni gene monitorato, quindi il fatto che per un dato recettore una delle probe presenti valori molto variabili (con una deviazione standard percentuale sopra il 50-70%) tra i donatori può essere segno che quella probe non rappresenta una sequenza specifica uguale in tutti i donatori. Al contrario, se una delle probe si lega in maniera specifica al trascritto monitorato in tutti i donatori, e quindi presenta una correlazione media alta con una bassa variabilità (sotto il 40%) questo vuol dire che la probe è disegnata bene, con una sequenza molto espressa e conservata rispetto a donatori diversi. Si osserva infatti che, tranne per il recettore  $\kappa$ , in tutti gli altri casi si può sempre individuare almeno una probe che in media correla nettamente meglio delle altre.

Infine, si può vedere che le correlazioni in Tabella 6.7 e 6.8 sono mediamente più alte rispetto alle tabelle degli oppioidi. Questo può essere dovuto alla specificità delle probe non costante, oppure alla diversa segmentazione operata dagli

atlanti. Nonostante questo, anche in questi casi comunque si può individuare una probe che correla meglio delle altre.

In Figura 6.1 si può vedere la rappresentazione a colori (in una scala da 0 ad 1) dei valori del coefficiente di correlazione ( $\rho^2$ ) dei dati di espressione per il recettore  $\kappa$  (probe 2) tra tutti i donatori del dataset (ogni riga e ogni colonna corrisponde ad un donatore). I valori percentuali presenti in ogni riquadro rappresentano la deviazione standard, moltiplicata per 100, risultante dall'applicazione del metodo Bootstrap, effettuando 100 ricampionamenti dei dati usati per calcolare il singolo valore di correlazione. Successivamente in questo capitolo, tutte le mappe di correlazione presenteranno le stesse caratteristiche appena descritte.

	<i>1 vs 2</i>	<i>1 vs 3</i>	<i>1 vs 4</i>	<i>2 vs 3</i>	<i>2 vs 4</i>	<i>3 vs 4</i>	$\rho^2$ media (SD%)
<i>Probe 1</i>	0.00	0.00	0.01	0.27	0.02	0.01	0.05 (203)
<i>Probe 2</i>	0.14	0.33	0.56	0.03	0.14	0.08	0.21 (93)

Tabella 6.4: Valori di correlazione ( $\rho^2$ ) dei dati di espressione dei recettori  $\delta$  per tutti i donatori e tutte le probe.

	<i>1 vs 2</i>	<i>1 vs 3</i>	<i>1 vs 4</i>	<i>2 vs 3</i>	<i>2 vs 4</i>	<i>3 vs 4</i>	$\rho^2$ media (SD%)
<i>Probe 1</i>	0.65	0.25	0.14	0.03	0.08	0.35	0.25 (90)
<i>Probe 2</i>	0.04	0.51	0.34	0.09	0.21	0.53	0.29 (73)
<i>Probe 3</i>	0.09	0.46	0.07	0.45	0.03	0.08	0.20 (102)

Tabella 6.5: Valori di correlazione ( $\rho^2$ ) dei dati di espressione dei recettori  $\kappa$  per tutti i donatori e tutte le probe.

	<i>1 vs 2</i>	<i>1 vs 3</i>	<i>1 vs 4</i>	<i>2 vs 3</i>	<i>2 vs 4</i>	<i>3 vs 4</i>	$\rho^2$ media (SD%)
<i>Probe 1</i>	0.12	0.42	0.25	0.60	0.57	0.78	0.46 (54)
<i>Probe 2</i>	0.61	0.71	0.75	0.77	0.66	0.84	0.72 (11)
<i>Probe 3</i>	0.36	0.82	0.31	0.51	0.05	0.16	0.37 (74)

Tabella 6.6: Valori di correlazione ( $\rho^2$ ) dei dati di espressione dei recettori  $\mu$  per tutti i donatori e tutte le probe.

	<i>1 vs 2</i>	<i>1 vs 3</i>	<i>1 vs 4</i>	<i>2 vs 3</i>	<i>2 vs 4</i>	<i>3 vs 4</i>	$\rho^2$ media (SD%)
<i>Probe 1</i>	0.77	0.68	0.69	0.50	0.53	0.88	0.67 (21)
<i>Probe 2</i>	0.19	0.34	0.57	0.01	0.25	0.47	0.31 (65)
<i>Probe 3</i>	0.74	0.68	0.91	0.54	0.66	0.77	0.72 (17)

Tabella 6.7: Valori di correlazione ( $\rho^2$ ) dei dati di espressione dei recettori  $5-HT_{1A}$  per tutti i donatori e tutte le probe (atlante a 73 regioni).

	<i>1 vs 2</i>	<i>1 vs 3</i>	<i>1 vs 4</i>	<i>2 vs 3</i>	<i>2 vs 4</i>	<i>3 vs 4</i>	$\rho^2$ media (SD%)
Probe 1	0.81	0.63	0.37	0.52	0.35	0.43	0.52 (34)
Probe 2	0.02	0.16	0.57	0.00	0.31	0.26	0.22 (97)
Probe 3	0.79	0.62	0.41	0.54	0.46	0.39	0.54 (28)

Tabella 6.8: Valori di correlazione ( $\rho^2$ ) dei dati di espressione dei recettori  $5-HT_{1A}$  per tutti i donatori e tutte le probe (atlante a 45 regioni).

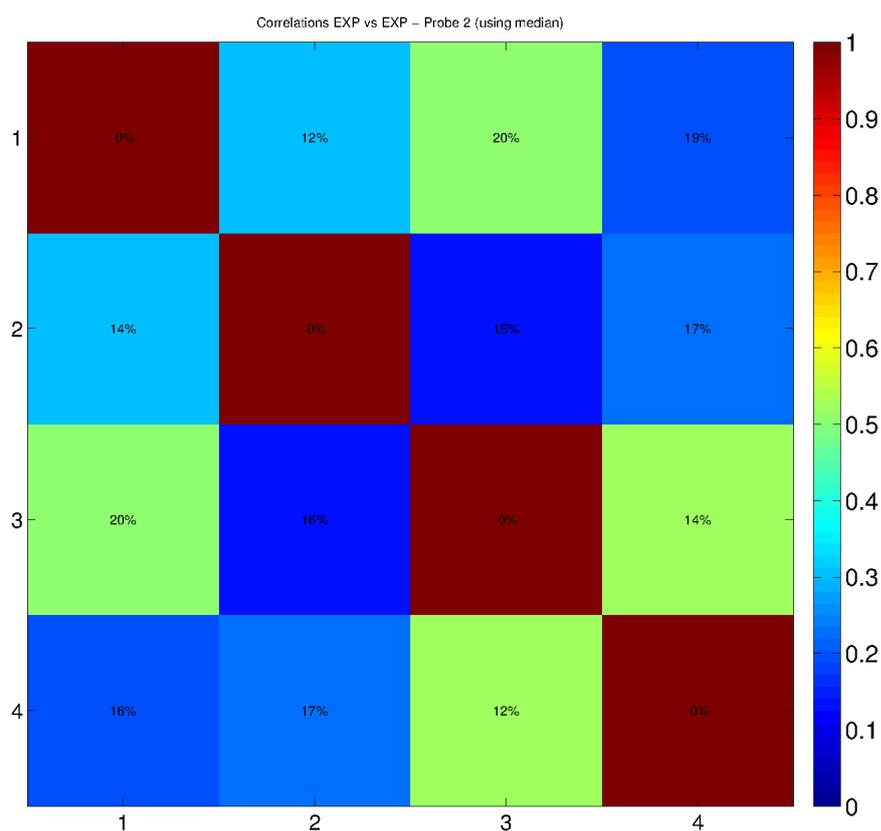


Figura 6.1: Rappresentazione a colori della correlazione ( $\rho^2$ ) dei dati di espressione dei recettore  $\kappa$  per la probe 2. Ogni riga e ogni colonna corrisponde ad un donatore.

## 6.2 Analisi dei dati PET

Le Figure 6.2, 6.3 e 6.4 sono una rappresentazione a colori dei valori di correlazione (in  $\rho^2$ ) di tutti i soggetti del dataset, in una scala colorata che va da 0 (nessuna correlazione) a 1 (correlazione massima).

Ogni riga e ogni colonna rappresentano un soggetto diverso, e come si può vedere le correlazioni sono molto alte per tutti i soggetti PET in tutti e tre i dataset. Questo risultato era atteso, perché i traccianti usati per creare le mappe parametriche sono traccianti specifici e affini per i recettori in esame, studiati per avere un'alta ripetibilità.

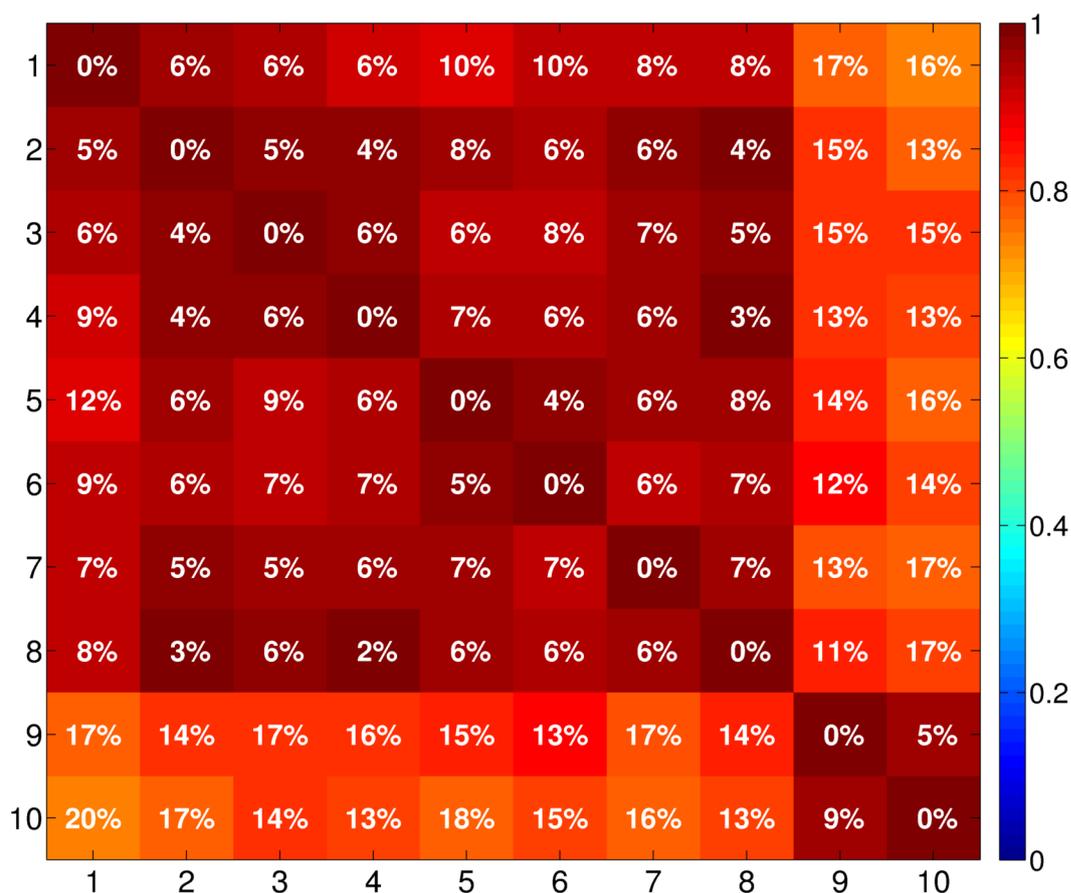
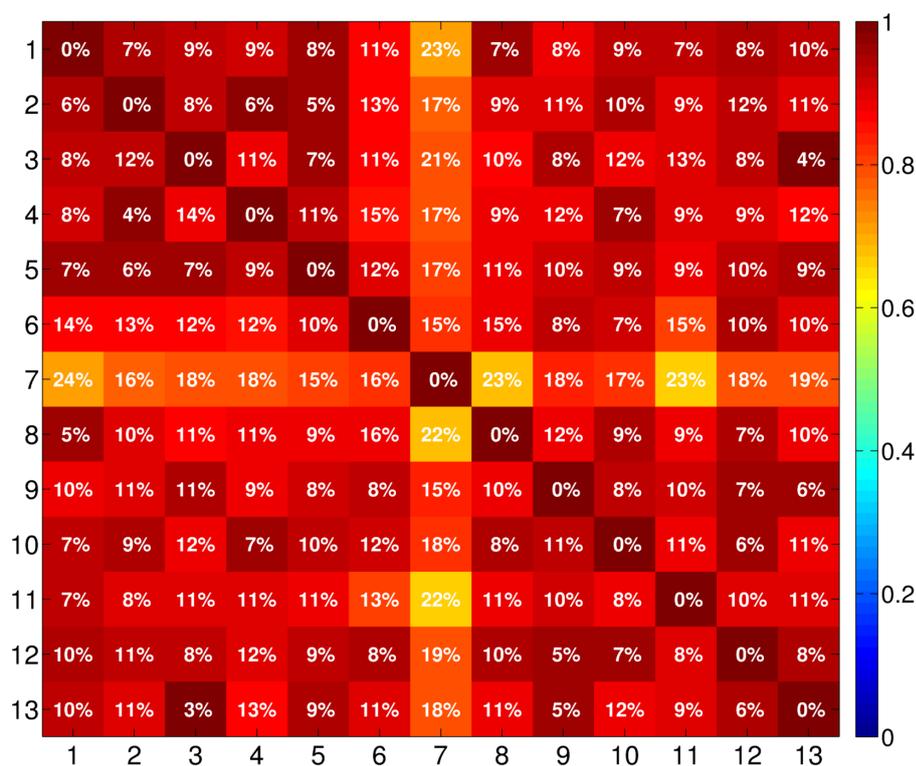
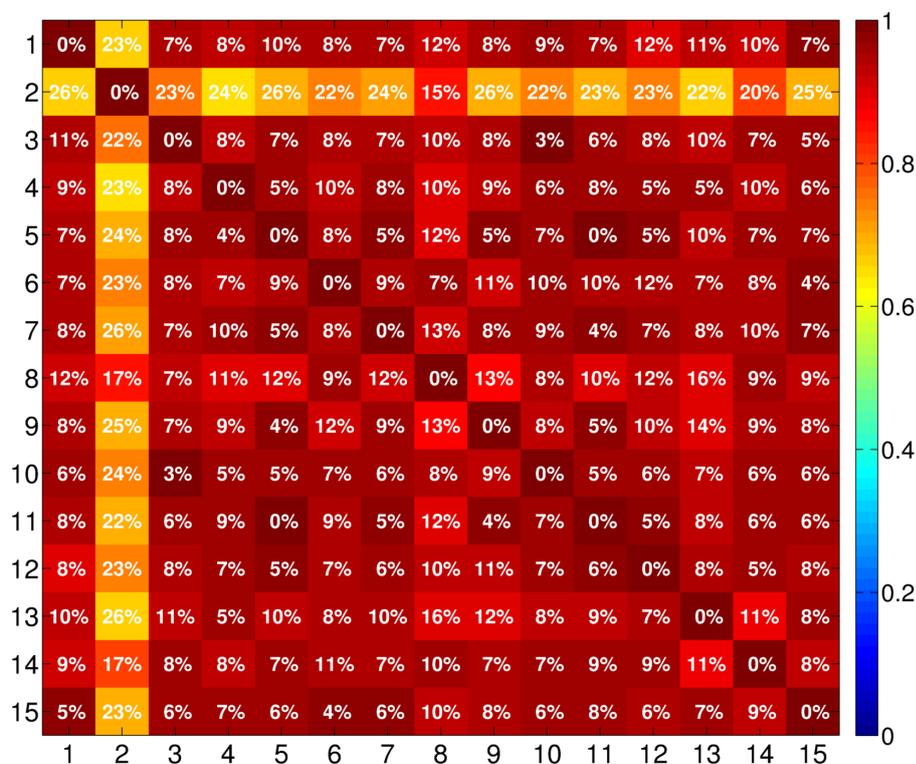


Figura 6.2: Rappresentazione a colori dei  $\rho^2$  per il tracciante  $[^{11}\text{C}]\text{DPN}$ .

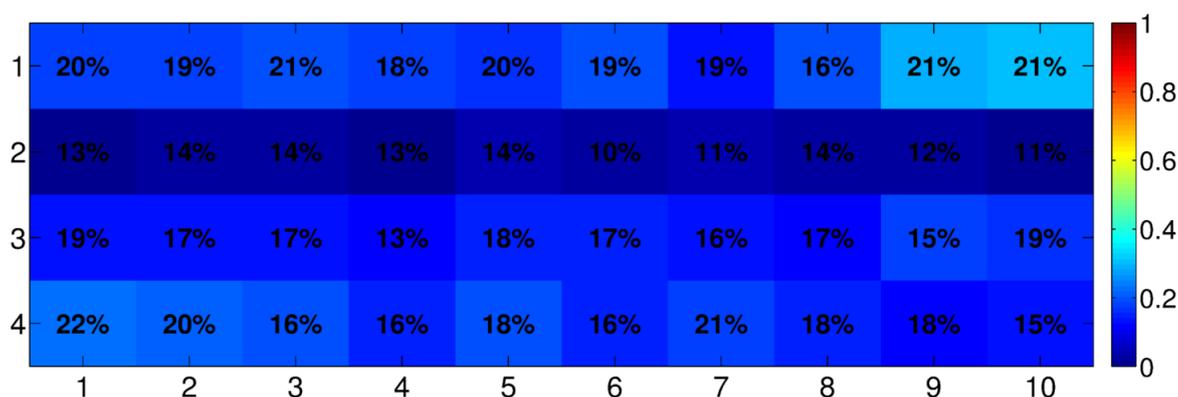
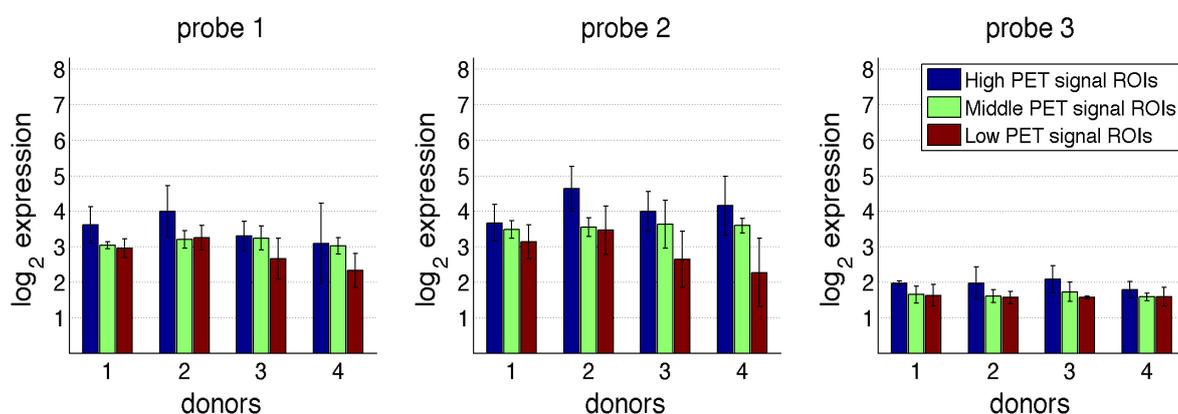
Figura 6.3: Rappresentazione a colori dei  $\rho^2$  per il tracciante  $[^{11}\text{C}]\text{CUMI}$ .Figura 6.4: Rappresentazione a colori dei  $\rho^2$  per il tracciante  $[^{11}\text{C}]\text{WAY}$ .

### 6.3 Correlazione per i recettori oppioidi

Dopo le analisi sui singoli dataset sono state eseguite le analisi di correlazione tra i donatori del dataset di espressione genica dei tre recettori per gli oppioidi ed il dataset PET del tracciante  $[^{11}\text{C}]\text{DPN}$ . Anche se erano disponibili sia le mappe del  $V_d$  che del BP per questo tracciante, si è deciso di usare le mappe del potenziale di legame per riportare i risultati. Nella Figure 6.5 e 6.6 sono presentati due esempi delle rappresentazioni a colori dei  $\rho^2$  (di seguito chiamate mappe di correlazione), sviluppate per poter visualizzare in maniera rapida e poter confrontare con semplicità i risultati di tutti i soggetti per tutte le probe. Ai fini della spiegazione e del commento dei risultati viene presentato il confronto dei dati di espressione per il recettore  $\kappa$  e i dati PET del tracciante  $[^{11}\text{C}]\text{DPN}$ . Le altre matrici di correlazione di tutti i recettori per gli oppioidi e per tutte le probe sono disponibili nell'Appendice A.

Per calcolare il coefficiente di correlazione di Spearman sono state confrontate le mediane delle 15-18 ROI estratte dopo l'analisi dei dati di espressione sia per la sorgente di espressione che per quella PET. In entrambi i casi, i dati di queste ROI sono stati normalizzati rispetto al loro massimo prima calcolare le singole correlazioni.

La Figura 6.5 rappresenta la mappa di correlazione tra i 10 soggetti PET per il tracciante  $[^{11}\text{C}]\text{DPN}$  (lungo le colonne) e i 4 donatori di espressione per il recettore  $\kappa$  (lungo le righe) per quanto riguarda la probe 2 di questo recettore. I valori di  $\rho^2$  variano tra 0.12 e 0.51, con una media di 0.36. In questo caso, tranne per il primo donatore di espressione (prima riga) per tutti gli altri donatori i valori del coefficiente di correlazione sono visibilmente più alti. In Figura 6.6 c'è lo stesso confronto di correlazione, ma per quanto riguarda i dati di espressione della probe 3. In questo caso i valori di correlazione variano tra 0.01 e 0.31, con una media di 0.13. E' evidente che nel primo caso c'è un discreto livello di correlazione, mentre nel secondo caso la correlazione è praticamente assente. Questo fatto è stato riscontrato anche nelle correlazioni del recettore  $\mu$ , in cui la probe 2 presenta un comportamento migliore rispetto alle altre. Nel caso di  $\delta$  invece si è riscontrata una bassa correlazione in entrambe le probe disponibili.

Figura 6.5: *KOP* (probe 2) vs. [ $^{11}\text{C}$ ]DPN (BP).Figura 6.6: *KOP* (probe 3) vs. [ $^{11}\text{C}$ ]DPN (BP).Figura 6.7: Analisi di espressione (recettore  $\kappa$ ) delle zone ad alto, medio e basso segnale PET ([ $^{11}\text{C}$ ]DPN BP)

Dopo aver eseguito le correlazioni suddette si sono andati ad indagare i livelli di espressione per il recettore in esame nelle zone in cui si sa esserci un alto numero di recettori affini al tracciante [ $^{11}\text{C}$ ]DPN. Per estrarre queste zone si valuta la lista di ROI di un super-soggetto, creata facendo la media per ogni mediana di tutte le ROI (la lista considera solo le regioni estratte nell'analisi di espressione) di ogni soggetto PET; successivamente si sono ordinati i valori delle ROI in ordine crescente e, usando come soglia il 20% del numero di regioni, si è preso il primo 20% delle regioni come ROI a basso segnale, l'ultimo 20% come ROI ad alto segnale, e per avere una sorta di valutazione intermedia si sono selezionate il 20% delle ROI rispetto al centro della lista come regioni a medio segnale PET. In Figura 6.7 si possono vedere i grafici a barre risultanti da questa analisi, eseguita individualmente su ogni donatore del dataset di espressione e per ogni probe. Si nota subito che il livello di espressione per le zone ad alto segnale PET (barre blu) è marcatamente più alto rispetto al livello per le zone a basso segnale PET (barre rosse) sia nella probe 2 che nella probe 1, invece per la probe 3 il livello di espressione è mediamente più basso rispetto alle altre due sonde, anche la differenza tra le zone ad alto e basso segnale è quasi assente. Le zone a medio segnale (barre verdi) si comportano in maniera variabile, mantenendo comunque un livello intermedio tra le altre due, ma senza delineare una scala precisa. Considerando che si stanno valutando valori in scala logaritmica il risultato di questi grafici per la probe 2 e la probe 3 rispecchiano ed enfatizzano quanto visto nel caso delle correlazioni in Figura 6.5 e 6.6, sembra infatti evidente che la probe 3 non è ottimale per il monitoraggio del gene a cui si riferisce.

## 6.4 Correlazione per i recettori serotoninergici

In questa sezione vengono analizzate le correlazioni fatte tra i dati di espressione dei recettore della serotonina  $5-HT_{1A}$  e i due traccianti PET  $[^{11}C]CUMI$  e  $[^{11}C]WAY$ . Si ricorda che per il tracciante  $[^{11}C]CUMI$  i dati delle mappe parametriche si riferiscono al potenziale di legame, mentre per il tracciante  $[^{11}C]WAY$  le mappe sono del volume di distribuzione. Anche in questo caso vengono riportate solo le immagini e i grafici necessari per esemplificare i risultati ottenuti, in Appendice A si trovano tutti i risultati conseguiti.

Anche in questo caso, per calcolare il coefficiente di correlazione di Spearman sono state confrontate le mediane delle 15-18 ROI estratte dopo l'analisi dei dati di espressione per entrambe le sorgenti di dati. I dati di queste ROI sono stati normalizzati rispetto al loro massimo prima calcolare le varie correlazioni.

In Figura 6.8 e 6.9 si possono vedere le mappe di correlazione per il dataset del tracciante  $[^{11}C]CUMI$  rispetto ai valori di espressione del recettore  $5HT_{1A}$ . La prima rappresenta la mappa rispetto alla probe 1, il coefficiente di correlazione di Spearman varia tra 0.21 e 0.77 con una media di 0.46; la seconda rappresenta la mappa di correlazione della probe 2, con valori tra 0.05 e 0.50 e media 0.26. Come nel caso dei recettori per gli oppioidi è chiara la differenza di correlazione tra le probe per uno stesso dataset, è quindi ancora verificato il comportamento diverso rispetto alle probe.

I grafici a barre della Figura 6.10 sono stati prodotti nello stesso modo descritto nella sezione precedente, usando sempre come soglia il 20% delle ROI totali analizzate. Si può vedere molto bene che la probe 2 presenta un livello di espressione molto basso rispetto alle altre due, mentre per la probe 1 e 3 le regioni ad alto segnale PET sono ben distinte da quelle a basso segnale. Anche le zone a medio segnale sono abbastanza distinte dalle altre due, con valori intermedi praticamente in tutti i casi. Questi grafici possono quindi essere usati per verificare l'attendibilità delle mappe di correlazione, ed identificare agevolmente la probe o le probe che non funzionano bene per il gene analizzato.

Di seguito, nelle Figure 6.11 e 6.12 a pag. 59 vengono riportate le mappe di correlazione per la probe 1 e 2 usando i dati del volume di distribuzione del

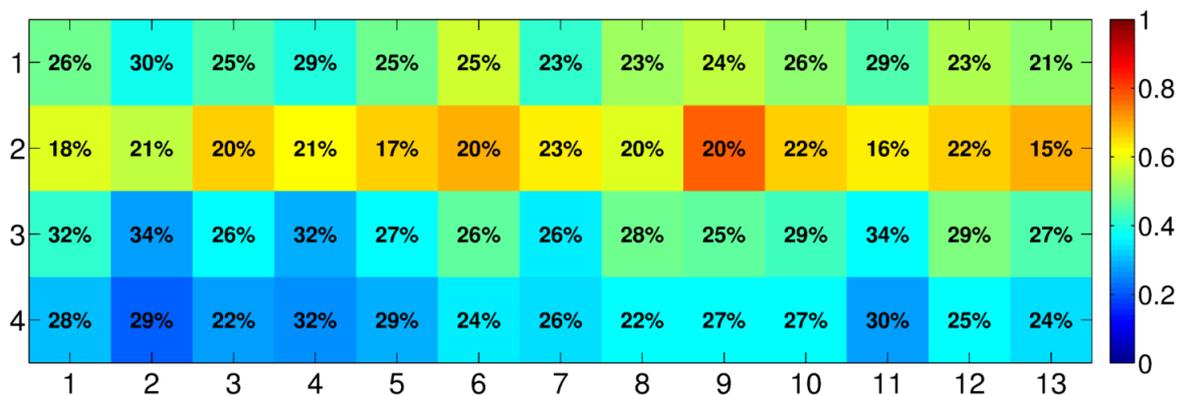


Figura 6.8:  $5HT_{1A}$  (probe 1) vs.  $[^{11}C]CUMI$  (BP).

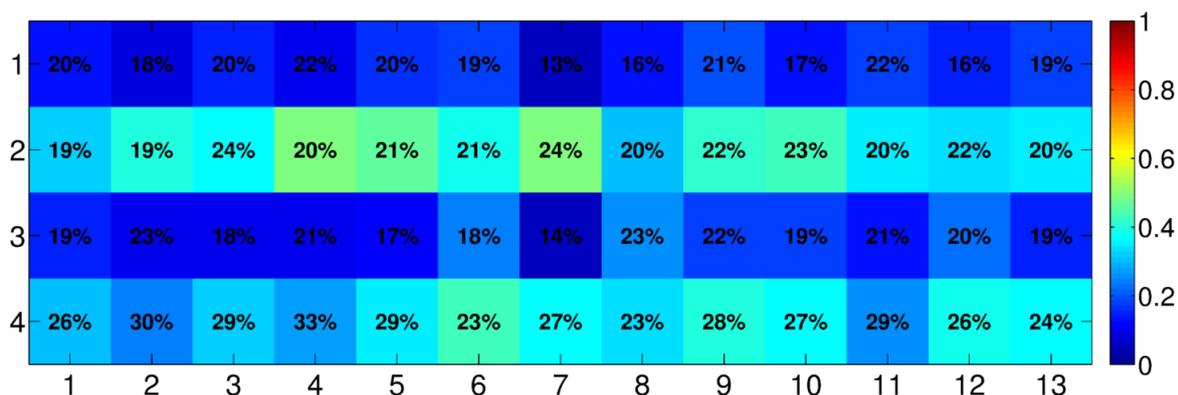


Figura 6.9:  $5HT_{1A}$  (probe 2) vs.  $[^{11}C]CUMI$  (BP).

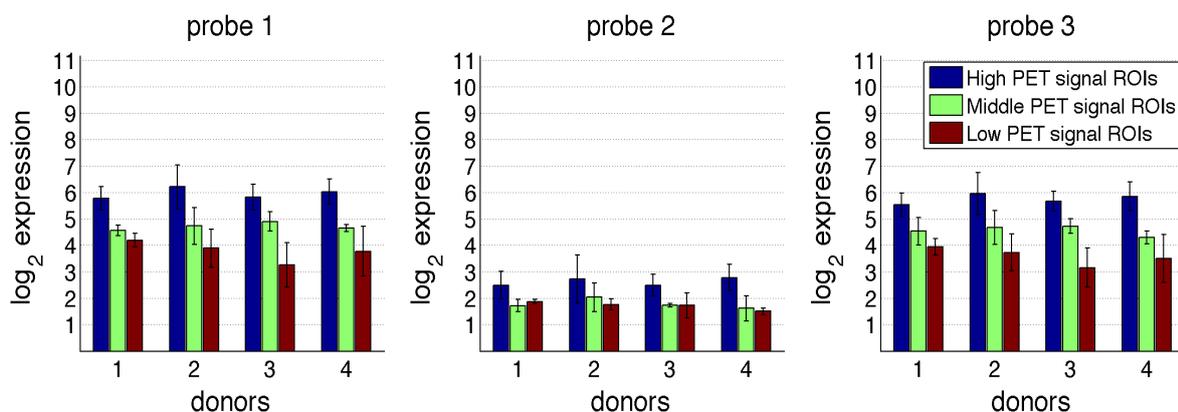
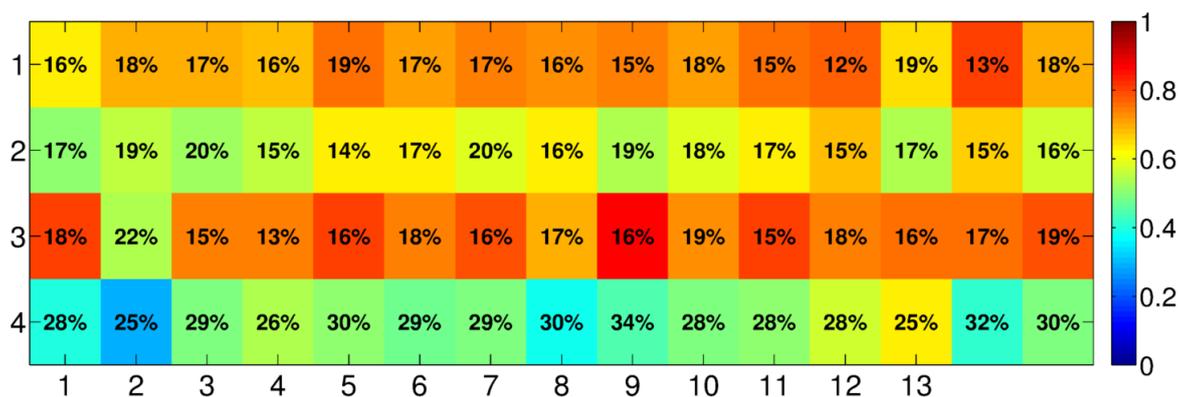
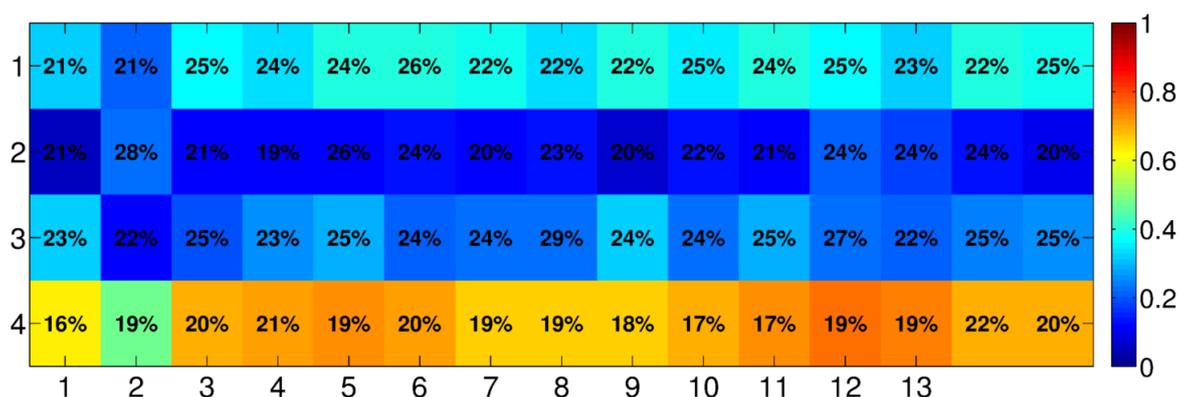
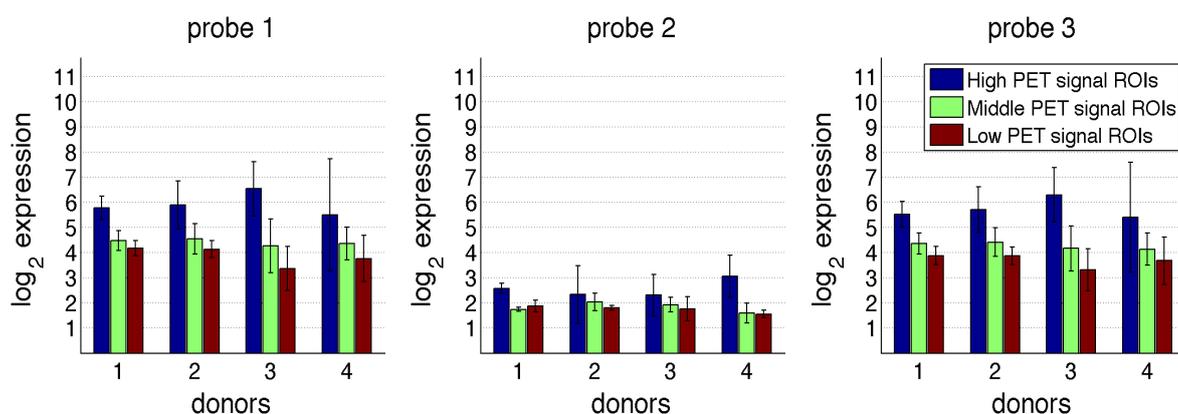


Figura 6.10: Analisi di espressione (recettore  $5HT_{1A}$ ) delle zone ad alto, medio e basso segnale PET ( $[^{11}C]CUMI$  BP)

Figura 6.11:  $5HT_{1A}$  (probe 1) vs.  $[^{11}C]WAY$  ( $V_d$ ).Figura 6.12:  $5HT_{1A}$  (probe 2) vs.  $[^{11}C]WAY$  ( $V_d$ ).Figura 6.13: Analisi di espressione (recettore  $5HT_{1A}$ ) delle zone ad alto, medio e basso segnale PET ( $[^{11}C]WAY$   $V_d$ )

tracciante [ $^{11}\text{C}$ ]WAY come sorgente PET e sempre i dati del dataset di espressione dei recettori  $5\text{HT}_{1A}$ . A differenza del tracciante [ $^{11}\text{C}$ ]CUMI, che è un parziale agonista altamente selettivo per i recettori  $5\text{HT}_{1A}$ , il tracciante [ $^{11}\text{C}$ ]WAY è un antagonista sempre altamente selettivo per gli stessi recettori. I bersagli di questi due radio-ligandi sono la stessa famiglia di recettori, e quindi anche per i traccianti [ $^{11}\text{C}$ ]WAY ci si aspettano dei risultati simili a quelli descritti per il tracciante [ $^{11}\text{C}$ ]CUMI.

Provenendo i dati da due studi diversi, i soggetti analizzati sono diversi, inoltre diverso è anche il parametro fisiologico considerato ( $V_d$  invece di BP). Questi fattori possono essere usati per considerare i risultati per quest'ultimo tracciante come controprova per validare i metodi usati per le analisi di correlazione fin qui condotte. Come si può vedere dalle mappe di correlazione, la probe 1 presenta alti valori di correlazione (tra 0.28 e 0.86, con una media di 0.63), mentre la probe 2 presenta valori di correlazione decisamente inferiori (tra 0.05 e 0.76, con una media di 0.36). Queste differenze corrispondono a quanto visto per le probe 1 e 2 dell'analisi con il tracciante [ $^{11}\text{C}$ ]CUMI. I grafici a barre in Figura 6.13 sono stati determinati con lo stesso criterio descritto in precedenza, e descrivono in maniera quasi simmetrica lo stesso comportamento del tracciante [ $^{11}\text{C}$ ]CUMI. Ancora una volta una delle probe si riconferma inadatta a monitorare il gene per cui è stata disegnata.

# Conclusioni

Lo scopo di questa tesi era quello di integrare dati PET con dati di espressione genica per investigare sulla possibile correlazione tra i livelli di espressione per un determinato gene e l'effettiva densità di distribuzione dei neuro-recettori per i quali tale gene codifica. Per fare questo sono stati analizzati due sistemi recettoriali, quello dei recettori per gli oppioidi e quello per i serotoninergici. Inizialmente sono stati considerati in maniera separata i dati di espressione e le misure PET per valutare la consistenza dei dati. Successivamente si sono calcolate delle mappe di correlazione che permettessero di capire se fosse presente un legame tra i livelli di espressione (post-mortem) per un determinato gene e le misurazioni PET (in-vivo) dei recettori codificati da tale gene.

I risultati di questo studio hanno evidenziato che esiste un buon livello di correlazione tra le misure post-mortem di espressione genica per una determinata proteina e la distribuzione spaziale della proteina stessa, rilevata in-vivo. In particolare, nel caso del sistema per la serotonina si è rilevata la presenza di una forte correlazione tra i livelli di espressione del gene che codifica per i recettori *5-HT<sub>1A</sub>* e le loro misure PET. Dopo l'analisi di correlazione tra i soggetti è stato applicato un approccio a livello di ROI per valutare l'espressione nelle regioni anatomiche rilevanti (ricavate dai dati PET) evidenziano che le zone con un alto valore del potenziale di legame (direttamente legato al numero di recettori affini presenti), o del volume di distribuzione, corrispondono a zone con alti livelli di espressione dei i geni codificanti per i sistemi recettoriali studiati.

In tutti i casi erano disponibili più probe per ogni recettore analizzato. Per ogni gene valutato almeno una delle probe disponibili portava a livelli di correlazione diversi dalle altre. E' stato quindi possibile determinare l'efficacia delle

probe nel monitorare il gene per cui sono state disegnate. Le analisi e i grafici proposti permettono infatti di identificare eventuali probe inadatte per il sistema recettoriale studiato.

Per la prima volta è stata fatta un'analisi esaustiva sul legame tra espressione genica e stime PET. Tale analisi ha confermato il comportamento ipotizzato, cioè è stato rilevato un legame diretto tra i livelli di espressione per un determinato gene e la distribuzione dei recettori per i quali tale gene codifica. Per i sistemi recettoriali studiati è stato confermato un buon livello di correlazione tra i livelli di espressione genica e le misure PET. Studi più estensivi sarebbero necessari per verificare la validità di tale relazione anche in sistemi neuro-recettoriali diversi. Per questo motivo è stata sviluppata un'interfaccia grafica (GUI, *Graphical User Interface*) che permette un'applicazione semplice ed interattiva delle analisi condotte, su qualunque altro dataset.

# Appendice A

In questa Appendice vengono presentate tutte le mappe di correlazione prodotte. Ogni figura è divisa in 4 quadranti: il quadrante in alto a sinistra rappresenta la mappa di correlazione tra i soggetti del dataset di espressione, il quadrante in basso a destra rappresenta le correlazioni tra i soggetti PET e il quadrante in alto a destra racchiude i valori di correlazione tra i soggetti PET e i soggetti di espressione. In tutti i casi, le numerazioni a lato rappresentano i soggetti con il seguente schema: lungo l'asse x e y, da 1 a 4 ci sono i soggetti del dataset di espressione, poi da 1 a 10 (o 13 o 15) ci sono i soggetti del dataset PET.

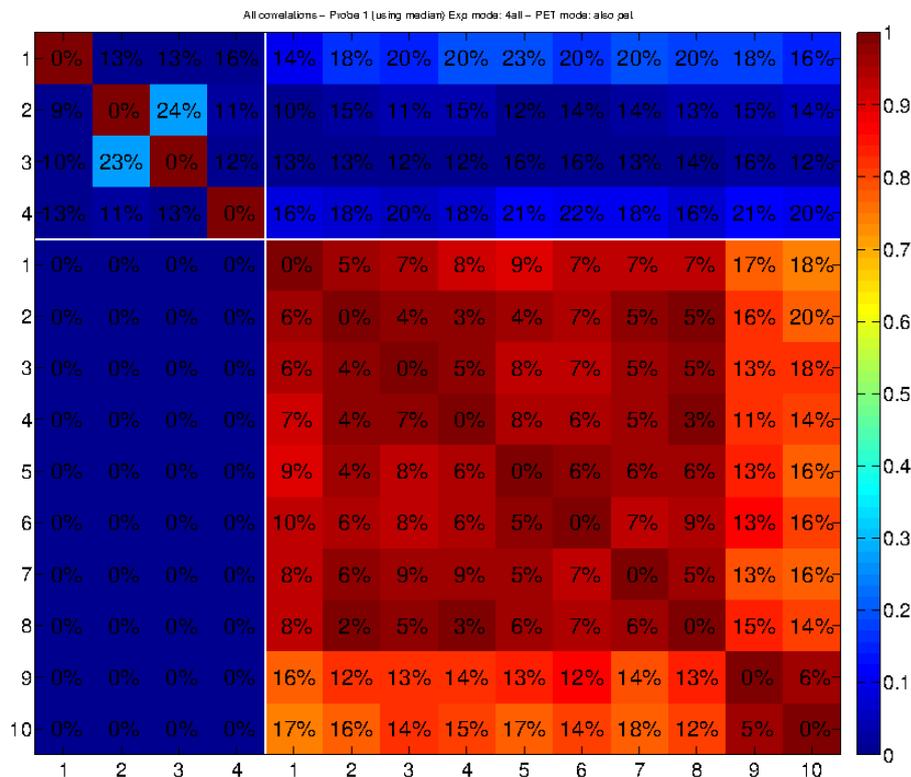


Figura 6.14: *DOP* (probe 1) vs.  $[^{11}\text{C}]\text{DPN}$  (BP).

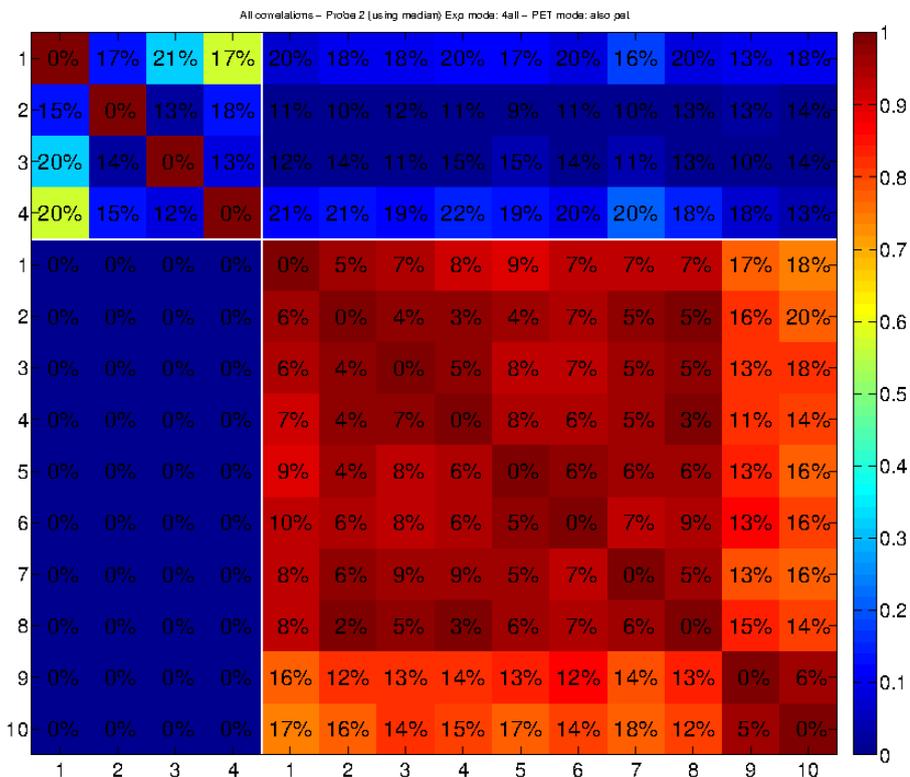


Figura 6.15: *DOP* (probe 2) vs.  $[^{11}\text{C}]\text{DPN}$  (BP).

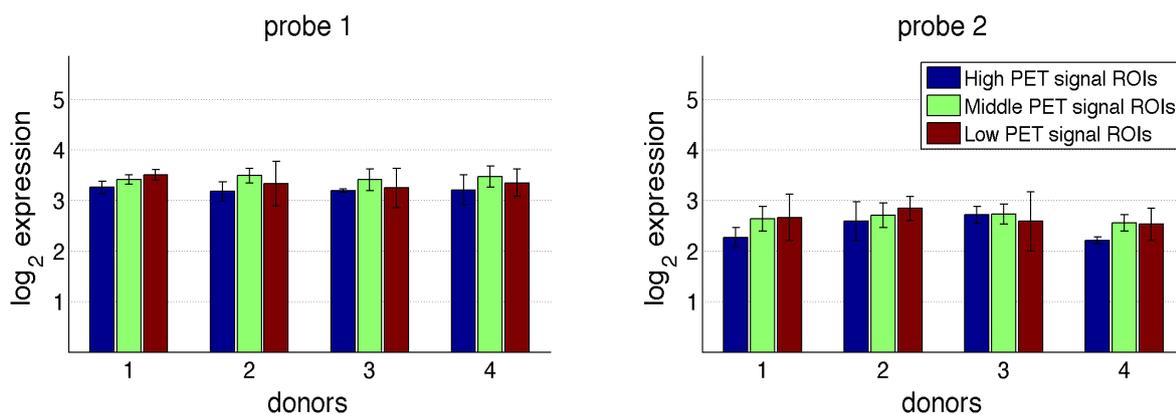


Figura 6.16: Analisi di espressione (recettore  $\delta$ ) delle zone ad alto, medio e basso segnale PET ( $[^{11}\text{C}]\text{DPN}$  BP)

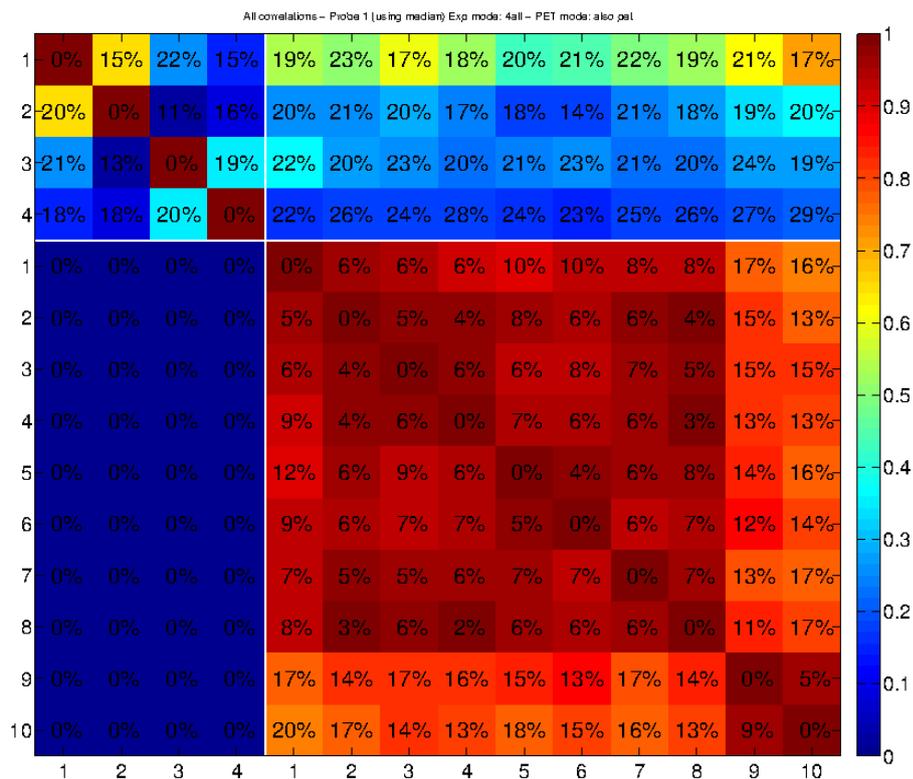


Figura 6.17: *KOP* (probe 1) vs.  $[^{11}\text{C}]\text{DPN}$  (BP).

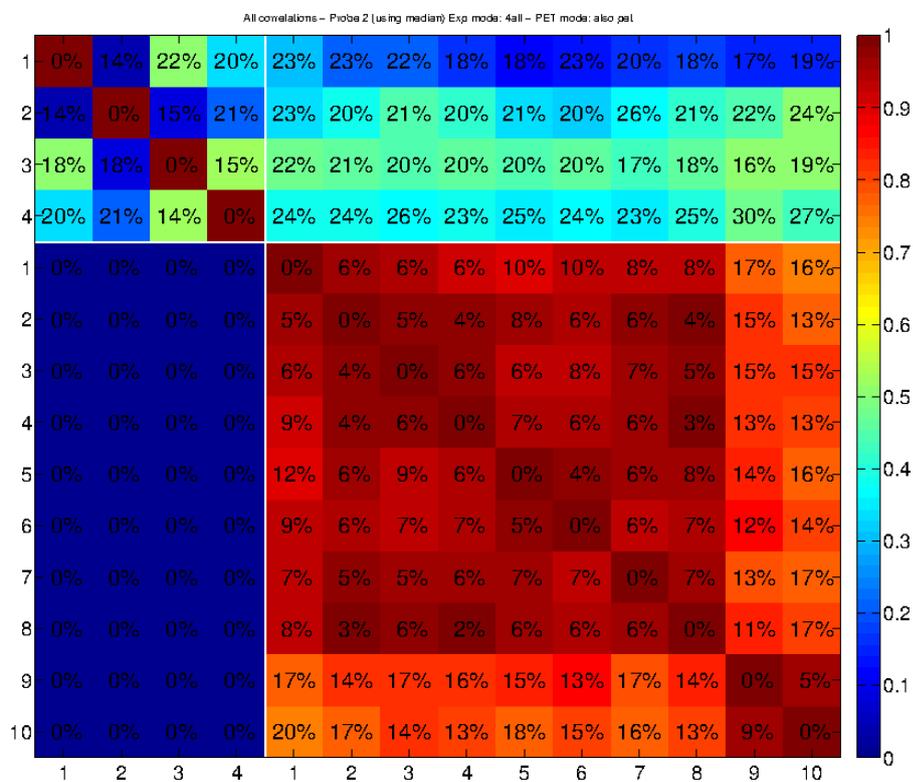


Figura 6.18: *KOP* (probe 2) vs.  $[^{11}\text{C}]\text{DPN}$  (BP).

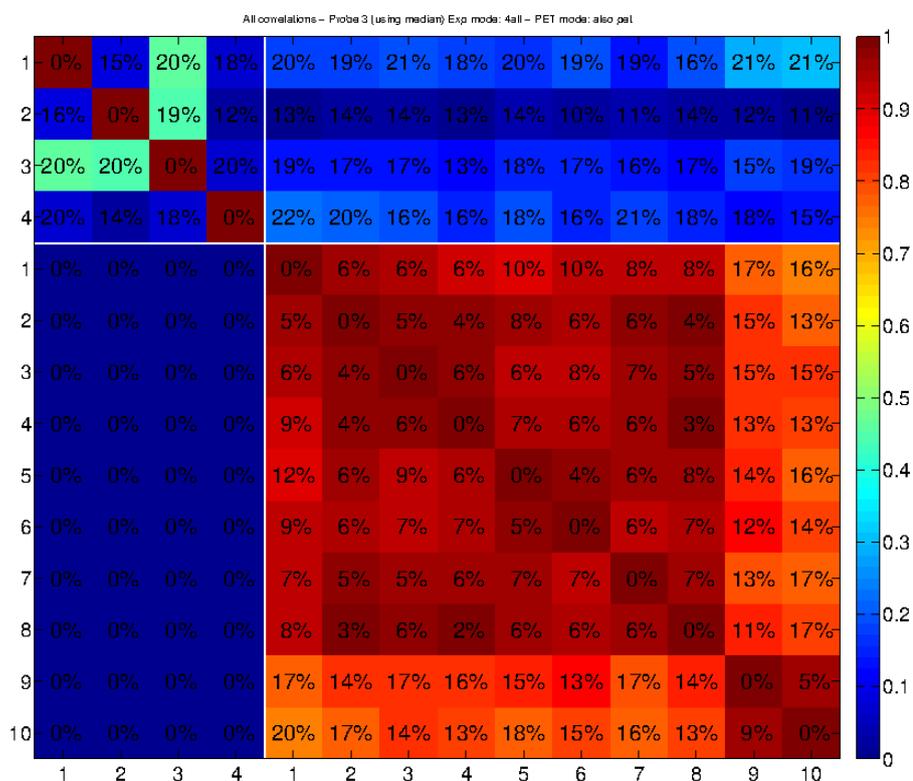


Figura 6.19: *KOP* (probe 3) vs.  $[^{11}\text{C}]\text{DPN}$  (BP).

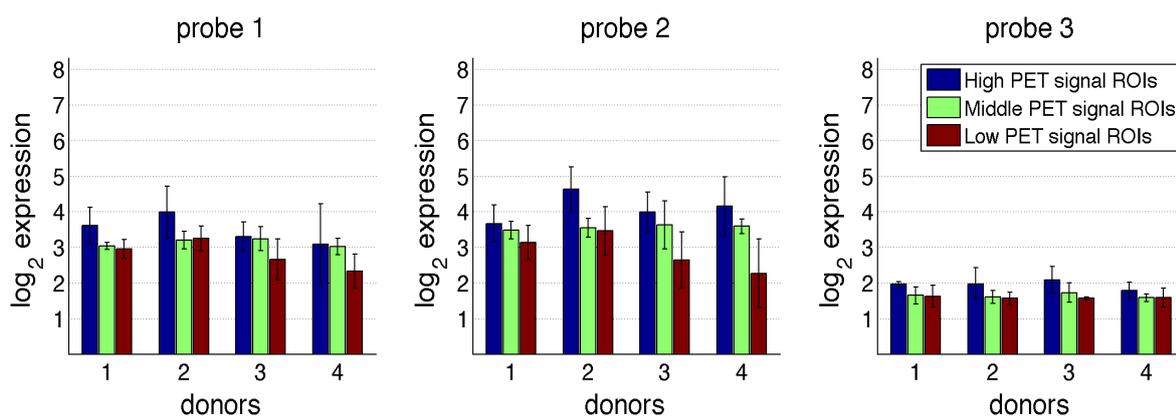


Figura 6.20: Analisi di espressione (recettore  $\kappa$ ) delle zone ad alto, medio e basso segnale PET ( $[^{11}\text{C}]\text{DPN}$  BP)

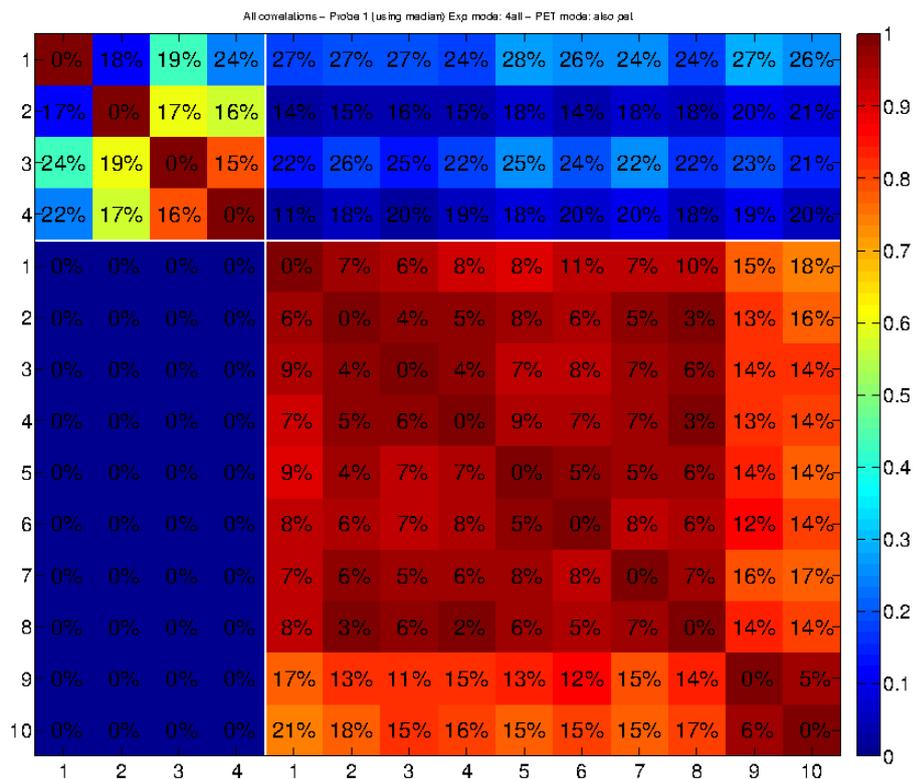


Figura 6.21: MOP (probe 1) vs.  $[^{11}\text{C}]$ DPN (BP).

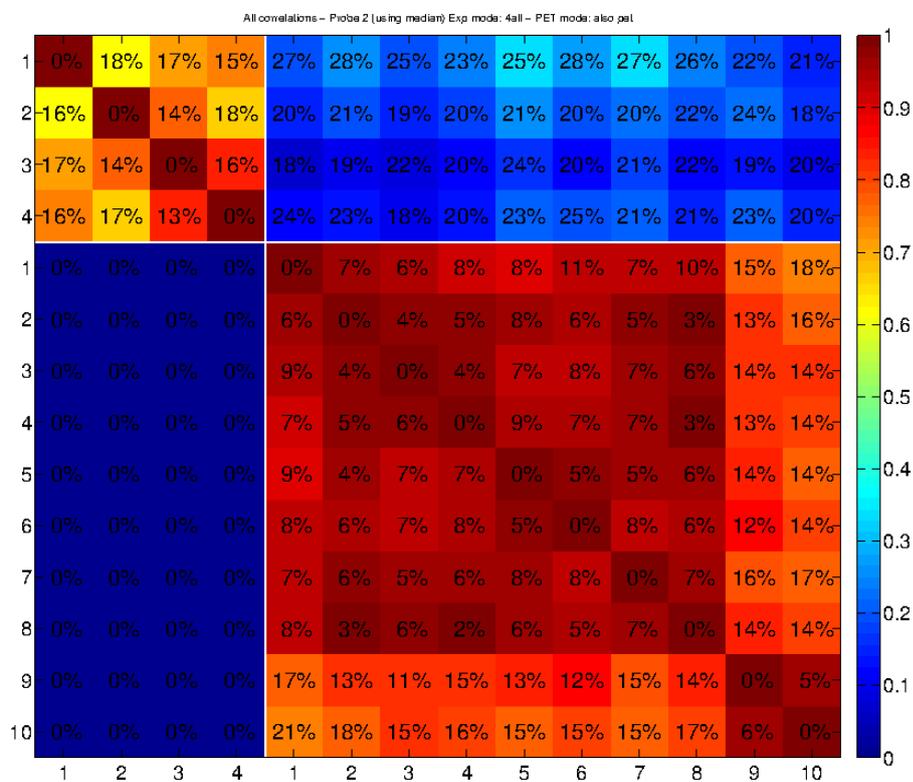


Figura 6.22: MOP (probe 2) vs.  $[^{11}\text{C}]$ DPN (BP).

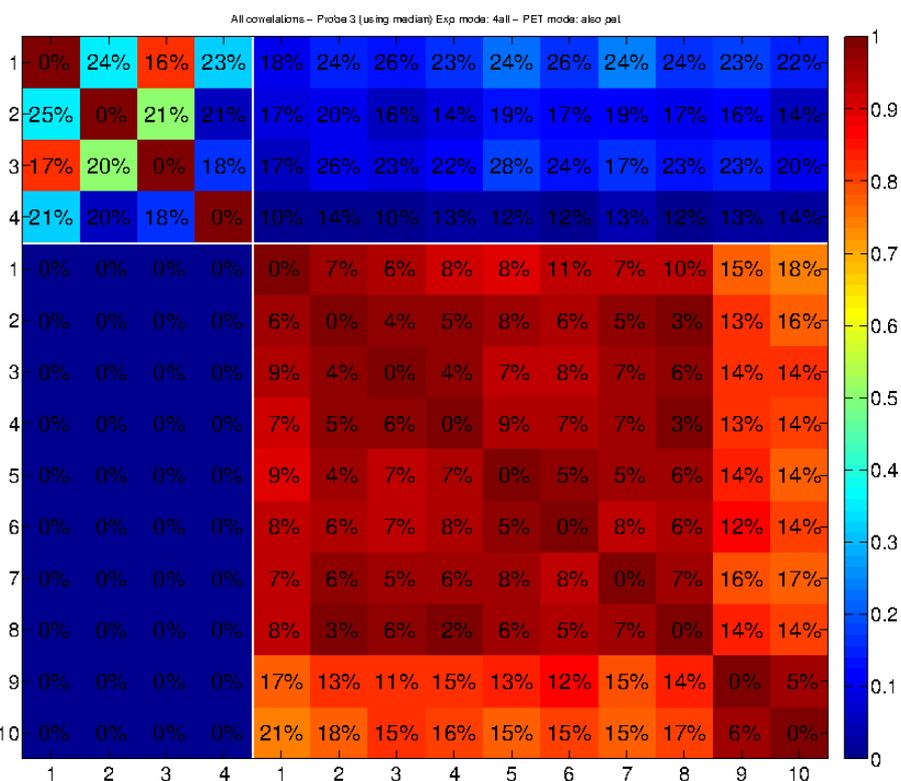


Figura 6.23: MOP (probe 3) vs.  $[^{11}\text{C}]\text{DPN}$  (BP).

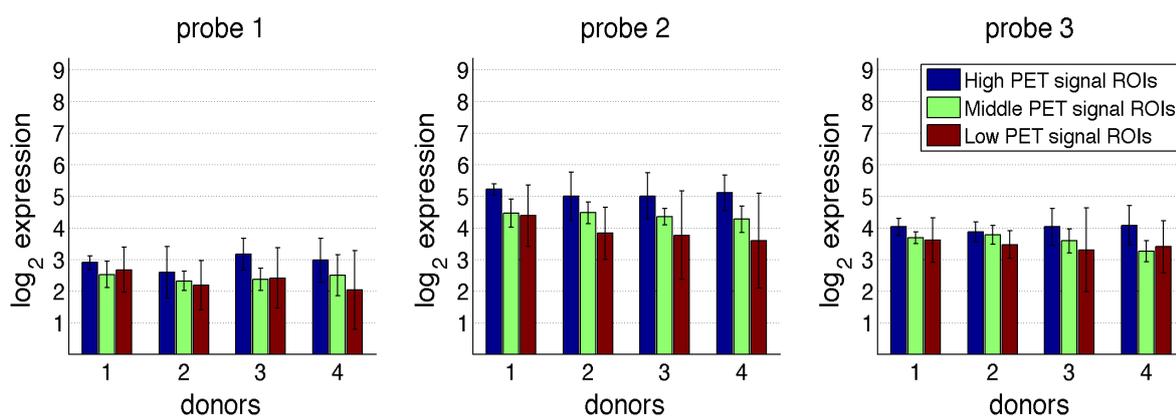


Figura 6.24: Analisi di espressione (recettore  $\mu$ ) delle zone ad alto, medio e basso segnale PET ( $[^{11}\text{C}]\text{DPN}$  BP)

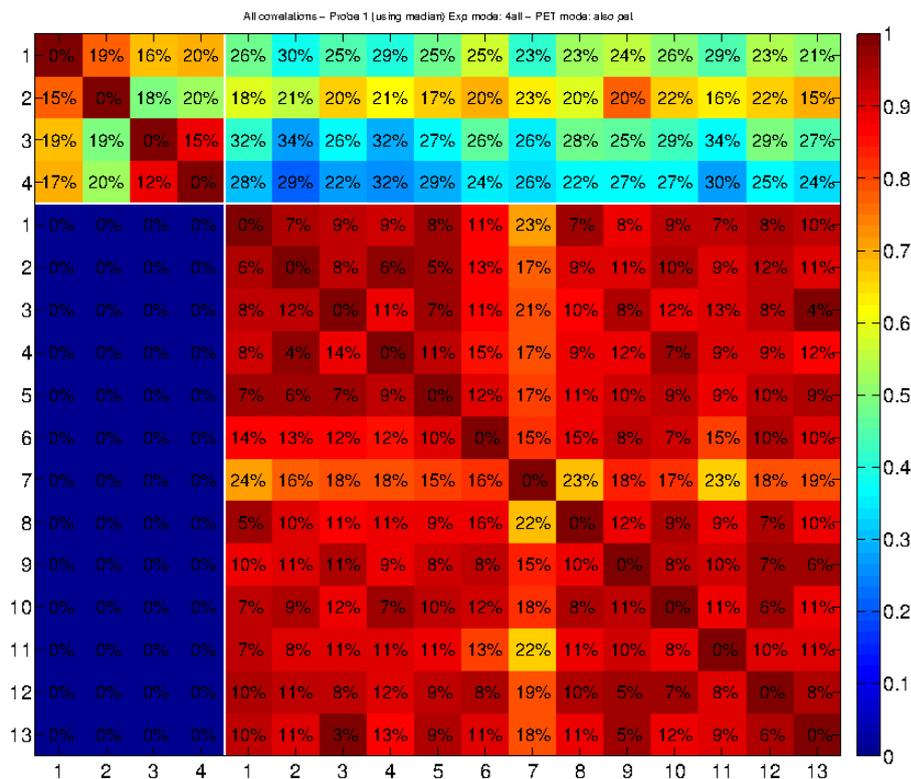


Figura 6.25:  $5HT_{1A}$  (probe 1) vs.  $[^{11}C]CUMI$  (BP).

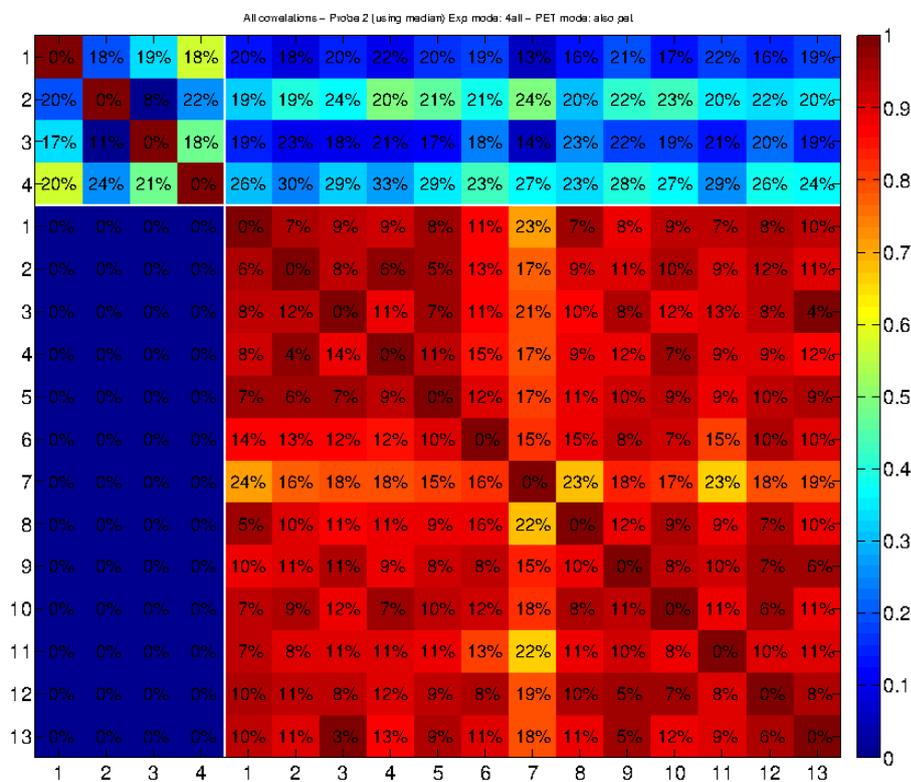


Figura 6.26:  $5HT_{1A}$  (probe 2) vs.  $[^{11}C]CUMI$  (BP).

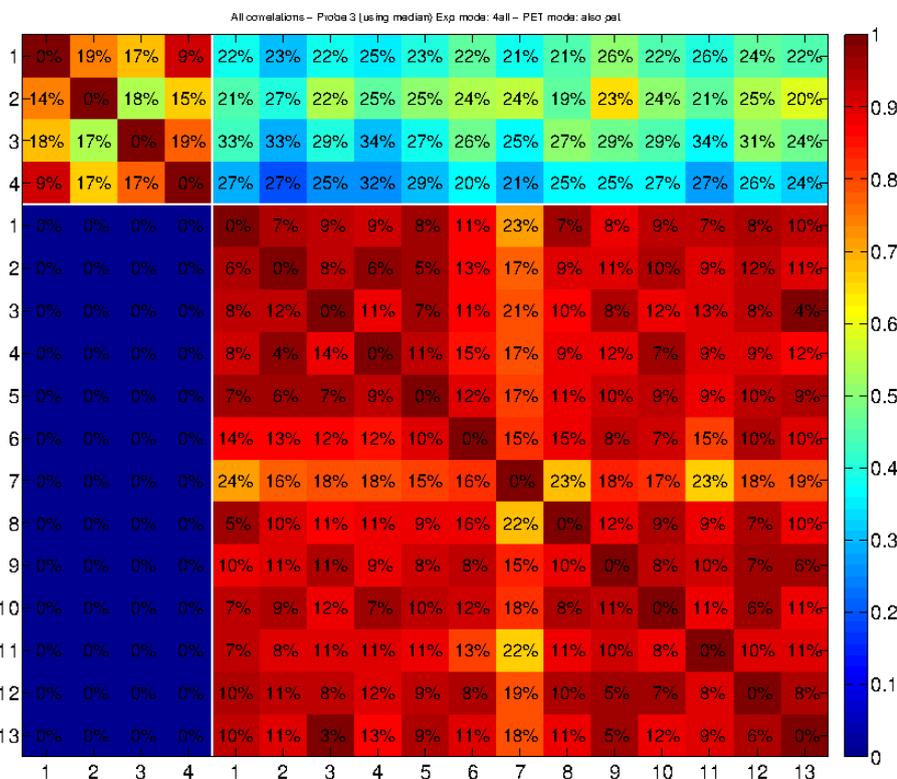


Figura 6.27:  $5HT_{1A}$  (probe 3) vs.  $[^{11}C]CUMI$  (BP).

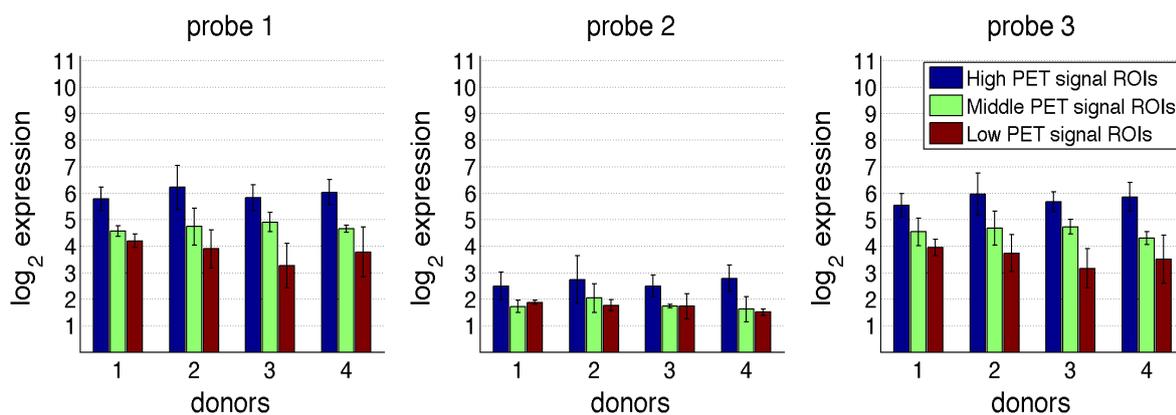


Figura 6.28: Analisi di espressione (recettore  $5HT_{1A}$ ) delle zone ad alto, medio e basso segnale PET ( $[^{11}C]CUMI$  BP)

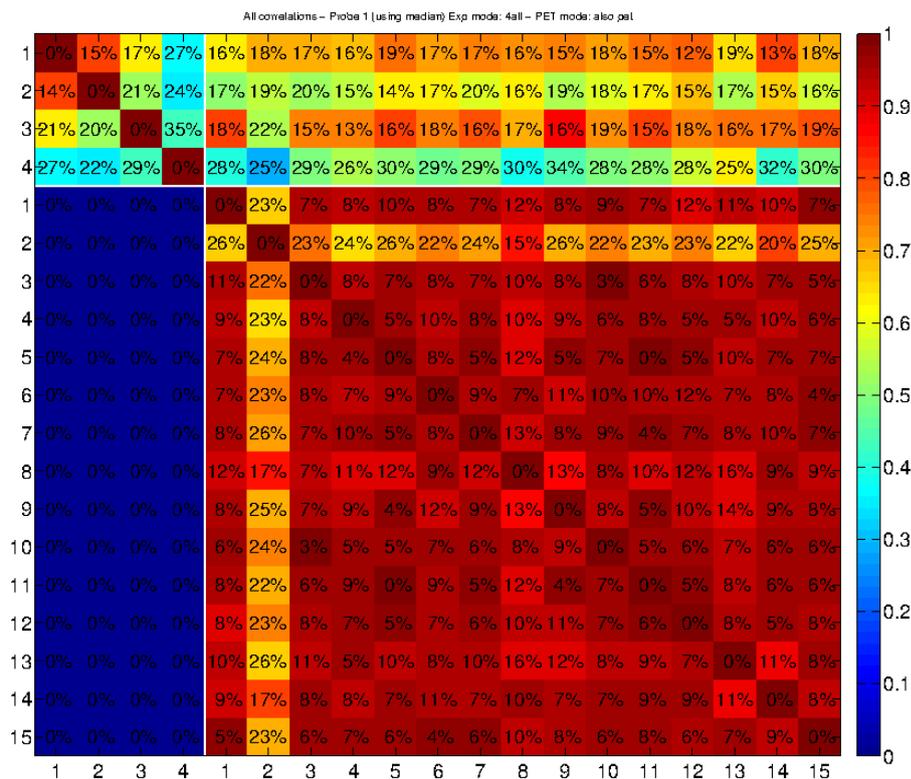


Figura 6.29:  $5HT_{1A}$  (probe 1) vs.  $[^{11}C]WAY (V_d)$ .

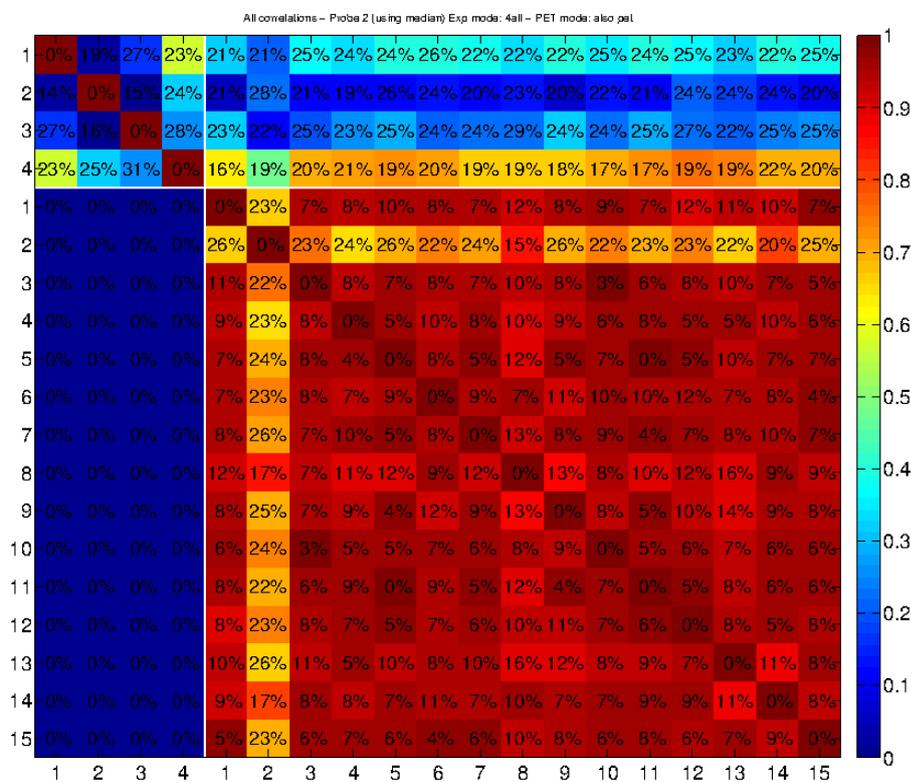


Figura 6.30:  $5HT_{1A}$  (probe 2) vs.  $[^{11}C]WAY (V_d)$ .

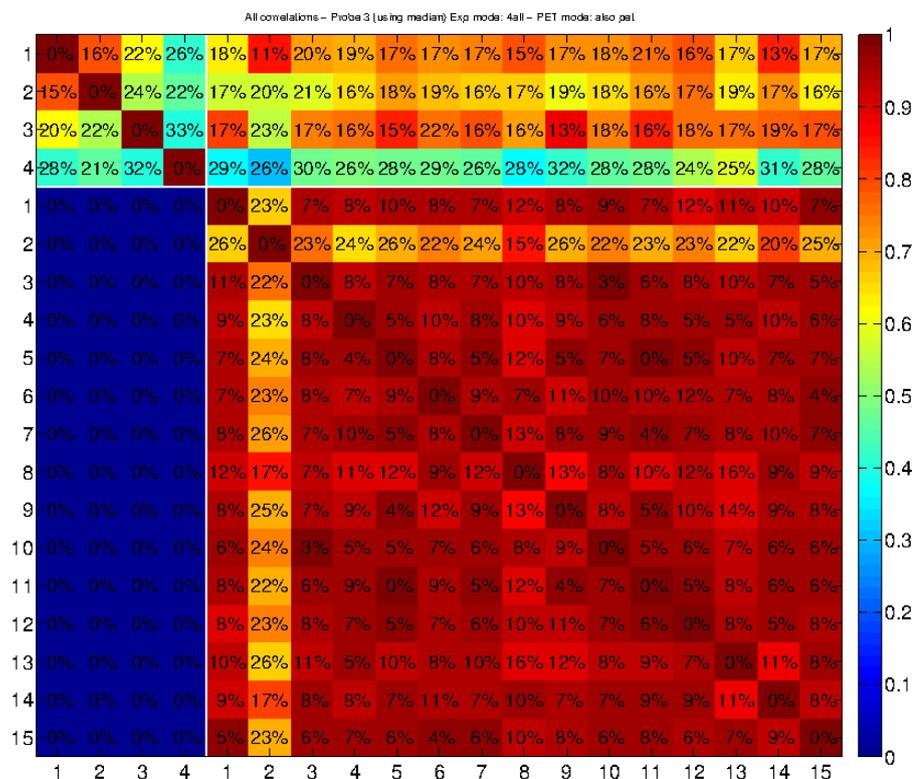


Figura 6.31:  $5HT_{1A}$  (probe 3) vs.  $[^{11}C]WAY$  ( $V_d$ ).

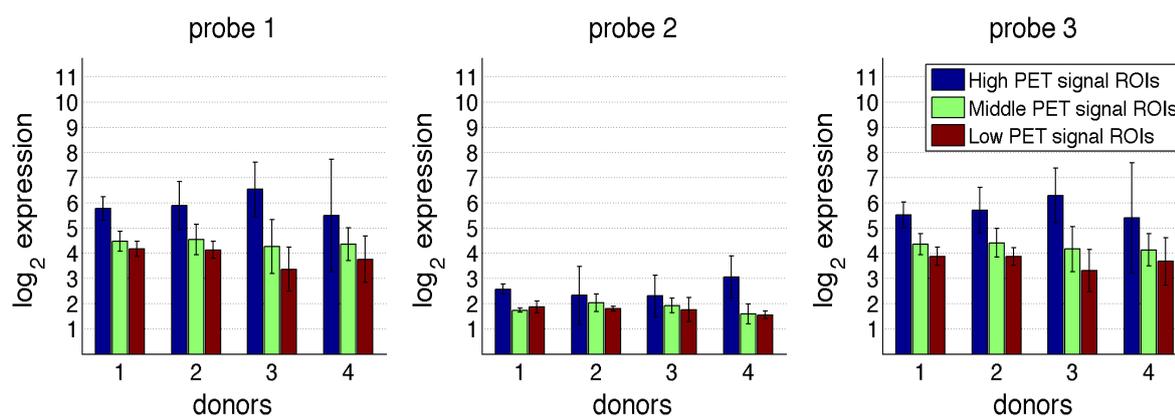


Figura 6.32: Analisi di espressione (recettore  $5HT_{1A}$ ) delle zone ad alto, medio e basso segnale PET ( $[^{11}C]CUMI$  BP)

# Appendice B

Nella Figura si può vedere uno screenshot della schermata principale dell'interfaccia sviluppata interamente in ambiente Matlab. La GUI permette sia di fare analisi sui singoli dataset di espressione e PET, sia di fare l'analisi combinata di entrambi. In output produce le figure esposte nell'Appendice A.

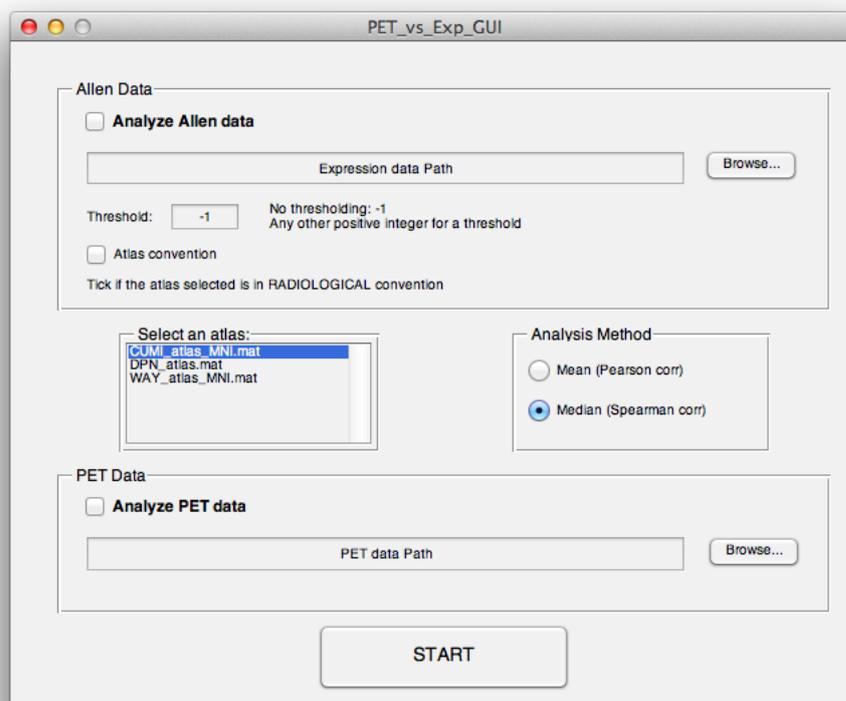


Figura 6.33: Screenshot della schermata principale della GUI: PET\_vs\_Exp\_GUI



# Bibliografia

- [1] E. Shumay, J. Logan, *et al.*, “Evidence that the methylation state of the monoamine oxidase A (MAOA) gene predicts brain activity of MAO A enzyme in healthy men,” *Epigenetics*, vol. 7, no. 10, pp. 1151–1160, Oct 2012.
- [2] A. Jacobs, J. Voges, R. Reszka, *et al.*, “Positron-emission tomography of vector-mediated gene expression in gene therapy for gliomas,” *Lancet*, vol. 358, no. 9283, pp. 727–729, Sep 2001.
- [3] M. A. Oquendo, R. S. Hastings, Y. Y. Huang, N. Simpson, R. T. Ogden, X. Z. Hu, D. Goldman, V. Arango, R. L. Van Heertum, J. J. Mann, and R. V. Parsey, “Brain serotonin transporter binding in depressed patients with bipolar disorder using positron emission tomography,” *Arch. Gen. Psychiatry*, vol. 64, no. 2, pp. 201–208, Feb 2007.
- [4] M. J. Hawrylycz, E. S. Lein, A. L. Guillozet-Bongaarts, *et al.*, “An anatomically comprehensive atlas of the adult human brain transcriptome,” *Nature*, vol. 489, no. 7416, pp. 391–399, Sep 2012.
- [5] J. McDonald and D. G. Lambert, “Opioid mechanisms and opioid drugs,” *Anaesthesia & Intensive Care Medicine*, vol. 9, no. 1, pp. 33–37, 2008.
- [6] A. D. Corbett, G. Henderson, A. T. McKnight, and S. J. Paterson, “75 years of opioid research: the exciting but vain quest for the Holy Grail,” *Br. J. Pharmacol.*, vol. 147 Suppl 1, pp. S153–162, Jan 2006.
- [7] Fine, P. G., and R. K. Portenoy, “Chapter 2: The endogenous opioid systems.” in *A Clinical Guide to Opioid Analgesia*. McGraw Hill, 2004.

- 
- [8] K. Hasebe, K. Kawai, T. Suzuki, K. Kawamura, T. Tanaka, M. Narita, H. Nagase, and T. Suzuki, "Possible pharmacotherapy of the opioid kappa receptor agonist for drug dependence," *Ann. N. Y. Acad. Sci.*, vol. 1025, pp. 404–413, Oct 2004.
- [9] R. Glennon, M. Dukat, and R. Westkaemper, "Serotonin Receptor Subtypes and Ligands," *American College of Neurophyscopharmacology*, Apr 2008.
- [10] M. M. Ter-Pogossian, M. E. Phelps, E. J. Hoffman, and N. A. Mullani, "A positron-emission transaxial tomograph for nuclear imaging (PETT)," *Radiology*, vol. 114, no. 1, pp. 89–98, Jan 1975.
- [11] M. E. Phelps, "Positron emission tomography provides molecular imaging of biological processes," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 97, no. 16, pp. 9226–9233, Aug 2000.
- [12] G. Rizzo, F. E. Turkheimer, S. Keihaninejad, S. K. Bose, A. Hammers, and A. Bertoldo, "Multi-Scale hierarchical generation of PET parametric maps: application and testing on a [11C]DPN study," *Neuroimage*, vol. 59, no. 3, pp. 2485–2493, Feb 2012.
- [13] A. Hammers, R. Allom, M. J. Koeppe, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, and J. S. Duncan, "Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe," *Hum Brain Mapp*, vol. 19, no. 4, pp. 224–247, Aug 2003.
- [14] J. Logan, J. S. Fowler, N. D. Volkow, *et al.*, "Graphical analysis of reversible radioligand binding from time-activity measurements applied to [N-11C-methyl]-(-)-cocaine PET studies in human subjects," *J. Cereb. Blood Flow Metab.*, vol. 10, no. 5, pp. 740–747, Sep 1990.
- [15] V. J. Cunningham and T. Jones, "Spectral analysis of dynamic PET studies," *J. Cereb. Blood Flow Metab.*, vol. 13, no. 1, pp. 15–23, Jan 1993.
- [16] A. Bertoldo, P. Vicini, G. Sambuceti, A. A. Lammertsma, O. Parodi, and C. Cobelli, "Evaluation of compartmental and spectral analysis models of

- [18F]FDG kinetics for heart and brain studies with PET,” *IEEE Trans Biomed Eng*, vol. 45, no. 12, pp. 1429–1448, Dec 1998.
- [17] X. Chen, S. T. Cheung, S. So, *et al.*, “Gene expression patterns in human liver cancers,” *Mol. Biol. Cell*, vol. 13, no. 6, pp. 1929–1939, Jun 2002.
- [18] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, “Quantitative monitoring of gene expression patterns with a complementary DNA microarray,” *Science*, vol. 270, no. 5235, pp. 467–470, Oct 1995.
- [19] T. K. Karakach, R. M. Flight, S. E. Douglas, and P. D. Wentzell, “An introduction to dna microarrays for gene expression analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 104, no. 1, pp. 28–52, 2010.
- [20] A. C. Pease, D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, and S. P. Fodor, “Light-generated oligonucleotide arrays for rapid DNA sequence analysis,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 91, no. 11, pp. 5022–5026, May 1994.
- [21] S. Selvaraj, F. Turkheimer, L. Rosso, P. Faulkner, E. Mouchlianitis, J. P. Roiser, P. McGuire, P. J. Cowen, and O. Howes, “Measuring endogenous changes in serotonergic neurotransmission in humans: a [11C]CUMI-101 PET challenge study,” *Mol. Psychiatry*, vol. 17, no. 12, pp. 1254–1260, Dec 2012.
- [22] S. K. Bose, M. A. Mehta, S. Selvaraj, O. D. Howes, R. Hinz, E. A. Rabiner, P. M. Grasby, F. E. Turkheimer, and V. Murthy, “Presynaptic 5-HT<sub>1A</sub> is related to 5-HTT receptor density in the human brain,” *Neuropsychopharmacology*, vol. 36, no. 11, pp. 2258–2265, Oct 2011.
- [23] Allen Institute for Brain Science, “Allen Brain Atlas.” [Online]. Available: <http://www.brain-map.org>
- [24] I. S. Gousias, D. Rueckert, R. A. Heckemann, L. E. Dyet, J. P. Boardman, A. D. Edwards, and A. Hammers, “Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest,” *Neuroimage*, vol. 40, no. 2, pp. 672–684, Apr 2008.



## Ringraziamenti

Innanzitutto ringrazio la Professoressa Alessandra Bertoldo per avermi dato la possibilità di svolgere questa tesi con lei e per i preziosi insegnamenti ricevuti nei suoi corsi e durante lo svolgimento di questa tesi.

Un grazie sentito a Gaia e Mattia per la grande disponibilità e il tempo dedicatomi durante questi mesi di lavoro. Dalle cose più semplici a quelle più complesse i loro consigli si sono sempre dimostrati validi.

Ringrazio con affetto i miei genitori, mio fratello e i miei nonni per il grande supporto e per essermi sempre stati vicini in tutti questi anni universitari.

Grazie agli amici che mi conoscono da una vita (Bajo, Bosca e Alex) per essere sempre rimasti uniti nonostante le e mille avversità e le diverse strade intraprese.

Un grazie speciale ai miei amici universitari: Albe per essersi dimostrato un insostituibile amico e compagno di mille giornate, di studio e non, passate al DEI; Chiara per essere la grandissima amica che è, sempre presente e disponibile, per aver sopportato il mio carattere e per le fondamentali sessioni di studio fatte assieme, che hanno reso piacevoli e semplici anche i corsi più difficili e noiosi; Ila per essere stata la voce della coscienza del gruppo e per avermi sopportato in tutti questi anni; Ele per essere stata un'amica sempre cara e gentile; Chiara-di-Albe per condiviso il suo Albe con me nelle lunghe giornate universitarie e per essere una buona amica con cui condividere la passione per il Giappone e le arti marziali; Marco e Roby per essere stati dei buoni compagni con cui fare squadra in questi anni. Infine, un grazie anche ad Anna, Bordi, Mauro, Alberto, Alessandro ed Elena per le belle chiacchierate che hanno reso i noiosissimi viaggi in treno decisamente più piacevoli.

Grazie a tutti voi per aver reso ogni giorno della mia vita universitaria degno di essere ricordato.