

Titolo.

Bioinformatica e data mining per la ricerca preclinica

Durata.

12 mesi

Piano di attività.

Questo progetto si propone di formare una figura professionale che, partendo da competenze bioinformatiche, sappia adattare e progettare nuovi algoritmi di analisi basati su modelli di data mining e intelligenza artificiale che trovino specifica applicazione nella ricerca preclinica dedicata al “drug discovery” compresi i test di tossicità previsti dal regolatorio e propedeutici alla traslazione in studi clinici di fase 1¹, con l’obiettivo primario di migliorare la predittività dei dati preclinici, e con l’obiettivo secondario di ridurre il numero di animali impiegati nella ricerca preclinica.

Per aumentare la predittività dei modelli preclinici ottenuti con sistemi in vitro (colture cellulari convenzionali, sferoidi e organoidi) e da sperimentazione animale, da alcuni anni si teorizza la possibile applicazione di strumenti di data mining e intelligenza artificiale sia nel disegno sperimentale che nell’analisi dei dati, ma tale teorizzazione si basa sostanzialmente su meta-analisi di dati pubblicati o modelli in silico, e deriva molto spesso dalla passiva applicazione di modelli di analisi sviluppati per altri contesti di “big-data analysis”².

Se l’applicazione di questi strumenti di analisi a dati empirici è ancora agli albori, è di sempre maggiore interesse visto che anche la raccolta dati nella ricerca preclinica, sia su sistemi cellulari che sull’animale, ricorre ormai sempre più diffusamente a tecnologia “high-throughput”, con produzione di una massa di dati non sempre correttamente analizzati, e di cui certamente si sfrutta una piccola parte del potenziale informativo. E’ quindi necessario sviluppare una metodologia per affrontare in modo sistematico l’estrazione di modelli predittivi e prescrittivi da dati preclinici che siano efficaci nell’ambito della medicina personalizzata. Questa metodologia, oltre ad individuare le features che influenzano i risultati, ed avere quindi un impatto significativo sul disegno sperimentale, deve produrre linee guida per la raccolta, la gestione e l’analisi di dati e fornire modelli che siano spiegabili per un esperto umano, siano in grado di unire conoscenza da esperti con modelli estratti dai dati, e siano traslabili su altri casi.

Nello specifico, obiettivi del progetto sono: (i) organizzare in maniera appropriata la raccolta e il repository dei dati sperimentali, da integrare nei sistemi di gestione qualità dei laboratori di ricerca preclinica; (ii)sviluppare nuovi modelli spiegabili e adattabili, disegnati per le finalità previste dalla ricerca preclinica per studi di sicurezza e di efficacia.

¹ Choukrallah MA, Martin F, Sierro N, Hoeng J, Ivanov NV, Peitsch MC. Computational Epigenomics: From Fundamental Research to Disease Prediction and Risk Assessment. In: Husi H, editor. Computational Biology [Internet]. Brisbane (AU): Codon Publications; 2019 Nov 21. Chapter 7.

² Hutson M. Artificial intelligence faces reproducibility crisis. Science. 2018 Feb 16;359(6377):725-726.

Come casi-studio, si analizzeranno i dati di trascrittomica (ottenuti da liquidi biologici e tessuto) e dati descrittivi del fenotipo neurologico (comportamenti complessi quali locomozione, coordinamento del passo, apprendimento e memoria) ottenuti da animali portatori di modelli di patologia neurologica, per correlare i dati molecolari a un fenotipo quantificabile, espressione anche della gravità di malattia. Saranno considerati i dati già disponibili nel repository storico del laboratorio (UO Medicina Traslazionale del CIRI-SDV), relativi al modello murino di malattia di Alzheimer e trattamenti con elettroceutica, per il quale sono disponibili dati da studi longitudinale di apprendimento e memoria e dati di trascrittomica, e la lesione contusiva del midollo spinale e trattamenti con farmaci, biomateriali, elettroceutica, per il quale sono disponibili dati da studi longitudinale di locomozione e analisi del passo e dati di trascrittomica. Il set di dati è incrementale, essendo composto dai dati già presenti nel repository del laboratorio, la cui integrità è garantita dalla raccolta e conservazione secondo i principi ALCOA³ previsti dai principi di “buona pratica” delle mouse clinics (vedi INFRAFRONTIERS, EUMODIC, EMMA), e recepito dal sistema di gestione qualità del laboratorio, a cui si aggiungeranno progressivamente dati prodotti da nuove sperimentazioni, che comprenderanno anche dati derivati da animali sottoposti a terapia (farmacologica, terapie avanzate).

Risultati attesi

Al termine del progetto si prevede di raggiungere i seguenti obiettivi:

- Aver formato una nuova figura professionale che, partendo da competenze bioinformatiche, abbia recepito i principi di buone pratiche per la conduzione di studi preclinici su modelli animali di malattia;
- Aver creato linee guida per costruire repository di dati con caratteristiche di interoperatività fra strumenti di misura di comportamenti complessi, di dati di trascrittomica e proteomica e software di analisi;
- Prodotto una metodologia per la costruzione di modelli di analisi dell'interazione fra fenotipo neurologico e dati di trascrittomica (tissutale) e proteomica (biomarkers plasmatici e liquorali e lo sviluppo di un prototipo su uno specifico set di dati.
- Stesura di un manoscritto sulla “working hypothesis” da proporre ad una rivista di Laboratory Animal Science, anche nell’ottica di una migliore gestione della power calculation mirata alla “reduction” del numero di animali da includere in sperimentazione.

³ Rattan AK. Data Integrity: History, Issues, and Remediation of Issues. PDA J Pharm Sci Technol. 2018 Mar-Apr;72(2):105-116.