

RD-CONNECT: AN INTEGRATED PLATFORM CONNECTING REGISTRIES, BIOBANKS AND CLINICAL BIOINFORMATICS FOR RARE DISEASE RESEARCH “

Sabina Gainotti, Claudio Carta, Paola Torreri, Domenica Taruscio
Centro Nazionale Malattie Rare, Istituto Superiore di Sanità, Roma

Necessità di integrazione della ricerca sulle malattie rare

L'80% delle malattie rare ha una componente genetica, e le nuove tecnologie Omics oggi disponibili stanno rendendo possibile lo sviluppo di trattamenti su base genetica per molte di queste malattie. Il sequenziamento del genoma umano completato per la prima volta nel 2003 ha richiesto il lavoro di centinaia di scienziati per più di 10 anni e per un costo di oltre € 3 miliardi. Oggi lo stesso risultato è ottenibile in pochi giorni mediante un unico strumento di sequenziamento ad un costo inferiore a € 10.000, e il costo dell'operazione è destinato ancora a diminuire. Le nuove tecnologie Omics permettono di generare una mole di dati senza precedenti nella storia della ricerca biomedica. Ad oggi è già stato sequenziato l'intero genoma di diverse migliaia di esseri umani da cui è derivata un'esplosione di dati di diversi ordini di grandezze. Eppure, la capacità di produrre dati non è stata ancora accompagnata da un'aumentata capacità di analizzarli. Inoltre, il trasferimento dei dati da un luogo all'altro e la condivisione di dati tra ricercatori è una pratica tutt'altro che comune. Non da ultimo, nella scoperta di nuove diagnosi e terapie il sequenziamento è solo una parte del processo e non può in alcun modo sostituire l'esperienza del clinico e la descrizione accurata del singolo caso, mentre diventa sempre più importante saper combinare i dati genetici con dati clinici e fenotipici il più possibile dettagliati secondo gli obiettivi della cosiddetta “precision medicine” (Robinson, 2012).

Il progresso della biomedicina richiede dunque la combinazione e l'integrazione di dati di genomica, trascrittomica, proteomica e metabolomica con informazioni fenotipiche dettagliate (fenomica), provenienti da diversi centri e progetti a livello globale. I registri di pazienti, le biobanche e i database di bioinformatica sono tutti strumenti necessari a costituire infrastrutture per la ricerca “omics”, traslazionale e clinica. Tuttavia, allo stato attuale continuano a moltiplicarsi gli sforzi individuali dedicati a singole malattie o gruppi di malattie, e gli stessi sforzi restano in buona parte isolati con scarsa interoperabilità e quasi nessun collegamento sistematico di informazioni cliniche con informazioni di tipo genetico, di disponibilità di biomateriali o insiemi di dati raccolti nei trial clinici, e questo limita le possibilità di ricerca diagnostica e lo sviluppo di terapie personalizzate per le persone con malattia rara.

Ruolo dei registri di pazienti nella ricerca sulle malattie rare

La creazione di registri di pazienti e patologie ha dei vantaggi molto evidenti nelle malattie rare, in particolare nelle malattie ultra rare per cui esistono conoscenze sparse e frammentate in

un numero limitato di centri al mondo. Nessuna singola istituzione, e in molti casi nessun Paese, ha un numero sufficiente di pazienti per condurre ricerche cliniche e traslazionali generalizzabili. Attualmente esistono registri di pazienti solo per una piccola parte delle malattie rare e, al contrario, per alcune malattie rare possono esistere più registri con obiettivi e dataset diversi. È dunque necessario promuovere un approccio sistematico e standardizzato per la loro creazione, manutenzione e accessibilità a livello internazionale e globale.

Secondo l'ultimo report di ORPHANET dedicato ai registri di malattie rare (ORPHANET, 2015) in Europa sono attivi 615 registri di malattie rare. Se il dato può sembrare incoraggiante, la maggior parte di queste banche dati di fatto non è ancora utilizzabile per la ricerca -Omics. Il progetto EPIRARE (coordinato dal Centro Nazionale Malattie Rare, CNMR) tra le altre cose aveva evidenziato una generale mancanza di armonizzazione nei sistemi di codifica usati dai registri e nel tipo di dati raccolti e una mancanza di standardizzazione e controllo di qualità nel processo di raccolta, conservazione e analisi dei dati stessi. Inoltre EPIRARE aveva evidenziato una tendenza alla non condivisione dei dati da parte dei registri e una difficoltà a garantire la propria sostenibilità nel medio e lungo periodo.

Eppure esistono esempi di registri condotti su scala internazionale che hanno permesso di ottenere grandi risultati in termini di ricerca diagnostica e di identificazione di nuove terapie. Spesso in questi registri i dati sono raccolti dai professionisti e dagli stessi pazienti (patient reported outcomes, PRO) con un rigoroso controllo di qualità e aggiornamenti regolari e cadenzati. Ad esempio, le reti di ricerca clinica dedicate alla fibrosi cistica, la Corea di Huntington e le malattie neuromuscolari hanno concentrato i propri sforzi dando luogo ad una "nuova generazione" di registri di pazienti stratificati per gene che comprendono una descrizione dettagliata del genotipo e fenotipo, procedure standardizzate per garantire la qualità dei dati stessi e meccanismi di governance mirati a mantenere un legame forte con la comunità dei pazienti in un quadro etico sicuro.

Questi registri di nuova generazione hanno reso possibile la conduzione di numerosi studi e trial clinici, generando investimenti da parte di partner commerciali, e assicurando trattamenti migliori e più mirati per i pazienti stessi.

Obiettivi del progetto

La piattaforma centrale di RD-Connect metterà a disposizione della comunità di ricerca sulle malattie rare informazioni di tipo clinico e fenotipico con informazioni di disponibilità di biomateriali e dati Omics, integrandoli con i più avanzati strumenti di analisi.

Contributo del CNMR: obiettivi e risultati raggiunti all'interno del WP2

Il CNMR coordina il Workpackage 2, dedicato ai database e registri di pazienti. Obiettivo principale del WP2 è contribuire ad armonizzare e standardizzare l'attività di raccolta dati dei database e dei registri di pazienti con malattie rare per definire coorti di pazienti ben caratterizzate e stratificate per genotipi specifici che permettano la conduzione di ricerche di Omics, ricerche traslazionali e trial clinici. La standardizzazione dei registri esistenti è ottenuta identificando e favorendo l'adozione di elementi di dati comuni e ontologie per la descrizione fenotipica e promuovendo la condivisione e lo scambio dei dati attraverso meccanismi di sostegno alla disseminazione dei risultati, necessaria a garantire la sostenibilità di lungo periodo dei registri stessi. Il WP2 è diviso in 8 "Tasks" svolti dallo stesso CNMR e da altri partner del progetto:

- 1) Mappatura dei registri/database esistenti e valutazione della loro utilità per la ricerca (*trial readiness*) rispetto alle migliori pratiche identificate.
- 2) Creazione di un “Core Implementation Group” di Registri che mettano a disposizione le procedure operative standard (SOP) elaborate per i propri obiettivi specifici e possano testare e adottare per primi le utilità della piattaforma (*early implementer*).
- 3) Creazione di un catalogo consultabile online di banche dati e registri di malattie rare.
- 4) Definizione di elementi comuni di dati (CDE), sistemi di codifica standardizzati e ontologie per la descrizione fenotipica all'interno dei registri e procedure operative standard per la raccolta, l'archiviazione e il recupero dei dati stessi.
- 5) Sviluppo di un registro pilota per la raccolta di dati clinici di pazienti senza una diagnosi genetica all'interno dei B-Projects.
- 6) Sviluppo di un'architettura per consentire l'interoperabilità con il *Global Rare Disease Registry* (GRDR) del *National Center for Advancing Translational Sciences* (NCATS) dei National Institutes of Health (NIH) Americani per condividere i dati de-identificati dei pazienti a livello globale.
- 7) Sviluppo di materiali formativi, laboratori di formazione e “cassette per gli attrezzi” o “toolkit” online per i nuovi registri e database.
- 8) Definizione di opzioni di sostenibilità per i database e registri.

Ad oggi sono stati raggiunti diversi obiettivi e in particolare:

- A) È stato creato e periodicamente aggiornato un database di tutti i registri/database esistenti di malattie rare a partire da diverse fonti di informazione, in particolare il portale Orphanet, EPIRARE, HQIP, EURORDIS, Treat NMD, CORDIS, E-RARE.
- B) Insieme al WP3 sulle biobanche, sono stati identificati i principali criteri di valutazione dei registri che ne definiscono l'utilità per la ricerca (*Trial readiness*), in particolare le caratteristiche generali dell'Organizzazione ospite, le fonti di finanziamento, i tipi di dati raccolti e l'uso di ontologie per la descrizione fenotipica, la copertura geografica, la raccolta di biomateriale e altro materiale clinico e di Imaging; i tipi di malattie di interesse del registro con le relative codifiche; gli standard adottati per la raccolta, archiviazione e analisi dei dati; le politiche e procedure per accedere ai dati stessi, la documentazione prodotta dal registro.
Questi criteri sono stati adottati per creare il profilo di descrizione del registro detto anche “ID-Card” che costituisce la base del catalogo online dei registri e biobanche sviluppato insieme al WP3.
- C) Il catalogo online di RD-Connect, pubblicamente disponibile e consultabile a partire da novembre 2014, è attualmente popolato da diversi registri (N=55) che hanno reso visibile il proprio numero di casi registrati alla comunità dei ricercatori. Questa informazione può essere importante per diversi tipi di utenti del catalogo, a partire dai singoli ricercatori che cercano casi simili al proprio, alle case farmaceutiche che hanno bisogno di sapere se esistono i numeri per condurre una sperimentazione clinica per singola malattia o mutazione ma anche ai registri di sanità pubblica “population based” con base regionale o nazionale per confrontare i propri numeri con quelli dei registri di patologia specifica. Rendendo disponibile questa informazione il catalogo di RD-Connect intende facilitare la condivisione di informazioni più dettagliate che i responsabili dei registri potranno caricare direttamente sulla piattaforma o condividere privatamente con i ricercatori che ne faranno richiesta.
- D) È stato costituito un Core Implementation Group di 18 registri e biobanche che hanno reso disponibili i materiali e documenti elaborati all'interno del proprio progetto e stanno

attivamente collaborando alla definizione di “use cases” per l’integrazione dei propri dati con quelli di altri registri e con biobanche di malattie rare all’interno della piattaforma.

- E) È stata condotta una rassegna accurata dei sistemi di ontologie disponibili per la descrizione fenotipica di diverse malattie rare, è stato proposto un sistema di integrazione delle ontologie stesse (in particolare la *Human Phenotype Ontology* e l’*ORPHANET Rare Disease ontology*) all’interno della piattaforma insieme ad un elenco di elementi di dati comuni (CDE) pensati per favorire l’integrazione tra registri e progetti dedicati a patologie differenti.
- F) È stata condotta una rassegna esaustiva dei sistemi esistenti per l’identificazione unica delle persone con malattia rara da adottare a livello globale. L’adozione di un unico codice identificativo per paziente permetterà di collegare tutti i dati clinici, di Omics e di disponibilità di biomateriali appartenenti alla stessa persona con malattia rara a livello globale, evitando le doppie registrazioni e la perdita di informazioni da un database all’altro. La rassegna ha tenuto in conto il problema fondamentale della sicurezza e del rischio di reidentificazione del paziente, che deve essere reso minimo e gestito attraverso procedure rigorose di sicurezza.
- G) È stato deciso di adottare la *Human Phenotype Ontology* (HPO) come principale strumento di descrizione fenotipica, in particolare per i casi senza una diagnosi genetica. In particolare è stato proposto di adottare il software PhenoTips, basato sull’HPO e sviluppato dall’Università di Toronto e di coinvolgere nel modo più attivo possibile i consorzi di ricercatori legati ai B-Projects nella definizione di termini sempre più accurati attraverso un flusso informativo il più possibile inclusivo e condiviso.
- H) Sono state condotte diverse iniziative per favorire l’avvio di nuovi registri e garantire l’interoperabilità di quelli esistenti.

Ad oggi Il CNMR all’interno del WP2 ha lavorato in stretta sinergia con il WP3 dedicato alle biobanche e coordinato da Telethon, in particolare per lo sviluppo del catalogo online di registri e biobanche e più di recente per la definizione di meccanismi di integrazione tra registri e biobanche. Inoltre ha coordinato l’attività di redazione e condivisione con il gruppo di ricerca delle linee guida per il consenso informato all’interno del WP6.

Attività degli altri Work Package

Al secondo anno di attività i diversi work package di RD-Connect hanno tutti sviluppato buona parte dei “task” assegnati e iniziano ad integrare le proprie attività con quelle di altri workpages.

Biobanche

L’accesso a campioni di biomateriali di qualità è un prerequisito necessario in ogni ricerca di Omics. Sulla base di reti di biobanking già esistenti tra cui EuroBioBank e BBMRI, RD-Connect sta sviluppando un nuovo sistema di catalogazione online di disponibilità di biomateriali e campioni biologici di malattie rare tra cui biopsie, tessuti, DNA, sieri, RNA e linee indotte di cellule staminali pluripotenti. Per facilitare questo processo RD-Connect sta implementando un flusso di lavoro semplificato e ottimale per la raccolta, conservazione e diffusione dei biomateriali.

Strumenti di bioinformatica

La moltitudine dei progetti di Omics esistenti sta generando una quantità di dati senza precedenti con la necessità di strumenti di analisi sempre più sofisticati per estrapolare la conoscenza. RD-Connect sta sviluppando strumenti di annotazioni di varianti in databases, previsione di patogenicità delle varianti stesse, analisi integrative di biomarker e profili molecolari di identificazione, identificazione di target terapeutici, ontologie e nanopublications. L'utilizzo di strumenti web e API insieme alla creazione e condivisione di strumenti operativi come i flussi di lavoro e le procedure operative standard (SOP) rende possibile l'uso di questi strumenti all'interno della piattaforma RD-Connect e nei sistemi locali. Una grande attenzione all'interoperabilità farà in modo che la piattaforma RD-Connect possa interagire con i sistemi futuri e gli strumenti sviluppati nei progetti collegati ad IRDiRC.

Piattaforma integrata

Il *Data Coordination Center* (DCC) consiste in un sistema per la memorizzazione, gestione e distribuzione dei dati. Questo include metodi per l'analisi primaria dei dati quali l'allineamento delle sequenze alle rispettive sequenze del genoma di riferimento e l'eventuale segnalazione della presenza di varianti. In questo primo livello saranno inclusi strumenti e standard per l'analisi non solo di dati provenienti dal sequenziamento genomico ma anche da esperimenti di proteomica, metabolomica e dati biochimici e fenotipici. I risultati uniformemente processati saranno successivamente messi a disposizione dal *Data Analysis Center* (DAC) per eventuali ulteriori analisi. Il DAC sarà anche potenziato per eventuali sviluppi di sequenziamenti genomici su larga scala. Il sistema terrà quindi conto dei numerosi dataset sviluppati in collaborazione con WP2 e WP3 con particolare attenzione alle esigenze delle malattie rare, e delle grandi moli di dati da analizzare che richiederanno un'elevata capacità di calcolo. Il sistema sarà sviluppato in stretta collaborazione con i B-projects per soddisfare in primis le loro esigenze e quelle dei futuri progetti IRDiRC.

Questioni etiche, legali e sociali

La ricerca Omics e, più in generale, la condivisione di dati e risultati di ricerca sollevano diverse problematiche etiche, legali e sociali. Tuttavia lo sviluppo di strumenti diagnostici e terapie innovative richiede di trovare soluzioni pragmatiche per non ostacolare la ricerca, pur tutelando gli interessi delle persone con malattia rara. Nell'affrontare queste problematiche RD-Connect tiene conto delle esperienze maturate in altri progetti su larga scala per sviluppare standard e linee guida il consenso informato e la condivisione dei dati che permettano alla ricerca sulle malattie rare di andare avanti.

Bibliografia

- ORPHANET (2015): Rare Disease Registries in Europe - January 2015. Orphanet Report Series. Disponibile all'indirizzo: <http://www.orpha.net/orphacom/cahiers/docs/GB/Registries.pdf>. Ultimo accesso: 1 febbraio 2015.
- Robinson PN. Deep phenotyping for precision medicine. Hum Mutat. 2012 May;33(5):777-80. doi: 10.1002/humu.22080.