



Ministero della Salute – Direzione Generale della Ricerca e dell'Innovazione in Sanità

Fondi 5 per mille ANNO 2018
Abstract ed elenco pubblicazioni scientifiche

Ente della Ricerca Sanitaria

Denominazione Ente:

Associazione La Nostra Famiglia – IRCCS “Eugenio Medea”

Codice fiscale: 00307430132

Sede legale: Via Don Luigi Monza n. 1 – Ponte Lambro (Co)

Indirizzo di posta elettronica dell'ente: segreteria.scientifica@pec.emedeia.it

Dati del rappresentante legale: Luisa Minoli nata il 14.01.1968 a Busto Arsizio (Va)
– CF: MNLLSU68A54B300V

Titolo del progetto: Il ruolo della diversità genetica in EBV nella suscettibilità alla Sclerosi Multipla

Abstract dei risultati ottenuti:

La sclerosi multipla (SM) è una malattia neurodegenerativa cronica ad eziologia sconosciuta. Le varianti di rischio genetico spiegano cumulativamente meno del 50% dell'ereditabilità della SM, mentre il fattore di rischio ambientale più forte è il virus di Epstein-Barr (EBV).

EBV è uno dei nove herpesvirus che infettano l'uomo ed è anche uno dei virus più diffusi nella nostra specie. Come tutti gli herpesvirus, EBV ha un genoma di grandi dimensioni e mostra una considerevole diversità genetica. Esistono due sottotipi di EBV (tipo 1 e tipo 2) che differiscono per la loro distribuzione e per la loro capacità di immortalizzare le cellule B. Inoltre, alcuni tumori associati ad EBV mostrano anch'essi incidenze notevolmente diverse in tutto il mondo. Diversi studi hanno, infatti, dimostrato che varianti nei genomi di EBV modulano il rischio e la predisposizione alla malattia. Ad esempio, il sequenziamento di genomi di EBV ha mostrato che alcune varianti virali spiegano una grande proporzione del rischio di sviluppare un carcinoma nasofaringeo. Tuttavia, per quanto riguarda i dati su varianti di rischio in EBV per malattie autoimmuni, in particolare SM, non si hanno molte informazioni. In generale, la variabilità genetica dei ceppi virali di EBV circolanti in Italia è praticamente sconosciuta.

L'obiettivo di questo progetto era quello di applicare strumenti di evoluzione molecolare per studiare l'evoluzione e l'adattamento alla nostra specie di un virus estremamente diffuso e identificare quei meccanismi molecolari che in determinate circostanze rendono questo virus responsabile di insorgenza di malattia.

Come prima analisi, abbiamo guardato la diversità genetica del virus in relazione alla sua struttura geografica, sfruttando le sequenze virali presenti in database pubblici.

Nel dettaglio, abbiamo utilizzato più di 300 genomi completi che avessero anche informazioni sul luogo e la data di campionamento. La generazione di un network filogenetico

ha indicato che i due gruppi principali (genotipo 1 e 2) si separano in maniera evidente e che la diversità genetica è chiaramente strutturata a livello geografico, con campioni dello stesso continente/area che si raggruppano insieme, formando dei cluster distinti rispetto alle altre aree geografiche.

Abbiamo quindi indagato in maniera più approfondita la struttura della popolazione del virus. Abbiamo utilizzato il programma STRUCTURE: questo tool si basa su un modello statistico bayesiano per raggruppare i genotipi in popolazioni senza fornire informazioni precedenti sulla loro relazione. Il programma è in grado di identificare sottopopolazioni distinte che compongono la popolazione globale. Le sottopopolazioni possono quindi essere correlate a caratteristiche specifiche come origine, genotipo o anche in base al fenotipo. Sebbene la maggior parte dei genomi abbia livelli variabili di ibridazione, i cluster di popolazione inferiti con STRUCTURE hanno mostrato una buona corrispondenza con l'origine geografica.

Alcune componenti genomiche hanno mostrato una preponderanza marcata a seconda del luogo di campionamento; ad esempio è stata trovata una componente genetica specifica per i ceppi africani, europei e americani, inoltre i genomi asiatici derivati dai casi di carcinoma nasofaringeo erano distinguibili dagli altri ceppi asiatici. Questo tipo di analisi ha permesso quindi di avere una più chiara rappresentazione delle relazioni filogenetiche che esistono all'interno di questa specie virale, suggerendo anche come sia probabile che le differenze geografiche a livello di prevalenza della varie malattie associate a EBV possano essere spiegate dalla variabilità genetica virale.

Successivamente è stata valutata la storia evolutiva del virus, in particolare abbiamo applicato un approccio filogenetico e di genetica di popolazione per indagare quali eventi selettivi abbiano accompagnato la divergenza di questo virus. Per prima cosa, abbiamo selezionato sequenze da database pubblici al fine di avere un pool equilibrato di ceppi virali. Per ogni genoma, abbiamo recuperato le sequenze codificanti annotate e stabilito il livello di ortologia tra i vari ceppi.

Per stimare la variazione a livello di singola ORF, abbiamo inferito il rapporto tra mutazioni non sinonime e sinonime (dN/dS). Il rapporto dN/dS per codone è indicativo del tipo di selezione che agisce sul singolo gene. Un eccesso di mutazioni non sinonime ($dN/dS > 1$) viene interpretato come presenza di selezione positiva, mentre un eccesso di mutazioni sinonime ($dN/dS < 1$) viene interpretato come selezione purificatrice. Un equilibrio tra mutazioni non sinonime e sinonime ($dN/dS = 1$) è inteso come selezione neutra.

Abbiamo quindi applicato questo tipo di analisi a circa 100 geni di EBV, suddivisi in differenti categorie funzionali: da geni coinvolti nella replicazione del DNA a geni responsabili nel modulare la risposta immunitaria dell'ospite. Abbiamo identificato 16 geni che presentano almeno un codone selezionato positivamente (con $dN/dS > 1$). La caratterizzazione funzionale di questi geni non ha rivelato una sovrarappresentazione di una classe funzionale, infatti questi geni si distribuivano in maniera uniforme nelle varie classi funzionali.

Abbiamo poi analizzato la distribuzione delle regioni disordinate nel genoma di EBV. Le regioni disordinate sono quei domini proteici che non adottano strutture tridimensionali compatte e che si ritiene siano fortemente arricchiti di segnali di selezione positiva. Utilizzando differenti software di predizione, abbiamo valutato il livello di disordine in tutte le proteine virali e abbiamo cercato una correlazione tra tale disordine e i siti identificati come selezionati positivamente. Non è stata riscontrata nessuna correlazione tra i livelli di disordine e la selezione positiva, suggerendo che questi segnali evolutivi non dipendono dalle caratteristiche intrinseche delle proteine analizzate.

In conclusione, abbiamo perfezionato e applicato approcci bioinformatici che consentono di analizzare l'evoluzione, la distribuzione geografica e l'origine di uno dei patogeni umani più diffusi e di grande rilevanza dal punto di vista della salute pubblica.

Infine, nel corso di questo progetto ci si è focalizzati anche nel sequenziamento di EBV con particolare attenzione al suo ruolo nella predisposizione a sviluppare Sclerosi Multipla. E' stato messo a punto un protocollo di sequenziamento NGS basato sull'analisi del genoma di EBV e sulla possibilità di poter distinguere i due sottotipi virali che caratterizzano questo patogeno. Nel dettaglio, si è scelto di adottare un sistema ad amplificazione del materiale genetico con ampliconi di circa 300 nucleotidi l'uno (Illumina AmpliSeq™), per un totale di 680 ampliconi per genoma, così da poter caratterizzare almeno l'80% del totale del genoma virale. Una prima analisi dei genomi virali ha confermato come la presenza di lunghe regioni di sequenze ripetute abbia un effetto importante sul sequenziamento di questa specie virale, rendendo complessa la loro analisi. Sono state quindi eseguite diverse prove di sequenziamento per mettere a punto il protocollo di sequenziamento utilizzando linee cellulari linfoblastoidi (che di routine vengono immortalizzate mediante infezione con EBV). I risultati ottenuti hanno permesso di predisporre un protocollo sperimentale in grado di ottenere le informazioni genomiche per il 75% del genoma con una profondità di copertura media di 350x. Nel frattempo, sono state anche testati differenti protocolli di estrazione del DNA virale utilizzando differenti materiali di partenza (e.g. sangue intero, plasma, saliva) allo scopo di verificare quale fosse la condizione migliore per poter ottenere la massima quantità di DNA virale possibile e contemporaneamente ridurre al minimo la contaminazione dovuta al DNA dell'ospite. I primi risultati di questo tipo di analisi hanno evidenziato come l'estrazione da plasma sia la strategia migliore da seguire.

Infine, è stata predisposta una pipeline bioinformatica per l'analisi dei dati grezzi generati dal sequenziamento. Anche in questo caso sono stati valutati differenti software e tool bioinformatici che permettessero di ottenere dei risultati più affidabili possibile. In particolare, si sono predisposte delle pipeline che permettessero di ricostruire la sequenza consenso del genoma virale, identificare le specifiche mutazioni associate ad un singolo strain e identificare anche la possibile presenza di infezioni miste. La predisposizione di tutti questi protocolli, sia sperimentali che computazionali, sarà di fondamentale importanza per futuri progetti di ricerca su questo virus, dove si effettueranno analisi su una numerosità campionaria elevata e che, grazie a questo progetto, non necessiteranno di messa a punto.

Prodotti della Ricerca (correlati al progetto):

I risultati delle analisi evolutive e sulla variabilità genetica del virus saranno utilizzati per la scrittura di articoli scientifici su riviste indicizzate e ad alto impatto.

I risultati ottenuti durante il sequenziamento virale hanno permesso di predisporre un protocollo sperimentale e una pipeline bioinformatica per l'analisi dei dati che verranno applicati in futuri progetti di ricerca.

Data , 23/01/2023

Il Responsabile del Progetto

Dr. Diego Forni



Il Legale Rappresentante

Dr.ssa Luisa Minoli



Si autorizza al trattamento dei dati ai sensi del d.lgs. 196/2003

Il Legale Rappresentante

Dr.ssa Luisa Minoli

