

Università degli Studi di Catanzaro "Magna Graecia"



Facoltà di Medicina e Chirurgia

*Scuola di Specializzazione in Medicina
Interna
"P.L. Mattioli"*

***Tecniche Bioinformatiche, Post-Genomiche
e Dislipidemie***

Gruppo Esami V anno

Dr. Vittorio Emanuele

Matricola 68292

Introduzione

Con la conclusione del progetto genoma la comunità scientifica internazionale ha a disposizione da pochi mesi il codice genetico dell'uomo¹. Questo ha prodotto l'archiviazione di molti Gbit di dati di diverso tipo. Pur trattandosi di un problema prettamente biologico (il sequenziamento e soprattutto l'interpretazione dei dati), infatti, la più importante sfida tecnica che gli scienziati si sono trovati a dover affrontare è stato di tipo informatico. Il problema era stato evidenziato già dalle prime sequenze nucleotidiche, ma la necessità di risolverlo si è presentata solo con l'accumularsi dei dati, cui si è aggiunta l'eterogeneità delle informazioni e la complessità delle loro relazioni. L'*IT* (*Information Technology*) ha ricevuto grazie a queste necessità un grandissimo impulso negli ultimi dieci anni. Diversi tipi di soluzione, che non sono oggetto di questa trattazione, hanno consentito, comunque, di ottenere dei database di tipo relazionale, veloci, facili da consultare e soprattutto molto ben correlati tra loro. Questi sistemi informativi consentono a chiunque (la maggior parte, infatti, sono liberamente accedibili) di cercare sequenze nucleotidiche e proteiche, di capire se si ci trova in un tratto di genoma codificante, promotore o altro; contengono, insomma, un'immensa mole di informazioni il cui fine ultimo dovrebbe essere quello di essere messo in relazione con le caratteristiche fenotipiche². In altre parole, la sequenza nucleotidica in sé è un'informazione di tipo linguistico – matematico (sequenza di quattro tipi diversi di caratteri), che

però contiene un numero enorme di altri dati (per esempio geni e loro regolazione espressiva) variamente codificati (per esempio codice genetico). Queste informazioni sono la base per la produzione di proteine e, di conseguenza, di cellule, tessuti e organi. La sfida della medicina moderna è quella di mettere in relazione i dati genetici con quelli clinici allo scopo di prevenire, diagnosticare ed infine curare le malattie³. L'approccio al problema è complesso per diverse ragioni. In primo luogo, l'analisi di questi dati richiede un'integrazione di conoscenze (matematiche, informatiche, biologiche e mediche) che non trova riscontro nelle figure professionali formate dalle nostre università. In secondo luogo, ancora oggi, le tecniche mediante cui questi studi si svolgono sono poche e in continua evoluzione. In questo mare di difficoltà, tuttavia, esistono dei punti fermi, metodiche e tecniche, che se ben condotte danno enormi soddisfazioni. Quest'ultime sono facili da ottenere quando ci si trova di fronte a malattie monogeniche, in cui l'assioma "un gene una malattia" è valido. In questi casi (per esempio Ipercolesterolemia Familiare) un'alterazione nota in un gene (recettore delle LDL) produce un fenotipo molto preciso (aumento plasmatico delle LDL), che a sua volta causa clinica e patologia facilmente identificabili (infarto del miocardio in età giovanile). Rari sono, però, i casi in cui l'assioma di cui sopra si può applicare, dal momento che, per lo più, si tratta di malattie plurifattoriali e multigeniche, in cui il fenotipo (per esempio infarto del miocardio) è la risultante di un insieme di alterazioni genetiche (diversi polimorfismi) ed un insieme di alterazioni

ambientali (fumo) e/o malattie concomitanti (diabete). La presenza di una sola delle precedenti alterazioni in un singolo individuo non è sufficiente a produrre il fenotipo patologico. In questa sede cercheremo di capire se e in che modo la bioinformatica può aiutare il clinico nella sua attività, quali sono i suoi vantaggi e limiti, quali i suoi campi di applicazione e i possibili benefici per l'utente finale, il paziente dislipidemico⁴.

Perché le Dislipidemie

L'interesse per le dislipidemie non è casuale. Nei Paesi occidentali, infatti, il maggior numero di morti è dovuto a eventi morbosi legati all'arteriosclerosi. Quest'ultima riconosce diversi fattori di rischio, alcuni non modificabili (il sesso, l'età), alcuni modificabili (ipertensione, ipercolesterolemica *etc.*) altri ambientali (fumo, dieta). Uno dei più importanti fattori di rischio per l'arteriosclerosi, nonché momento patogenetico iniziale, risulta essere l'ipercolesterolemia e più esattamente l'aumento del colesterolo LDL. A sua volta, quest'ultimo può essere espressione di una miriade di patologie (familiari e non familiari) genericamente inquadrate come dislipidemie. Risulta evidente, quindi, come vi sia notevole attenzione da parte della comunità scientifica nel tentare di diagnosticare e curare queste malattie. Molti di questi difetti non hanno cause genetiche note e risentono fortemente dell'ambiente. A questo punto entrano in gioco le tecniche bioinformatiche⁵. Diversi scienziati ritengono che l'applicazione di queste tecniche possa essere di estrema utilità per il progredire delle conoscenze in oggetto. Ad oggi, però, pochissimi se ne occupano o hanno dei progetti di ricerca in tal senso.

Sul piano teorico una delle prospettive più allettanti sarebbe quella di utilizzare le informazioni provenienti dal progetto genoma, per potere identificare quali sono i geni coinvolti nelle alterazioni del metabolismo lipidico. Le tecniche classiche hanno come presupposto quello di correlare l'espressione

del fenotipo con la presenza di *markers* sul genoma mediante studi di linkage. Questi studi hanno permesso l'identificazione di alcune dislipidemie (per esempio Ipercolesterolemia Autosomica Recessiva⁶). Sono, in genere, molto utili quando il numero di geni coinvolto nella patologia è limitato. Risultano inutili e, peggio ancora, a volte fuorvianti, invece, nella patologia in cui i geni coinvolti sono molti o in cui la loro stessa espressione è fortemente regolata da interazioni gene-ambiente (per esempio iperlipidemia familiare combinata). Saranno, con molta probabilità, quest'ultime patologie a giovare delle tecniche di analisi bioinformatiche⁷.

L'acquisizione dei dati

Risulta evidente che il primo passo per qualsivoglia analisi mediante supporti informatici è la raccolta dei dati. Esistono diverse tecniche per la raccolta dei dati che non saranno oggetto di questa trattazione². E' importante, però, dire che esistono dei canoni universali ed indipendenti dall'infrastruttura informatica utilizzata. La caratteristica più importante è quella dell'universalità. Una banca dati è molto più utile se i dati in essa contenuti possono essere condivisi e/o mischiati con altre banche dati. Alla luce di questo la loro quantità passa in secondo ordine. Un'altra caratteristica importante dei database è la loro relazionalità. I dati raccolti devono essere correlati tra loro e la natura della relazione deve essere studiata in maniera tale da poter associare elementi diversi di uno stesso record nel minor tempo possibile. Quello che deve essere sempre conservato, quindi, è il flusso logico-informativo. Questo inizia con l'inserimento dei dati o *Input*, il processamento, la memorizzazione, la distribuzione che avviene mediante l'ultima fase del processo informativo, cioè la trasmissione. L'implementazione di questi processi nel loro insieme ha permesso alla bioinformatica moderna la strutturazione di sistemi informativi eccezionali, senza i quali oggi nessun medico e nessun ricercatore potrebbero mai lavorare o studiare. Pensiamo, per esempio, alla differenza che esiste fra un laureando di oggi e uno di soli 10-15 anni fa. Oggi un laureando, per effettuare una ricerca bibliografica in merito ad un argomento, impiega un tempo

compreso fra 0,1 e 5 secondi, necessario per la risoluzione della query (domanda che viene effettuata al database e che si basa su operatori logici universali booleani) da parte del motore di ricerca del database e per l'invio delle informazioni al computer terminale da cui l'interrogazione viene effettuata (per esempio internet). Fino a 10-15 anni fa, invece, la ricerca veniva effettuata manualmente sugli *Index Medicus* o pubblicazioni simili e poteva, quindi, impiegare anche giorni. Questo esempio è emblematico: lo stesso livello di sviluppo (e in alcuni casi anche maggiore) ha riguardato tutti i campi della ricerca scientifica, con ripercussioni sul piano della produttività che è inutile sottolineare.

Processo fondamentale, subito successivo all'inserimento, come abbiamo visto nel flusso informativo, è la memorizzazione. E' impressionante la velocità con cui cresce la capienza dei dispositivi per la memorizzazione di massa e anche degli algoritmi informativi che permettono alla macchina di funzionare. Tale velocità correla perfettamente con quella di crescita della capacità di calcolo dei processori, descritta nella famosa legge di Moore (il numero di transistor in un *microchip* raddoppia ogni 18 mesi).

In questo momento nella comunità scientifica internazionale esistono pochissimi esempi (pubblici) di questo tipo (database delle mutazioni per il gene del recettore per le LDL, gestito dalla *College University* di Londra, e sistema MEDPED con scopo simile al precedente, ma in fase di costituzione presso la *University of Utah*, USA). Questi esempi, comunque, non sono orientati allo

studio clinico, ma solo ad una funzione di supporto per gli studi di biologia molecolare (per esempio raccolta di *primer* utilizzati per la ricerca delle mutazioni). Per potere essere utile anche sul piano clinico (per l'assistenza e l'epidemiologia), un database di questo tipo dovrebbe contenere informazioni di correlazione genotipo-fenotipo al variare del tipo di popolazione. Sicuramente è utile collezionare tutte le mutazioni di un recettore che producono una malattia, per sapere in che modo una mutazione può essere evidenziata nei pazienti, ma è sicuramente più utile, per esempio, conoscere in che modo queste mutazioni si distribuiscono nella popolazione dal punto di vista geografico. In questo modo un paziente con diagnosi fenotipica di FH (Ipercolesterolemia Familiare), che proviene da una regione in cui il 90% delle mutazioni sono del tipo G528D (per esempio la zona di Palermo, in Sicilia), permetterà uno screening molecolare rapido e nella maggior parte dei casi poco costoso. Affinché questo possa avvenire, però, è necessario che vi sia la possibilità di raccogliere, strutturare, conservare e rendere accessibili informazioni di questo tipo (oggi in Italia non esiste un sistema di questo tipo che, invece, è ben sviluppato in Inghilterra e soprattutto in Olanda, paese pioniere in questo campo). La necessità di sistemi simili è quanto mai indispensabile soprattutto nel centro-sud della nostra penisola, in particolar modo in Calabria e Sardegna. In queste due regioni, infatti, si verificano condizioni rarissime sul pianeta, che consentono di classificare queste zone geografiche come "isole genetiche". In altre parole, la bassissima immigrazione, abbinata, in queste zone, con l'elevata emigrazione e

l'elevata prevalenza e incidenza di matrimoni fra consanguinei (anche cugini di I grado), fa sì che la mescolanza genetica sia molto bassa e che, di conseguenza, la prevalenza di malattie genetiche sia elevata. Questo effetto, di recente, ha consentito di utilizzare dati in gran parte provenienti proprio dalla Sardegna per l'identificazione di un nuovo tipo di iperlipidemia familiare (Ipercolesterolemia Familiare Recessiva⁶), che, per il tipo di trasmissione, difficilmente si sarebbe potuta identificare in zone con mescolanza genetica più elevata. Ciò nonostante, però, ad oggi siamo in grado di stimare solo approssimativamente la prevalenza delle diverse dislipidemie in Sardegna, mentre nulla o poco si sa sulle popolazioni calabresi. In queste regioni, per complicare e abbellire il quadro dal punto di vista scientifico, è presente una bassa incidenza di malattie cardiovascolari rispetto ai paesi del nord Europa (per esempio rispetto all'Olanda, dove pure vi è una discreta mappatura delle dislipidemie almeno per quanto riguarda le ipercolesterolemie⁸), per esempio, a significare che non esiste la mutazione di un recettore (LDL) ma il paziente in cui questa mutazione si verifica. Proprio il principale gruppo di ricerca olandese in merito ha recentemente pubblicato dati di questo tipo, in cui si dimostra come, a parità di mutazioni, la mortalità anche per le sole malattie cardiovascolari varia al variare delle famiglie e come, anche in presenza di gravi forme di ipercolesterolemia, alcune famiglie esprimono una longevità inaspettata. Risulta evidente, quindi, l'utilità che un sistema organico di raccolta di questi dati clinici e biochimici potrebbe avere. Informazioni sul fenotipo lipidico espresso a livello biochimico

(ipercolesterolemia isolata piuttosto che mista con ipertrigliceridemia), sul fenotipo clinico (eventi cardiovascolari, xantomi, aterosclerosi carotidea, *etc.*), sul tipo di familiarità se presente sulle caratteristiche genetiche (mutazioni, satelliti ed altri marcatori fisici genetici), sul tipo di risposta alla terapia (sensibilità al trattamento, *non responders*, farmacogenetica, *etc.*) sono indispensabili per disegnare un quadro d'insieme su queste patologie in una data popolazione. Ottenere un'informazione di questo tipo è una cosa estremamente importante, perché questi dati possono servire a tantissimi scopi, dalla stima di prevalenza e incidenza di malattie, alle scelte farmaco-economiche ed epidemiologiche di assistenza sanitaria.

Analisi dei dati: il valore della genomica in medicina clinica

Sebbene sia stata da poco completata la sequenza dell'intero genoma umano, questo non è che il primo passo sulla strada di una “medicina genomica”, in cui le informazioni di sequenza possono essere utilizzate per la diagnosi e terapia delle malattie⁹. Le annotazioni sulle sequenze sono, sotto questo punto di vista, fondamentali. E' noto, infatti, che gli individui di una stessa specie condividono il 99,9% del materiale genetico. Parenti e discendenti hanno, ovviamente, affinità superiori. La variabilità genetica, quindi, è affidata ad una piccolissima parte del genoma, lo 0,01% appunto. Questa quota è comunque sufficiente ad offrire un'enorme biodiversità intraspecie. Ad oggi, tra l'altro, sono state identificate decine di migliaia di queste variazioni. Grossolanamente e senza avere la pretesa di descriverle nei dettagli possiamo dividerle, in primo luogo, in due grosse famiglie. Vi sono le variazioni puntiformi e quelle non puntiformi. Le variazioni puntiformi, a seconda della frequenza con la quale si trovano nel genoma, si differenziano in polimorfismi e mutazioni. I primi, per essere tali, devono essere presenti nell'1% della popolazione; le seconde, invece, sono del tutto rare. Le mutazioni sono generalmente associate ad una malattia e rispondono ad una legge che in genetica è nota come “ un gene una malattia”. In questi casi la proteina per la quale il gene codifica ha una funzione estremamente importante sul piano fisiologico, per cui l'evoluzione tende a conservarla (per esempio il recettore per le LDL, la cui mutazione è responsabile

dell'Ipercolesterolemia Familiare¹⁰). I polimorfismi, invece, generalmente appartengono a geni che codificano per proteine la cui importanza, sul piano clinico, è collegata all'espressione di altri geni e di determinate condizioni ambientali (è il caso, per esempio, del gene per IRS-1, proteina fondamentale per la trasduzione del segnale insulinico e il cui polimorfismo 972 predispone all'insulino-resistenza e, quindi, al diabete). Le mutazioni non puntiformi (traslocazioni, inversioni, duplicazioni), invece, se pur importanti sul piano patologico, svolgono un ruolo secondario in questo ambito clinico. Le moderne tecniche di identificazione di queste variazioni genetiche hanno permesso il fiorire, in particolare negli ultimi dieci anni, di studi di associazione fenotipo-genotipo. Questi studi, però, sono sempre particolarmente contraddittori. Un polimorfismo, infatti, essendo per definizione una variazione normale e frequente del genoma, è facilmente associabile ad un fenotipo sul piano statistico quando il campione è piccolo; di contro, l'associazione è difficilmente rilevabile al crescere del campione in esame. Il discorso è diverso nel caso delle mutazioni responsabili direttamente di una malattia, ma questi casi sono purtroppo l'eccezione, essendo la maggior parte delle malattie il risultato dell'espressione di diversi geni, che a loro volta sono in relazione con diversi fattori in diversi ambienti (per esempio il diabete o l'obesità). Gli studi di polimorfismi, quindi, dopo un acme raggiunto alla fine degli anni Novanta, attraversano recentemente una piccola crisi. Contestualmente sono in enorme espansione le annotazioni funzionali genomiche. Conoscere la sequenza

nucleotidica, come abbiamo già detto, ha poca importanza se non si conosce la sua funzione. I ricercatori che hanno completato il progetto genoma ora stanno arricchendo queste informazioni, proprio cercando di associare una funzione per ognuno dei circa 20-30.000 geni presenti nel genoma. Mediante studi classici di linkage, infatti, è stato possibile scoprire la funzione di poche migliaia di geni, la cui alterazione è associata ad una malattia (esiste una banca dati mondiale gestita dal *National Institute of Health*, nota come OMIN). Gli studi di linkage, infatti, si basano sul concetto che individui correlati sul piano genetico (parenti) tendono a condividere grosse parti di genoma. Una quota compresa intorno ai 500 marcatori sparsi lungo tutto il genoma, quindi, è più che sufficiente per effettuare studi di questo tipo, sempre che sia un solo gene a provocare una malattia o, per lo meno, che vi sia un gene principale. Nella maggioranza dei casi, però, e anche nei casi di malattie importanti sul piano sanitario (per esempio il diabete), sono più geni che interagiscono fra loro e con l'ambiente a provocare le malattie. Per quest'ultimi casi, solo da poco e grazie al completamento del progetto genoma i ricercatori hanno disegnato un approccio di ricerca noto come *whole-genome association studies* o studi di associazione di tutto il genoma, che promette di essere molto utile sotto questo punto di vista. I vantaggi consistono nel poter utilizzare soggetti non correlati (cioè non familiari), nel poter reclutare molte persone e nel poter studiare più geni, che sono in grado di dare anche un lieve aumento di rischio per una determinata patologia. Lo svantaggio principale consiste nel fatto che è richiesta una densità

di polimorfismi di molti ordini di grandezza superiore a quella che si usa per il *linkage* (in genere decine di migliaia). A breve saranno, infatti, disponibili le prime mappe di associazione genomica, in cui sarà presente una densità sufficiente di polimorfismi per condurre gli studi. A questo punto entra in campo un altro aspetto fondamentale: l'informatica. È impensabile, infatti, analizzare decine di migliaia di polimorfismi con i tradizionali software, che si utilizzano oggi per gli studi di *linkage*. Non bisogna trascurare, inoltre, le altre innumerevoli applicazioni (per esempio le analisi dei dati provenienti dai *Micro Array*) in cui l'informatica è dominante. Bisogna, a tal proposito, inventare tutto, dalle interfacce per l'inserimento dei dati agli algoritmi per l'analisi degli stessi. In questo lavoro gli informatici sono fortemente indietro, soprattutto perché non in possesso di nozioni di biologia, medicina, clinica e genetica. Nasce, quindi, l'esigenza, di creare una nuova classe di scienziati, che siano in grado di fondere, o per lo meno di coordinare, i due tipi di conoscenze. Con un *insight* di *Nature*, pubblicato nel novembre del 2002 e intitolato “*Computational Biology*”¹¹, la comunità scientifica internazionale ha preso coscienza di questo problema e negli ultimi anni sono nati una nuova disciplina, nota come Bioinformatica, una nuova classe di scienziati, nuovi corsi di laurea e di dottorati di ricerca. La post-genomica, dunque, è e sarà di importanza vitale per l'analisi dei dati, insieme alla bioinformatica. Come accade per gli esami strumentali, gli interventi chirurgici o i semplici esami ematochimici, però, c'è bisogno primariamente di un'indicazione. Esami fatti senza indicazione, è noto,

hanno uno scarsissimo significato sul piano clinico e diagnostico, proprio come un intervento chirurgico, se mal indicato, può non servire a niente o essere addirittura dannoso. E' necessario, allora, vedere quali sono le domande a cui la bioinformatica in futuro potrebbe essere in grado di fornire risposte adeguate.

Aspettative delle analisi bioinformatiche applicate alle dislipidemie

Partendo dal presupposto di essere riusciti a strutturare un sistema informativo in cui sia possibile inserire in maniera sistematica tutti i dati fenotipici (eventi e caratteristiche cliniche dei pazienti) e genotipici (polimorfismi, mutazioni) relativi alle diverse dislipidemie presenti in una popolazione, sarebbe enorme la quantità di informazioni da acquisire, allorquando la banca dati avesse raggiunto la grandezza e l'accuratezza ideali. In primo luogo sarebbe possibile, come già sottolineato, stimare in maniera accurata la prevalenza e l'incidenza di alcune patologie di cui attualmente si sa poco (per esempio ad oggi non è nota con esattezza la prevalenza dell'Ipercolesterolemia Familiare Recessiva, delle Sitosterolemie e anche di altre dislipidemie più frequenti, ma non per questo meglio caratterizzate, come l'Iperlipidemia Familiare Combinata). Sarebbe possibile anche associare un profilo di rischio ad ogni malattia. E' noto che l'Ipercolesterolemia Familiare, per esempio, aumenta l'incidenza di malattie cardiovascolari in giovane età, ma diversi studi hanno già dimostrato come, indipendentemente dal difetto genetico (stessa mutazione), dalle interazioni ambientali (stessa popolazione), dal livello di assistenza sanitaria (stesso stato), dai fattori di rischio associati (ipertensione, fumo, età, sesso) e persino dal fenotipo espresso (stesso livello di colesterolemia), diversissima sia la risposta individuale alla malattia (per esempio infarto a 30 piuttosto che a 60 anni). In che modo e soprattutto da cosa

il difetto genetico preso in considerazione (la mutazione del recettore LDL), dunque, viene modulato per essere espresso sul piano clinico in maniera così eterogenea? L'aiuto di un sistema informativo, come quello ideale di cui stiamo parlando, potrebbe essere di estrema utilità per chiarire questi punti. Anche i dati di risposta alla terapia potrebbero essere analizzati e riutilizzati per cure personalizzate e/o individualizzate. In fondo, la farmacogenetica non è altro che l'utilizzo delle informazioni genetiche (e delle loro correlazioni ai fenotipi clinici) per lo studio della risposta ai farmaci. Non ci sarebbe da stupirsi se si scoprisse che particolari popolazioni rispondono meglio a certi tipi di statine piuttosto che ad altre. Da sempre, infatti, l'evoluzione seleziona attività enzimatiche e strutture cellulari che sono "utili" proprio sul piano evolutivistico (per esempio la resistenza alla malaria dei portatori dell'anemia falciforme). Quest'ultimo aspetto potrebbe, tra l'altro, essere di estrema utilità per il disegno di nuove molecole e, quindi, per la ricerca farmacologica. Un altro risvolto che questo tipo di studio potrebbe implementare è quello delle procedure terapeutiche. L'analisi di dati, anche retrospettivi, sui pazienti inseriti nel sistema informativo potrebbe fornire informazioni, per esempio, su quando e se intervenire con una procedura di rivascolarizzazione miocardica per cutanea piuttosto che chirurgica. Sempre dall'analisi di questi dati si potrebbero creare dei *biomarkers* anche compositi. Sarebbero, questi, dei complessi fattori di rischio formati singolarmente da un misto di informazioni genetiche e fenotipiche. In altre parole, un fumatore con una determinata serie di

polimorfismi (che, per esempio, espongono più di altri al danno da fumo) potrebbe essere più a rischio di un altro fumatore che non presenta quell'*array* di polimorfismi. In questi due individui, esemplari, l'astensione dal fumo avrebbe un significato molto diverso, con un'importanza di gran lunga superiore nel gruppo con *array* polimorfico più a rischio. Quest'ultimo aspetto è tutt'altro che da sottovalutare, specie se visto nell'ottica della medicina genomica. Oggi, infatti, l'iter che il medico segue è quello della diagnosi della malattia seguita dalla terapia. Fra una decina d'anni, o forse meno, con l'avvento delle prime mappe di associazione e la disponibilità di sistemi informativi dedicati per patologie e/o popolazioni, sarà possibile, già alla nascita di un soggetto, studiare il suo profilo di rischio genetico. Questo individuo seguirà, durante la sua vita, una prevenzione mirata alle malattie per le quali risulta geneticamente predisposto; per esempio, deciderà di iniziare a fumare sapendo che per lui quella decisione sarà fatale con molte più probabilità che per altri. Nello studio genetico effettuato alla nascita del soggetto saranno inclusi studi di farmacogenetica, così da evitare di somministrare una cefalosporina piuttosto che un'aspirina ad un paziente che nasce allergico a questo farmaco. Per far ciò, però, bisogna aver studiato con sufficiente accuratezza che quel determinato marker genetico (o genetico-ambientale) è realmente associato con lo sviluppo di shock anafilattico da cefalosporina, per esempio. Una delle priorità di questi sistemi informativi, naturalmente, dovrà essere la riservatezza, dal momento che informazioni di questo tipo non possono che essere utilizzate solo ed

esclusivamente per migliorare la durata e la qualità della vita. Tornando alle dislipidemie, una banca dati ben strutturata, unita a studi di associazione genomica avanzati, potrebbe finalmente spiegare come mai esistono persone che, pur con valori di colesterolemia plasmatica molto elevati, arrivano in perfetta salute in tarda età. Potrebbe non essere solo l'assenza di altri fattori di rischio (ipertensione, fumo, *etc.*), infatti, ad allungare la vita di questi soggetti, dal momento che non mancano esempi di individui morti a 40 anni con un infarto del miocardio fatale in assenza di fattori di rischio cardiovascolari e/o patologie trombotiche. La possibilità di analizzare grosse moli di dati clinici e di poterli associare a marcatori genetici è, probabilmente, l'unico modo per svelare questi misteri ^{12,13}.

Esempio di sistema informativo: Il database del Day Hospital di Medicina

Presso il nostro reparto, da quando è stato istituito il servizio di DH, i dati dei pazienti vengono raccolti in un sistema informativo creato appositamente. In due anni di attività questo sistema, oltre che ai fini assistenziali, ci ha consentito di monitorare e promuovere la nostra attività di ricerca. Di seguito sono riportati i dati relativi a questi due anni di attività (Tab 1).

Tab.1

<i>Patologia</i>	<i>N.Famiglie</i>	<i>N.Pazienti</i>	<i>Note</i>
FCH	48	100	
FH	10	24 (2 OMO)	5 Famiglie genotipizzate
ARH	1	1	
IperTRI	3	5	
IperCHI	2	2	
DisBeta	2	2	
IpoAlfa	1	6	
IpoBeta	2	2	
Poligenica		405	

In totale, nel corso di questi due anni, sono giunti alla nostra osservazione per dislipidemie 878 pazienti. Di questi 543 (il 62%) hanno ricevuto diagnosi di iperlipidemia di tipo familiare, 335 avevano una dislipidemia secondaria o metabolica (Sindrome Metabolica).

Conclusioni

Le dislipidemie, per il loro fondamentale ruolo nello sviluppo e progressione dell'aterosclerosi che, come già sottolineato, da sola è responsabile di circa la metà delle morti ogni anno nei paesi occidentali e occidentalizzati, dovrebbero essere campo fondamentale della moderna ricerca, bioinformatica e non solo¹⁴. Troppo spesso, infatti, si assiste a campagne di sensibilizzazione per altre patologie (oncologiche, infettive, genetiche), che mietono, comunque, meno vittime dell'aterosclerosi, o alla creazione di associazioni e gruppi per malattie che, per quanto importanti per il fatto stesso di far soffrire qualcuno, non rivestono, comunque, alcun interesse sul piano della sanità pubblica. E' ovvio che un medico non può e non deve occuparsi di economia nella gestione di un paziente e/o di una patologia. Non si può, però, rimanere insensibili di fronte al fatto che la maggior parte delle risorse scientifiche di un Paese sia destinata a malattie che, nella migliore delle ipotesi, sono "soltanto" la seconda o la terza (dipende dalla fascia di età) causa di morte nella popolazione. E' senz'altro auspicabile, quindi, che vi sia una ricerca scientifica più "democratica", che preveda lo stanziamento di una maggiore quantità di fondi per il problema di salute pubblica più importante. Si può affermare, in conclusione, che in questo momento l'enorme mole di dati provenienti dal progetto genoma è simile ai geroglifici egizi prima che fosse scoperta la stele di Rosetta: un'immensa quantità di informazioni che ancora non siamo in grado di comprendere. La

bioinformatica¹⁵, nel prossimo futuro, avrà la funzione di decifrare questi dati¹⁶ e correlarli con quelli clinici¹⁷, la cui stessa raccolta e organizzazione razionale è sempre in ugual misura affidata alla bioinformatica¹⁸⁻¹⁹. Le scoperte così ottenute potranno tornare utili nella medicina di ogni giorno, quella a cui i pazienti di continuo si rivolgono per i loro piccoli e grandi problemi.

BIBLIOGRAFIA

1. Human Genomics and Medicina. Insight. Nature 2004;429:439.
2. SUSHMITA MITRA, TINKUACHARYA. Data Mining Multimedia, Soft Computing, and Bioinformatics. JOHN WILEY & SONS, INC. 2003.
3. Functional Genomics. Insight. Nature 2000;405:819.
4. Ilene Mizrahi. GenBank: The Nucleotide Sequence Database. NCBI Book 2004.
5. Cynthia Gibas. Developing Bioinformatics Computer Skills. O'Reilly Inc. 2001.
6. Christine Kim Garcia, Kenneth Wilund, Marcello Arca, Giovanni Zuliani,⁴ Renato Fellin,⁴ Mario Maioli, Sebastiano Calandra, Stefano Bertolini, Fausto Cossu, Nick Grishin, Robert Barnes, Jonathan C. Cohen, Helen H. Hobbs. Autosomal Recessive Hypercholesterolemia Caused by Mutations in a Putative LDL Receptor Adaptor Protein. SCIENCE 2001;292:1394.
7. James Tisdall. Beginning Perl for Bioinformatics. O'Reilly Inc. 2001.

8. Marina A.W. et al. Low-Density Lipoprotein Receptor Gene Mutations and Cardiovascular Risk in a Large Genetic Cascade Screening Population. *Circulation*. 2002;106:3031-3036.

9. Human Genome. Special. *Nature* 2001;409:813.

10. Dalya Marks et al. A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia. *Atherosclerosis* 2003;168:14.

11. Computational Biology. Insight. *Nature* 2002;420:205.

12. Dalya Marks et al. Cost effectiveness analysis of different approaches of screening for familial hypercholesterolaemia. *BMJ* 2002;324:1303.

13. Marks D. et al. Screening of hypercholesterolemia versus case finding for familial hypercholesterolemia: a systematic review and cost-effectiveness analysis. *HTA* 2000;4:29.

14. Proteomics. Insight. *Nature* 2003;422:191.

15. G. Valle et al. *Introduzione alla Bioinformatica*. Zanichelli. 2003.

16. Andreas D. Baxevanis, B.F. Francis Ouellette. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Second Edition. 2001 John Wiley & Sons, Inc.

17. A. Camussi et al. Metodi Statistici per la Spermitantazione Biologica. Zanichelli. 2002

18. Arthur M. Lesk. Introduction to Bioinformatics. OXFORD UNIVERSITY PRESS 2002.

19. D. Mount. Bioinformatics: sequence and genome analysis. CSHL Press. 2003.