

**COMITATO NAZIONALE PER LA BIOSICUREZZA
E LE BIOTECNOLOGIE**

Gruppo di lavoro Bioinformatica

**Linee guida per la definizione di una strategia per lo sviluppo
del settore
della bioinformatica in Italia
con particolare attenzione all'ambito biomedico**

Dicembre 2005

Componenti la Commissione

Maria Teresa ANNECCA	Ufficio del Garante per la protezione dei dati personali
Francesco BELTRAME	Dipartimento di Bioingegneria, Università degli Studi di Genova
Giovanni BELTRAMI	Chiron Vaccines
Giovanni BRIGANTI	Ministero degli Affari Esteri
Fabrizio GAGLIARDI	CERN
Giovanni GIOVANNINI	Osservatorio Tuttimedia, Media duemila
Luciano MILANESI	Istituto Tecnologie Biomediche, Consiglio Nazionale delle Ricerche
Giovanni PAOLELLA	CEINGE, Università degli Studi di Napoli "Federico II"
Anna Maria RAVIOLA	Marconi, Selenia Communications
Fabrizio RICCI	Istituto di Ricerca sulla Popolazione e le Politiche Sociali, Consiglio Nazionale delle Ricerche
Leonardo SANTI	Comitato Nazionale Biosicurezza e Biotecnologie
Lorenzo SILENGO	Dipartimento di Biologia Università degli Studi di Torino
Anna TRAMONTANO	Dipartimento di Scienze Biochimiche "Rossi Fanelli", Università degli Studi di Roma "La Sapienza"
Settimio VINTI	Dipartimento Innovazione Tecnologica, Presidenza del Consiglio dei Ministri

Ufficio Comitato Nazionale Biosicurezza e Biotecnologie

Mario BEBI
Silvana CAMILLERI
Alessandra CIFANI
Giulia NITTI

Sommario

Con i recenti progressi nell'ambito biomedico, è diventato essenziale assicurare un supporto adeguato alle ricerche nell'ambito della medicina e delle scienze della vita. Infatti una caratteristica dell'era post-genomica dipenderà da nuove modalità di trattare l'enorme mole di dati generata quotidianamente al fine di correlare le informazioni genotipiche con quelle fenotipiche e cliniche.

In questo ambito la Bioinformatica si occupa dell'acquisizione, memorizzazione, distribuzione, analisi e interpretazione dei dati prevalentemente nell'ambito della biologia molecolare, della genetica e della biochimica con collegamenti sempre più importanti con la medicina.

Scopi della Bioinformatica sono quelli di fornire in tempi rapidi le informazioni e le metodologie d'indagine, che consentano ad esempio di fornire alla scienza medica le necessarie informazioni per comprendere i meccanismi alla base di tutte le possibili disfunzioni. In quest'ottica, i metodi e gli strumenti messi a disposizione dalla Bioinformatica giocano un ruolo fondamentale nello sviluppo delle biotecnologie.

L'investimento di Stati Uniti, Canada, Giappone e Russia nella Bioinformatica è mirato soprattutto alla creazione di centri, reali o virtuali, che provvedano servizi e conoscenza alla comunità mentre, allo stesso tempo, spingono in nuove direzioni grazie ad attività di ricerca avanzate.

Lo sviluppo di piccole e medie aziende con finalità bioinformatiche in Europa è, come ci si può aspettare, proporzionale all'investimento in ricerca dei vari Stati e quindi molto attivo in Inghilterra, Francia, Germania, in crescita in Spagna, Olanda e paesi nordici.

L'Italia è ben collocata nel panorama europeo dal punto di vista scientifico, anche se le risorse limitate rispetto soprattutto a Germania e Regno Unito hanno finora reso difficile l'implementazione di progetti indipendenti a grande respiro e di risonanza internazionale. Oggi si è creata un'enorme attesa, soprattutto negli Stati Uniti, in Canada e in Giappone per un impatto economico crescente della bioinformatica. Stime conservative fanno assumere che la Bioinformatica solo negli Stati Uniti permetterà di generare almeno 7 miliardi di dollari nei prossimi tre anni e che il mercato crescerà del 20% entro il 2006.

L'Italia per poter mantenere il passo in un settore altamente strategico deve prevedere misure di potenziamento della bioinformatica nel settore biotecnologico, biomedico e sanitario operando sui seguenti piani distinti:

- la ricerca,
- lo sviluppo dei sistemi,
- l'infrastruttura operativa,
- le applicazioni bioinformatiche nella clinica quotidiana,
- la formazione.

L'obiettivo è quello di creare le condizioni per uno sviluppo omogeneo del settore della bioinformatica ed è quindi necessario che le azioni che si intraprendono su questi piani siano a ricaduta immediata.

Indice

Indice.....	4
Introduzione	5
Stato dell'arte	7
Cenni storici	7
Situazione Internazionale.....	7
Situazione Europea	8
Situazione Nazionale.....	9
L'impatto economico prevedibile	10
Le necessità italiane nel campo della bioinformatica	11
Le possibili linee d'intervento.....	13
La ricerca in Bioinformatica	14
Lo sviluppo dei sistemi	16
L'infrastruttura operativa	17
L'infrastruttura informatica.....	17
L'infrastruttura concettuale.....	18
Le biotecnologie nella clinica quotidiana	21
La formazione	23
Conclusioni	24
Bibliografia	25

Introduzione

La biologia e la medicina hanno da sempre cercato di studiare l'organismo umano a tutti i livelli sia morfologico e sia funzionale per il miglioramento della salute umana.. Con l'introduzione delle più moderne tecnologie di biologia molecolare si sono iniziati a comprendere meglio le regole d'espressione genetica ed i vari passaggi metabolici e sia la struttura del DNA, RNA e delle proteine.

Questa enorme complesso d'informazioni, consentirà di comprendere il flusso delle informazioni che governano il passaggio dal Genoma al Fenotipo di un organismo. Un utilizzo appropriato di questi dati associate ad appositi programmi d'analisi, porterà nuove possibilità per la comprensione dell'espressione dei geni, la loro regolazione e le malattie genetiche a loro correlate in caso di mutazioni o disfunzioni metaboliche. La disciplina che si occupa di queste problematiche per trattamento dell'informazione biologica a tutti i livelli è la *Bioinformatica*.

La Bioinformatica si occupa, quindi, dell'acquisizione, memorizzazione, distribuzione, analisi e interpretazione dei dati prevalentemente nell'ambito della biologia molecolare con collegamenti sempre più importanti con la medicina. Questa nuova disciplina scientifica utilizza metodi di matematica, informatica, biologia, medicina, fisica allo scopo di migliorare la comprensione dei fenomeni biologici. Di seguito, si elenca a titolo alcuni dei principali obiettivi della Bioinformatica:

- Sviluppo di strumenti per la generazione ed il mantenimento dell'informazione proveniente dalle varie fonti: mappa fisica, mappa genetica, mappa cromosomica, mappa citogenetica, polimorfismi, e l'informazione relativa alle sequenze genomiche e proteiche etc.;
- Raccolta e organizzazione delle informazioni genetiche associate alle patologie mediche;
- Sviluppo di programmi di calcolo per l'analisi delle sequenze;
- Sviluppo d'interfacce grafiche in grado di visualizzare in maniera efficace l'informazione richiesta;
- Sviluppo di metodi software che consentano di agevolare tutte le fasi dei progetti;
- Sviluppo di strutture per database specializzate ed integrate;
- Sviluppo di strumenti informatici, includendo software e hardware e algoritmi per l'organizzazione e l'analisi dei dati;
- Realizzazione di standard per lo scambio e la descrizione dei dati;
- Realizzazione di una rete dati per la raccolta la distribuzione e l'aggiornamento costante di tutta l'informazione prodotta;
- Raccolta della bibliografia, brevetti e altri database di supporto all'informazione specifica;
- Predizione dei geni nelle sequenze di DNA;
- Predizione delle strutture tridimensionali delle proteine partendo dalle sequenze primarie;
- Predizione delle funzioni biologiche e biofisiche sia dalle sequenze e sia dalle strutture;
- Simulazione dei processi metabolici e cellulari basati su queste funzioni.
- Realizzazioni di sistemi per la correlazione dell'informazione in sistemi biologici complessi.

Da questo elenco risulta evidente che uno dei principali scopi della Bioinformatica è di fornire in tempi rapidi le informazioni e le metodologie d'indagine, che consentano ad esempio di fornire alla scienza medica le necessarie informazioni per comprendere i meccanismi alla base di tutte le possibili disfunzioni d'origine genetica. Dalle mutazioni in regioni funzionali del DNA, alla mancata produzione di una certa proteina a causa di un anomalo funzionamento dei fattori di trascrizione, fino a comprendere il funzionamento d'ogni singolo gene in relazione con gli altri geni nei diversi processi metabolici che continuamente avvengono nel corso dell'intera vita.

L'industria farmaceutica utilizza queste informazioni e metodologie e conoscenze al fine di produrre medicine, proteine specifiche e terapie geniche in grado di intervenire in maniera selettiva per risolvere le possibili cause di malattie.

Inoltre la gran parte dei prodotti genici del genoma umano ha più di una funzione, in alcuni casi addirittura in antagonismo con altre. E' per questo necessario identificare i modelli quantitativi che stanno alla base di questi processi.

In quest'ottica i metodi e i programmi messi a disposizione dalla Bioinformatica giocano un ruolo fondamentale nello sviluppo delle biotecnologie.

Stato dell'arte

Cenni storici

Già alla fine degli anni Settanta risultava chiaro che non ci poteva più affidare alla capacità dei singoli gruppi di ricerca per immagazzinare e organizzare i dati biologici e che occorreva uno sforzo informatico per sfruttarne le potenzialità.

L'Europa coglie quest'occasione con grande tempismo e per prima: la banca dati dell'EMBL nasce nel 1980, mentre gli Stati Uniti dovranno attendere il 1986 per fondare il National Center for Biotechnology Information, depositario della banca dati GenBank.

Negli anni immediatamente successivi, anche il primo Dipartimento di Ricerca in Bioinformatica, il Biocomputing Group dell'EMBL, nasce in Europa. Il programma di cui fanno parte Arthur Lesk, Chris Sander e Pat Argos sfocerà nel 1996 nella fondazione dell'EBI (European Bioinformatics Institute di Hinxton, Cambridge) che è oggi tra i punti di riferimento più importanti della bioinformatica mondiale, ma soprattutto addestra giovani ricercatori in questa nuova disciplina. Al Biocomputing dell'EMBL e ai gruppi di Bioinformatica strutturale di Cambridge, guidati, tra gli altri, da Sir Tom Blundell, Janet Thornton e Cyrus Chothia, devono il loro ingresso alla bioinformatica molti, se non tutti, i gruppi di riferimento europei.

Le competenze dei gruppi pionieri di questa disciplina sono oggi affiancate da scuole più "informatiche" che le completano e rendono la bioinformatica un punto di forza della ricerca Europea che va assolutamente coltivata e incoraggiata.

Situazione Internazionale

L'investimento di Stati Uniti, Canada e Giappone nella Bioinformatica è mirato soprattutto alla creazione di centri, reali o virtuali, che provvedano servizi e conoscenza alla comunità mentre, allo stesso tempo, spingono in nuove direzioni grazie ad attività di ricerca avanzate.

Gli Stati Uniti riconoscono che anche solo il mantenimento dei sistemi attualmente in uso richiede aumenti di investimenti. L'NCBI (National Center for Biotechnology Information) è un esempio eclatante, un incremento degli investimenti di circa il 10% annuo è necessario solo per mantenere l'attuale livello di servizi e ricerca. Il personale ad oggi impiegato presso l'NCBI è vicino alle 400 unità.

Negli Stati Uniti soltanto, l'investimento economico nel settore della bioinformatica è valutato intorno ai 500 milioni di dollari all'anno (la metà circa di quello mondiale) di cui i due terzi sono delle grandi industrie farmaceutiche che si aspettano di ridurre i costi e i tempi di sviluppo di un farmaco di circa il 30% grazie alle tecniche bioinformatiche. Questo ha evidentemente aperto la strada ad un gran numero di società che forniscono software e servizi alle grandi industrie farmaceutiche, tra cui per esempio Accelrys, Applied Biosystems, DNASTar, Informax, Genamix, Rosetta (poi acquisita dalla Merck), ecc.

Situazione Europea

L'EBI European Bioinformatics Institute rimane oggi un punto focale della bioinformatica europea, ma gruppi di dimensioni e impatto notevole sono presenti in praticamente tutti i paesi dell'Unione.

Il **Regno Unito** ha per esempio tenuto alto il suo livello negli aspetti strutturali, con gruppi di grande competenza e fama all'MRC di Cambridge, al King's College e allo University College di Londra, affiancati da gruppi che si occupano di analisi genomiche non solo in Inghilterra ma anche in Scozia ed Irlanda. Alcuni dei più noti gruppi di ricerca hanno dato vita a vari spin-off (vedi Inpharmatica). Lo sviluppo della bioinformatica in Inghilterra avviene anche grazie al regolare interesse dell'agenzia di finanziamento Wellcome Trust per la genomica che si traduce in sostanziali investimenti in bioinformatica ad esempio presso il Sanger Centre a Cambridge.

La **Germania** sta dando vita ad un aggressivo programma di training con lauree di primo e secondo livello in Bioinformatica in varie città. Ad oggi sono stati attivati 34 nuovi cattedre universitari in Bioinformatica. Ci si aspetta che questo dia un grande impulso alla ricerca che, in questo paese, si sta concentrando sull'analisi genomica, sulla genomica funzionale e sulla biologia dei sistemi. Esistono Max Planck Institut e grandi dipartimenti interamente dedicati alla Bioinformatica, per esempio il Max Planck di Berlino, Monaco e il Centro di Bioinformatica di Saarbrueken.

In **Spagna** vi sono alcuni gruppi estremamente interessanti, soprattutto all'interno del Consiglio Nazionale delle ricerche a Madrid e del Genome Informatics Research Laboratory di Barcellona, centro che si sta affermando come uno dei punti cardine per la ricerca e caratterizzazione dei geni umani per via bioinformatica. In Spagna è stato istituito un "Istituto Virtuale di Bioinformatica" che coinvolge la maggior parte dei gruppi spagnoli. L'Istituto, pur essendo virtuale dal punto di vista logistico, è organizzato con Dipartimenti, direttori, progetti, ecc. e rappresenta un interessante esempio di cooperazione nazionale strutturata e gestita managerialmente.

Anche il **Portogallo** ha iniziato un programma educativo in Bioinformatica, ancora agli inizi ma che sta coinvolgendo docenti da tutta Europa.

In **Francia**, presso l'Istituto Pasteur ed il CNRS, sono attivi diversi gruppi dedicati alla Bioinformatica.

La **Svizzera** si è conquistata un posto centrale nel panorama bioinformatico grazie al SIB - Swiss Institute for Bioinformatics, iniziatore della lodevole iniziativa della banca dati SwissProt, e che oggi è un partner essenziale dell'EBI nell'area della banche dati nell'area della proteomica.

In **Olanda, Svezia, Finlandia, Belgio** operano gruppi di bioinformatica di ottimo livello scientifico, generalmente derivanti da quell'iniziale gruppo dell'EMBL di cui si è discusso precedentemente.

Un discorso a parte merita la **Danimarca** che, avendo dato il via al "Centre for Biological Sequence Analysis", è riuscita a conquistarsi un ruolo centrale nella bioinformatica europea partecipando alla maggior parte dei progetti di finanziamento comunitari nella disciplina, ancora una volta dimostrando l'importanza di una sinergia di sforzi coordinati.

Va inoltre menzionato quanto sta accadendo nei paesi dell'est europeo (specialmente **Polonia** ma anche **Lituania**) il rientro di alcuni ricercatori formati negli Stati Uniti ha permesso di organizzare gruppi di ricerca che, nell'arco di qualche anno, sono già entrati a far parte dell'establishment scientifico in bioinformatica.

Lo sviluppo di piccole e medie aziende con finalità bioinformatiche in Europa è, come ci si può aspettare, proporzionale all'investimento in ricerca dei vari Stati e quindi molto attivo in Inghilterra e in Germania, e in crescita in Spagna.

L'Unione Europea con il II VI PQ dell'UE, nella linea Scienze della Vita, Genomica e Biotecnologie per la Salute - Genomica Avanzata e sue Applicazione per la Salute, ha come obiettivo migliorare la comprensione e la decifrazione delle informazioni genomiche, individuandone la funzione, la struttura e la relazione con la salute umana. In tale ambito, uno dei risultati più attesi è creare gli strumenti informatici che permettano alla comunità medico-scientifica e al mondo industriale di analizzare e utilizzare i dati sui genomi per lo sviluppo di nuove terapie, medicinali e diagnosi sanitarie sempre più accurate. Ciò rientra in un orientamento generale, che parla oramai di strumenti necessari a supportare la cosiddetta medicina personalizzata (*personalized healthcare*).

Pertanto l'Unione Europea sta attuando un programma di finanziamento della bioinformatica sia nel campo dello sviluppo di metodi (BioSapiens, IntAct, EMSD, ORIEL, E-BIOSCIENCE, ecc.), sia delle tecnologie di supporto (BIOINFOGRID, EMBRACE, EGEE), sia dell'applicazione della bioinformatica all'area medica (INFOBIOMED, SYMBIOMATICS).

Situazione Nazionale

L'Italia è ben collocata nel panorama europeo dal punto di vista scientifico, anche se le risorse limitate rispetto soprattutto a Germania e Regno Unito hanno finora reso difficile l'implementazione di progetti indipendenti a grande respiro e di risonanza internazionale.

Poli molto attivi sia nella ricerca sia nello sviluppo in bioinformatica in Italia sono, per esempio, Napoli, Roma, Milano, Bologna, Genova, Bari, Palermo, Padova, Trento, Trieste e Cagliari.

L'aspetto forse più interessante del panorama italiano è che esso copre un largo spettro di attività bioinformatiche, che vanno dall'analisi genomica alla genomica funzionale, dalla predizione di strutture tridimensionali, dall'assegnazione funzionale alla predizione di proprietà di prodotti genici, dai sistemi basati su GRID a metodi di intelligenza artificiale, etc.

In altre parole, dal punto di vista delle competenze, la bioinformatica italiana costituisce un'opportunità forse unica per lo sviluppo di ricerca e tecnologia.

Molti dei gruppi italiani sono coinvolti in iniziative internazionali di vasto respiro, come società internazionali, gruppi di lavoro, commissioni di valutazioni scientifiche, progetti europei, iniziative educative. Sarebbe impossibile elencarle tutte, e soprattutto difficile tenere traccia di tutte le attività in cui i bioinformatici italiani vengono coinvolti continuamente.

Il MIUR ha investito negli ultimi tre anni risorse finanziarie per un valore complessivo di 32 Meuro destinate a iniziative scientifiche in ambito nazionale e internazionale strettamente connesse all'area "Bioinformatica".

Queste iniziative attualmente in corso possono essere così strutturate:

- Realizzazione di due laboratori nazionali in Bioinformatica pubblico-privato (FIRB 2003-2004) ai quali sono collegate unità di ricerca dislocate in varie località italiane. Il finanziamento pubblico per queste iniziative programmate per un quinquennio d'attività è pari a 8,404 Meuro.
- Il MIUR, nell'ambito degli interventi CIPE per il Mezzogiorno ha destinato 10 Meuro per la realizzazione di Laboratori pubblico-privati (il bando è in corso di espletamento).
- Finanziamento di progetti di ricerca congiunti con centri e/o istituti di eccellenza di altri Paesi (USA, Canada, UK, India, Cina, Russia, Israele), con una previsione di finanziamento dell'ordine di 8 Meuro.
- Realizzazione del laboratorio di Bioinformatica previsto dall'accordo tra MIUR, Ministro per l'Innovazione Tecnologica, Università degli Studi di Trento, Provincia di Trento e Microsoft. Per tale iniziativa si prevede un impegno finanziario del FIRB pari a circa 4 Meuro.

Nonostante questo panorama interessante, le iniziative di trasferimento tecnologiche sono rimaste eccessivamente ristrette e non si è finora sfruttato appieno questo patrimonio per mancanza, a nostro parere, delle appropriate infrastrutture, di coordinazione e di adeguate risorse.

Il patrimonio di conoscenze accumulato da questi gruppi di ricerca e i giovani ricercatori che sono impiegati e formati nei vari progetti, corre il rischio di essere completamente vanificato nel giro di poco tempo, in mancanza di un programma a medio lungo periodo. Le società che svolgono attività bioinformatiche in Italia sono infatti pochissime, fra cui:

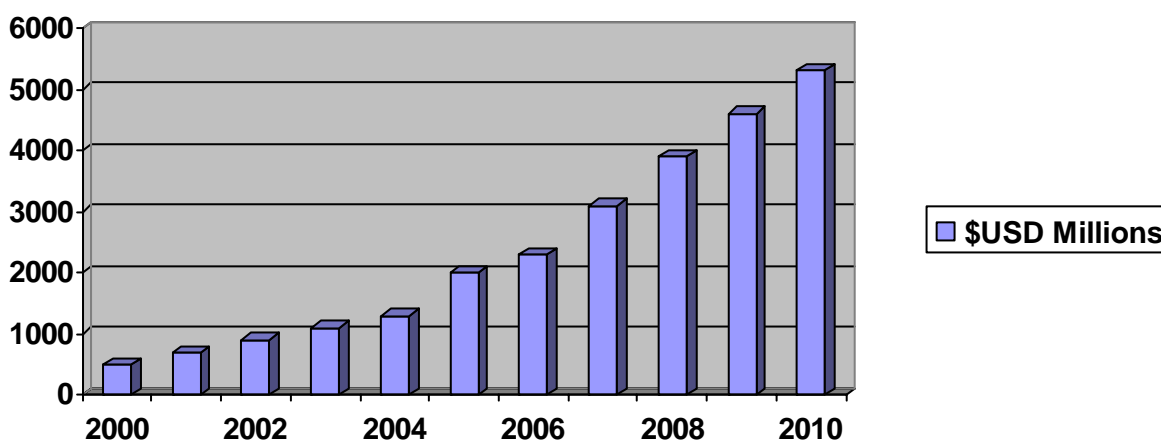
- **SharDNA** a Cagliari (che si occupa principalmente di genomica e quindi necessita di un supporto bioinformatica per la gestione e l'analisi dei dati provenienti dallo studio di popolazioni isolate).
- **ProteoGenBio** a Pisa .
- **BioDec** a Bologna.

Alcune industrie e centri di eccellenza universitaria hanno poi sviluppato al loro interno delle competenze in Bioinformatica, tra queste la **Chiron**, l'**IRBM** (Laboratorio Merck in Italia), la **Glaxo** , centri di eccellenza di ricerca **CNR**, **Università**, **CISI**, e poche altre realtà.

L'impatto economico prevedibile

C'è un'enorme attesa, soprattutto negli Stati Uniti, in Canada, Giappone e Europa per un impatto economico crescente della bioinformatica, per esempio grazie all'individuazione di nuovi agenti terapeutici e farmacologici ci si aspetta che la bioinformatica sia in grado di aumentare il numero di bersagli terapeutici da circa 400 a circa 4000 nei prossimi vent'anni e quindi, già solo grazie a questo, ci si aspetta che i proventi dell'industria farmaceutica aumentino di un fattore mille. Stime conservative fanno assumere che la Bioinformatica solo negli Stati Uniti permetterà di generare almeno 7 miliardi di dollari nei prossimi tre anni e che il mercato crescerà del 20% entro il 2006.

Nella figure seguente è mostrato l'andamento del mercato delle industrie bioinformatiche in USA dal 2000 al 2010 (Frontline Research).



E questa è solo una delle possibili applicazioni della bioinformatica che ci si aspetta abbia un impatto anche sullo sviluppo di vaccini, nuove biotecnologie, nanotecnologie, etc.. Gli investimenti industriale in bioinformatica saranno almeno triplicati entro il 2010, prevalentemente nel settore che cresce più velocemente, cioè quello del software di analisi e banche dati specializzate.

Le necessità italiane nel campo della bioinformatica

Il mondo della bioinformatica italiana ha le seguenti necessità:

- rafforzare il collegamento dell'industria con il mondo della ricerca sia universitaria che degli Enti Pubblici di Ricerca come il CNR;
- sviluppare nuove metodologie per la gestione delle biobanche fra loro integrate;
- consentire un costante aggiornamento delle conoscenze in un settore altamente innovativo come quello della System Biology applicata all'analisi dei dati generati dalla Genomica e Proteomica di vari organismi;
- stimolare la crescita di nuove imprese nel settore della Bioinformatica applicata al settore Biomedicale e Biotecnologico;
- offrire una base strategica per agevolare la formazione interdisciplinare di nuovi profili, bioinformatici, biotecnologici e biomedici;
- stimolare l'interazione e l'integrazione delle iniziative già finanziate con i laboratori pubblico-privati finanziati dal FIRB, FAR e dalle infrastrutture finanziate ad hoc per la Bioinformatica;
- integrare le infrastrutture di calcolo adeguate alle esigenze che si stanno profilando nel campo della Biologia, delle Biotecnologie, della Genetica e della Medicina;
- finanziare l'adesione dell'Italia a iniziative internazionali;
- finanziare un progetto congiunto per l'accessibilità di banche dati bibliografiche di interesse internazionale;
- favorire l'interazione dei laboratori pubblico-privati con iniziative per lo sviluppo delle applicazioni di GRID computing e calcolo avanzato nell'ambito Biomedico e ospedaliero in Italia;

Per ottenere i precedenti obiettivi il mondo della bioinformatica italiana ha necessità di reperire personale giovane con un'adeguata formazione scientifica. Si riscontra, infatti, un gap in biologia e informatica; la creazione di un percorso di studi che istruisca in entrambi gli aspetti darebbe una risposta concreta alle esigenze del comparto industriale.

Le industrie operanti nel comparto delle biotecnologie hanno una forte dipendenza dalla ricerca e ed il settore della bioinformatica e l'informatica medica non fanno eccezione. Un importante vantaggio competitivo sarebbe rappresentato dalla possibilità di incrementare in maniera sostanziale le collaborazioni sia a livello di ricerca di base che applicata. Aree di interesse sono, ad esempio, la ricerca di nuovi sistemi ed algoritmi per l'identificazione funzionale di geni e proteine, l'adozione di tecniche e strumenti per l'estrazione automatica dell'informazione da basi di dati testuali quali ad esempio letteratura scientifica, classificazione automatica delle proteine, sviluppo di strumenti per la definizione/esecuzioni di esperimenti "in silico" e la gestione automatica di flussi di lavoro soprattutto in collegamento a sistemi di calcolo distribuiti.

Pertanto l'industria ha necessità di luoghi dove imprese, laboratori universitari, istituzioni economiche e di ricerca operano a stretto contatto al fine di:

- Sviluppare ricerca fondamentale e a carattere applicativo;
- Mettere a punto nuove tecnologie, prodotti e processi;
- Valorizzare i risultati della ricerca e trasferire tecnologie e innovazioni al mondo produttivo;
- Creare reti di cooperazione nazionali e internazionali;
- Sviluppare attività economiche e imprenditoriali ad alta intensità di conoscenza.
- In parole povere, occorrono centri di eccellenza capillarmente distribuiti sul territorio con una serie di servizi (cataloghi del know-how e delle tecnologie pronte all'uso, panel di esperti e consulenti, seminari e workshop, etc.) per estendere e sviluppare i cosiddetti CLUSTER e Reti tematiche.

Il sempre maggior costo per lo sviluppo di nuovi farmaci e le barriere economiche all'accesso a tali prodotti per i paesi poveri stanno forzando la ricerca di nuove vie per risolvere questi problemi. Ad esempio, l'adozione dei concetti alla base della "Translational Medicine", dove tutti i processi industriali di ricerca, sviluppo, test e produzione sono orientati al paziente/gruppo/etnia oggetto del prodotto, necessitano in sempre maggior misura della integrazione fra dati genomici, proteomici e le informazioni contenute nelle biobanche. Inoltre sviluppare ed adottare considerazioni etiche volte a minimizzare ed ottimizzare l'uso della sperimentazione sia sulla gestione dei dati sensibili nell'uomo e sia nell'utilizzo della sperimentazione animale.

La presenza di infrastrutture operative di riferimento e l'adozione di standard internazionali sono fondamentali sia per la ricerca che per l'industria. Un ambiente di riferimento permette una pianificazione degli investimenti più precisa, facilita le collaborazioni sia tra industrie che con il mondo della ricerca accademica, assicura i necessari livelli di sicurezza nel trattamento ed accesso ai dati garantendo la sicurezza della Proprietà Intellettuale delle singole aziende.

Infatti per mantenersi al livello presente non basta però più il lavoro di gruppi indipendenti: l'aumento enorme dei dati disponibili rende oggi essenziale la presenza di infrastrutture e, soprattutto di risorse che oggi sono più carenti in Europa rispetto agli Stati Uniti e al Giappone.

L'industria ha necessità di reperire personale giovane con un'adeguata formazione scientifica. Si riscontra, infatti, un gap tra biologia ed informatica; la creazione di un percorso di studi che cumuli entrambi gli aspetti darebbe una risposta concreta alle esigenze del comparto industriale.

Nell'ambito di una sanità "Informata e Formata" sulle specifiche ricerche di Bioinformatica e' doveroso sensibilizzare i medici di medicina generale, biologi e altri operatori interessati alla materia, su argomenti che possano essere applicati alla pratica clinica.

Da qui nasce la necessità di sviluppare un adeguato programma di formazione per gli operatori del Servizio Sanitario, che potrebbe essere collegata al sistema dell'ECM Educazione Continua in Medicina programma voluto dal Ministero della Salute.

Le possibili linee d'intervento

Al fine di potenziare il ruolo dell'informatica nelle biotecnologie della salute occorre operare sui seguenti piani distinti:

- la ricerca,
- lo sviluppo dei sistemi,
- l'infrastruttura operativa,
- le biotecnologie nella clinica quotidiana,
- la formazione.

L'obiettivo è quello di creare le condizioni per uno sviluppo omogeneo del settore della bioinformatica ed è quindi necessario che le azioni su questi piani siano quelle che abbiano una ricaduta immediata. Pertanto i temi prioritari riguardano:

- Stabilire un'effettiva cooperazione pubblico-privato fra la comunità di ricercatori in ambito nazionale ed internazionale;
- La ricerca vista sia come sviluppo di sistemi software per la biologia computazionale che come messa a punto di metodologie innovative per l'analisi e la ricerca di correlazioni tra i dati biologici.
- Lo studio di applicazioni legate alle metodologie per la realizzazione delle biobanche.
- La creazione di una infrastruttura per favorire la collaborazione e il coordinamento tra i vari centri.
- Favorire la realizzazione di ricerche basate su accordi internazionali al fine di agevolare lo scambio di ricercatori (Research HUB), e la creazione di laboratori congiunti (Joint Research LAB);
- Introdurre della Bioinformatica e le biotecnologie nella pratica clinica quotidiana.
- La formazione estesa all'aggiornamento professionale.

Va tenuto conto che le ricadute economiche di un investimento congruo nel comparto dovranno devono valutate in termini di profitti provenienti non solo dalla diretta commercializzazione dell'invenzione, ma anche da una maglie posizionamento del sistema paese nel contesto internazionale.

La ricerca in Bioinformatica

E' in atto una progressiva raccolta, organica, in basi di dati, di informazioni genetiche e mappature del Genoma Umano, prodotte dai tanti gruppi di ricerca che, in regime di autonomia, cooperano al progredire dell'analisi e dello sviluppo della Genomica.

Obiettivo di tale raccolta è migliorare la comprensione e la decifrazione delle informazioni genomiche, individuandone la funzione, la struttura e la relazione con la salute umana.

Pertanto si mira a creare strumenti informatici che permettano alla comunità medico-scientifica e al mondo industriale di:

- analizzare e utilizzare i dati sui genomi per lo sviluppo di nuove terapie, medicinali e diagnosi sanitarie sempre più accurate;
- raccogliere le informazioni al fine di analizzare e utilizzare i dati sui genomi per lo sviluppo di nuove terapie, medicinali e diagnosi sanitarie sempre più accurate.

Sul piano della ricerca è necessario lo sviluppo di sistemi software per la Bioinformatica, diretti ad esempio:

- alla modellizzazione della struttura tridimensionale a livello atomico delle molecole biologiche;
- all'analisi, lo studio e la predizione dei modi di interazione delle molecole biologiche;
- alla classificazione di dati basati su metodi di intelligenza artificiale quali le reti neurali, gli Hidden Markov Models, le Support Vector machines, etc. mirati all'assegnazione di funzione o di caratteristiche funzionali ai prodotti genici di interesse;
- alla visualizzazione, l'analisi, il confronto e la ricerca in banche dati di interazioni tra molecole biologiche e di pathway metabolici;
- alla correlazione dei dati di interazione con dati disponibili sulle basi genetiche di patologie complesse;
- all'identificazione di elementi regolativi della trascrizione all'interno e nell'intorno di geni;
- all'identificazione delle reti di regolazione genetica a partire da dati di espressione dalla conoscenza degli elementi regolativi. Questi algoritmi permetteranno di capire come i geni agiscono in maniera concertata e pertanto di ricostruire i pathway biologici in cui essi sono coinvolti e quindi capire la loro funzione.
- all'identificazione di siti di legame di fattori di trascrizione ha lo scopo di fornire un insieme di strumenti flessibili per prelevare i dati da sorgenti multiple.
- all'identificazione di reti di regolazione genetica. Una rete genetica può essere rappresentata matematicamente utilizzando equazione differenziali ordinarie i cui parametri sono incognite che bisogna stimare utilizzando i dati sperimentali quali profili di espressione genetica per identificare le interazioni tra geni.
- all'interpretazione automatica di letteratura scientifica e di divulgazione delle più note riviste internazionali.
- alla modellistica biomedica e all'apprendimento automatico nella systems biology. A tal fine occorre impiegare metodologie innovative per la generazione di ipotesi su reti di regolazione genica e per lo studio e la simulazione di particolari processi cellulari.
- allo sviluppo di metodi basati su modelli qualitativi e metodi probabilistici, in grado di tener conto di informazioni incerte e conoscenza imperfetta sul dominio allo studio.
- allo sviluppo di sistemi informativi sanitari correlati a dati genetici. In questa attività verranno utilizzati ed adattati metodi ed approcci sviluppati nell'ambito dei sistemi informativi sanitari per tener conto delle esigenze specifiche delle applicazioni che trattano dati genetici, genomici e di proteomica.

- allo sviluppo di cartelle cliniche elettroniche orientate alla gestione ed all'integrazione dei dati clinici con dati genetici, di genomica funzionale e di proteomica.
- alla progettazione di sistemi informativi per la conduzione di studi di caratterizzazione genetico-clinica
- alla gestione di dati e database orientati agli aspetti di privacy e sicurezza informatica.
- a costruire workflow per il design e lo sviluppo di applicazioni di supporto al lavoro collaborativi in bioinformatica.
- a sviluppare metodologie basate sul GRID computing e calcolo avanzato specifiche per la bioinformatica per fare fronte a calcoli di notevole complessità algoritmica.

Lo sviluppo dei sistemi

Oggi si è in presenza di una situazione dove il rapido avanzamento della ricerca e delle tecnologie applicate alla medicina ha portato ad un considerevole aumento di interesse verso le collezioni di materiali biologici umani. Le raccolte di campioni di tessuti e linee cellulari, da cui si ottengono acidi nucleici e proteine, sono definite biobanche e rappresentano un'importante fonte di risorse per la diagnosi e per la ricerca da quella di base fino alla sperimentazione di terapie per le malattie genetiche. Il successo delle ricerche per l'identificazione dei geni malattia, dei geni di suscettibilità e delle possibili applicazioni terapeutiche, compreso lo sviluppo di nuovi e specifici farmaci, si avvale della possibilità di disporre di campioni biologici di persone affette o portatrici o predisposte a patologie su base genetica, o che manifestano variabilità alla risposta dei farmaci.

Il materiale biologico è diventato una risorsa molto preziosa e deve essere quindi favorito lo sviluppo e la organizzazione di tale attività e regolamentato l'utilizzo, il mantenimento e l'interazione tra varie biobanche.

Queste collezioni sono numerose e di diversa entità ed è quasi impossibile quantificarle e avere informazioni sul loro stato di conservazione e sulla loro gestione. Pertanto di fronte al crescente aumento delle biobanche genetiche e degli interessi ad esse collegati, si ritiene necessario che in Italia si giunga rapidamente ad una regolamentazione per la loro istituzione ed utilizzo, e venga realizzato un censimento di quelle già esistenti.

E' da potenziare il coordinamento delle biobanche secondo le indicazioni fornite dall'OCSE, provvedendo alle autorizzazioni e alle certificazioni.

Sul piano dello sviluppo dei sistemi occorre studiare:

- Metodologie di progetto di una biobanca;
- Linee guida per la progettazione di una biobanca, interoperabile con le altre biobanche;
- Valutazione della qualità di una biobanca.

Infatti, uno degli obiettivi perseguibili è la definizione del protocollo e delle modalità di popolamento delle biobanche e le specifiche per una progettazione orientata alla efficienza dell'utilizzo. Occorre infatti, grazie all'informatica, affrontare la multidisciplinarietà:

- in modo da rappresentare tutte le numerose tipologie di dati per la definizione del profilo genomico di un individuo che richiede spesso l'analisi simultanea di molti geni e lo studio della loro attività in modo parallelo (e.g. nell'uomo numerosi geni possono avere decine o centinaia di mutazioni diverse responsabili della stessa malattia);
- lo studio di queste mutazioni in parallelo sottolinea proprio la capacità di fornire informazioni su numerose alterazioni o stato di attività di geni nel medesimo istante.

La tipologia dei dati deve permettere di modellare i seguenti dati:

- sequenze di DNA;
- SNP (Single Nucleotide Polymorphism), ovvero variazioni di singole basi di DNA (consentono di studiare a livello molecolare le differenze individuali tra individui e sviluppare profili genetici caratteristici per ogni persona o gruppi di persone);
- annotazioni manuali o automatiche (e.g. SNPs);
- livelli di espressione genica, mappe proteomiche;
- interazioni tra geni e proteine;
- pathway metabolici;

L'infrastruttura operativa

Lo sviluppo tecnologico ha reso sempre più semplice e rapida la possibilità di sequenziare genomi di organismi viventi, dai più semplici come i virus a quelli più complessi come l'uomo. Nei prossimi anni sarà possibile ottenere sequenze genomiche in modo molto rapido e preciso. Questo fa nascere la necessità di definire i modi in cui memorizzare, trattare e sfruttare questa immensa quantità di dati. In analogia ad altri campi della scienza, come ad esempio la fisica delle particelle o l'astronomia, la genomica e le biotecnologie dovranno affrontare il problema del sovraccarico di informazione. Diventa quindi essenziale la creazione di schemi logici e di modalità di scambio tra le molteplici entità per sviluppare le basi di dati appropriate.

Il primo passo è necessariamente costituito dalla definizione della semantica dei dati, ossia la definizione del significato delle entità che definiscono il "dominio" sotto osservazione e le relazioni che le legano. La definizione delle modalità di scambio delle informazioni rappresentano il secondo passo, infatti è fondamentale poter trasferire l'informazione tra le varie entità che la trattano in maniera precisa e sicura.

La crescita esponenziale della quantità di dati da trattare sta inoltre creando un problema di capacità di elaborazione. Saranno sempre di più necessari sistemi di calcolo potenti e veloci e quindi sarà sempre più indispensabile l'impiego di sistemi distribuiti capaci di mettere assieme le risorse necessarie ovunque si trovino nel mondo, il tutto in maniera semplice e sicura.

Tutto questo potrà permettere, ad esempio lo studio delle ragioni alla base della diversa efficacia di un farmaco nei vari pazienti o di capire le basi genetiche di malattie molto importanti. Oppure sarà più semplice studiare il comportamento dei microrganismi e comprenderne a pieno la loro interazione le cellule.

Altrettanto importante è l'osservanza e l'adesione agli standard internazionali già definiti o in via di definizione in modo da facilitare sempre più l'integrazione della bioinformatica italiana con il resto delle comunità scientifiche ed industriali.

L'infrastruttura informatica

La quantità di dati attualmente disponibile non è che una minima parte di quella che è lecito attendersi nel prossimo futuro. Questo pone un problema di capacità di elaborazione e di gestione non banale che implica necessariamente l'uso di sistemi distribuiti.

Nel campo dell'infrastruttura informatica per il supporto di attività scientifiche che richiedono accesso a grosse potenze di calcolo ed a risorse geograficamente distribuite si parla sempre più spesso della GRID. Questo termine è stato coniato in USA alla fine degli anni 90 [ref. The Grid a Blue print for a new computing infrastructure. Foster and Kesselman ed.] per analogia con la rete di distribuzione elettrica che è un ottimo esempio di un'infrastruttura pervasiva ed altamente distribuita che ha generato a suo tempo una rivoluzione nei metodi di produzione industriale e nella società.

Per GRID s'intende un'infrastruttura informatica, formata da risorse di calcolo (in genere centri di calcolo di piccole, medie e grosse dimensioni), altre risorse di tipo vario, banche dati, strumenti scientifici (sensori, satelliti, acceleratori di particelle), sistemi di visualizzazione e strumenti per il supporto di lavoro condiviso come video conferenze e librerie digitali interconnesse da una rete di comunicazione ad alte prestazioni ed alta affidabilità come in effetti Internet sta sempre più diventando ed un software adeguato a creare con tutti questi elementi un centro di calcolo virtuale che dia l'impressione all'utente finale in fronte del suo calcolatore personale di avere accesso alle stesse facilità e risorse come se si trattasse di un centro locale.

Le funzioni di questo software si possono riassumere ad un sistema di identificazione dell'utente che permetta un accesso a risorse distribuite senza il bisogno per l'utente di essere riconosciuto ed autorizzato in tutte le risorse remote. Dietro queste funzioni c'è un sistema molto complesso e distribuito implementato sempre più spesso sotto forma di servizi Web (altra analogia

con il Web) che permette di localizzare le risorse necessarie per l'elaborazione del programma richiesto, ottimizzare il carico dei vari sistemi di calcolo coinvolti ed assicurare il buon funzionamento di tutta l'infrastruttura di Grid.

La bioinformatica ha molto in comune con altre scienze come la Fisica delle particelle elementari che hanno già da qualche anno fatto ricorso ad un'infrastruttura di Grid per i loro bisogni informatici. Gli elementi più importanti per consigliare l'adozione di questo approccio informatico sono i seguenti: grandi mole di dati distribuite geograficamente e non necessariamente di tipo ed accesso uniforme; possibilità di parallelizzare il calcolo e distribuirlo in unità di elaborazione indipendenti; esistenza di collaborazioni organizzate a livello internazionale.

Innumerevoli esempi di progetti che prevedono o stanno già utilizzando tecnologia di Grid per applicazioni di bioinformatica sono disponibili tramite organizzazioni come Healthgrid. Il progetto europeo di e-Infrastructure EGEE e BIOINFOGRID consentirà di supportare un notevole numero di applicazioni di biomedicina e bioinformatica.

Pertanto la bioinformatica è l'area dove le tecnologie Grid possono dare il maggiore contributo. L'obiettivo della bioinformatica è lo sviluppo ed il mantenimento di una infrastruttura per la conservazione, l'accesso, il trasferimento e la simulazione di processi ed informazioni biomediche; gli attuali sforzi di estesa computazione su dati biologici rientrano, quindi, negli obiettivi generali delle tecnologie Grid.

Al fine di favorirne l'uso occorre l'istituzione di fondi nazionali per applicazioni di Grid in questo dominio. Questi fondi dovrebbero avere un ruolo complementare ai fondi messi a disposizione dalla EU in questo campo. Favorire la partecipazione industriale compensando la parte non coperta dalla EU con fondi nazionali ogni qualvolta si identifichi un progetto interessante; creare una commissione nazionale che verifichi e promuova questi progetti.

L'infrastruttura concettuale

Nei prossimi anni analizzando il corredo genetico di ogni individuo, si attendono passi in avanti nella ricerca sul perché alcuni si ammalano, altri sono suscettibili ad ammalarsi o a resistere a una infezione virale o batterica o, ancora, a reagire in un determinato modo a un preciso farmaco.

Questo significa che i ricercatori sono chiamati a lavorare su un insieme vasto ed eterogeneo di dati disponibili in differenti basi di dati accessibili in rete.

Tutto ciò implica la possibilità di vedere tutti questi dati in un unico ambiente, sfruttando al massimo le possibilità offerte dall'interoperabilità.

I motivi sono tanti alla base dello sviluppo armonioso ed interoperabile delle basi di dati, legati principalmente all'esigenza sempre sentita:

- da un lato di una normalizzazione concettuale e coerente di questi Repository (usare la stessa rappresentazione),
- dall'altra di strumenti informatici pensati per consentire le indispensabili Modalità di Scambio e Cooperazione (parlare la stessa lingua).
- In tale situazione, è evidente la necessità di tool informatici in grado di:
- incrociare efficacemente in una Cooperazione Integrata le principali caratteristiche genetiche, epidemiologiche, etniche e sociali per identificare i geni-malattia (unico strumento di raccolta dati),
- attribuire a ciascuna sequenza genica uno score di Rischio Genomico Individuale (analisi statistiche).

Questo implica enormi sforzi nell'integrazione di informazioni cliniche e genomiche distribuite a livello geografico. Inoltre lo sviluppo di vocabolari standard facilita la condivisione di dati e risorse.

Sul piano di questa infrastruttura gli aspetti da studiare riguardano:

- Il modello semantico dei processi associato ad analisi biomediche.
- Il modello di standardizzazione delle biobanche.
- Lo sviluppo dei sistemi informativi sanitari per tener conto delle esigenze specifiche delle applicazioni che trattano dati genetici, genomici e di proteomica.
- Lo sviluppo di Cartelle cliniche elettroniche orientate alla gestione ed all'integrazione di dati clinici con dati genetici, di genomica funzionale e di proteomica.
- Il progetto di sistemi informativi per la conduzione di studi di caratterizzazione genetico-clinica di particolari sottopopolazioni, anche in riferimento agli aspetti di privacy e sicurezza informatica.
- Utilizzo di sistemi formali di workflow per il design e lo sviluppo di applicazioni di supporto al lavoro collaborativi in bioinformatica.

Semantica dei dati

Il primo passo nello studio di un particolare dominio della conoscenza dovrebbe essere la definizione del significato stesso dei dati, delle entità che costituiscono il dominio in questione e delle relazioni che intercorrono tra esse. Questo consiste nella definizione di una "normalizzazione concettuale" delle banche dati, la definizione di ontologie che definiscano le entità (oggetti) e le relazioni tra esse basate su vocabolari condivisi e controllati, la definizione degli schemi delle basi dati derivati dagli schemi concettuali che rappresentano il dominio in esame.

Il risultato di questo lavoro di definizione semantica è la possibilità di definire in maniera chiara e precisa il significato degli oggetti e dei dati che rappresentano l'informazione eliminando alla base gli errori dovuti alla loro sbagliata interpretazione ovvero ad una rappresentazione incoerente tra le varie basi dati. In questo ambito dovranno essere definiti gli standard di rappresentazione dei dati utilizzando il linguaggio di marcatura XML e sulla base di questo gli standard di descrizione dei dati quali BSML, MAGE-ML, ecc.

Protocolli di scambio informazioni

La definizione dei metodi di scambio dell'informazione e di collaborazione tra diversi soggetti rappresenta il secondo passo fondamentale. E' necessaria la definizione di protocolli di comunicazione che permettano la trasmissione dei dati in maniera definita e sicura. Sulla base di questi protocolli sarà quindi possibile definire servizi aggiuntivi quali ad esempio la sincronizzazione delle basi dati, la creazione di indici comuni delle basi dati disponibili e la loro ricerca, la pubblicazione delle caratteristiche delle singole basi dati e le modalità di accesso ed utilizzo, ecc.

Tutti questi servizi devono avere il preciso obiettivo di facilitare la collaborazione e l'integrazione sia tra le basi dati che tra i soggetti che vi operano. Un'analogia nel mondo informatico che ben si adatta a questo aspetto e' rappresentata dall'utilizzo dei "Web Services".

Basandosi su un protocollo di comunicazione ben definito come l'HTTP, alla base del World Wide Web, sono state definite le modalità di pubblicazione, indicizzazione, ricerca, accesso ed utilizzo di servizi disponibili sulla rete Internet.

Sviluppo di standard in Biomedicina e Bioinformatica

In Italia sono da tempo attivi gruppi di riferimento per l'adattamento, allo sviluppo e alla diffusione dello standard Health Level 7 (HL7) nel settore dell'informatica sanitaria.

HL7 è una organizzazione di crescente successo, che da oltre 20 anni produce standard per l'informatica sanitaria. Negli USA è accreditata dall'Ente normatore nazionale (ANSI) e copre il 90% dei maggiori fornitori di sistemi informativi per la sanità. HL7 è stato adottato dall'azione di e-government USA, dal National Program for Information Technology (NPfIT) in Inghilterra, dal Programma federale australiano OpenConnect, dal Programma federale canadese Infoway. Negli ultimi anni sono state create circa 30 sezioni nazionali fuori dagli USA; le più attive sono in Australia, Canada, Gran Bretagna, Germania, Olanda.

Informazioni sulle attività, la struttura organizzativa e i materiali prodotti sono disponibili sul sito ufficiale (www.hl7.org) e per l'Italia (www.hl7italia.it).

Inoltre sono iniziate diverse collaborazioni per favorire la convergenza con altre organizzazioni di standard, quali CEN/TC251, DICOM, CORBA. HL7 – tramite l'ANSI, ente normativo americano – ha richiesto un canale preferenziale per la sottoporre il proprio standard al Comitato Tecnico TC215 dell'ISO.

Fanno parte di HL7 Italia numerose grandi aziende nazionali, piccole e medie imprese, strutture sanitarie, centri di ricerca ed università.

HL7 Italia è consulente del Ministero della Salute nel progetto denominato “Mattoni del SSN” per la corretta progettazione e lo sviluppo del Nuovo Sistema Informativo Sanitario (NSIS), che richiede la costruzione di un linguaggio comune per consentire l'interscambio informativo tra il livello nazionale del NSIS e i sistemi sanitari regionali. HL7 Italia collabora con il Ministro per l'Innovazione e le Tecnologie nel tavolo di lavoro denominato “Sanità Elettronica” coordinato da Innovazione Italia per puntare a costruire nei prossimi anni un moderno sistema sanitario a rete tra tutti i soggetti ed i cittadini, in grado di modificare concretamente il funzionamento della sanità pubblica e migliorare l'importante azione della prevenzione proattiva.

In particolare, HL7 Italia ha un Gruppo di lavoro “Bioinformatica”, attivo per il supporto, la divulgazione e l'applicazione dello standard HL7 per la normalizzazione e condivisione dei dati clinici e genomici delle biobanche.

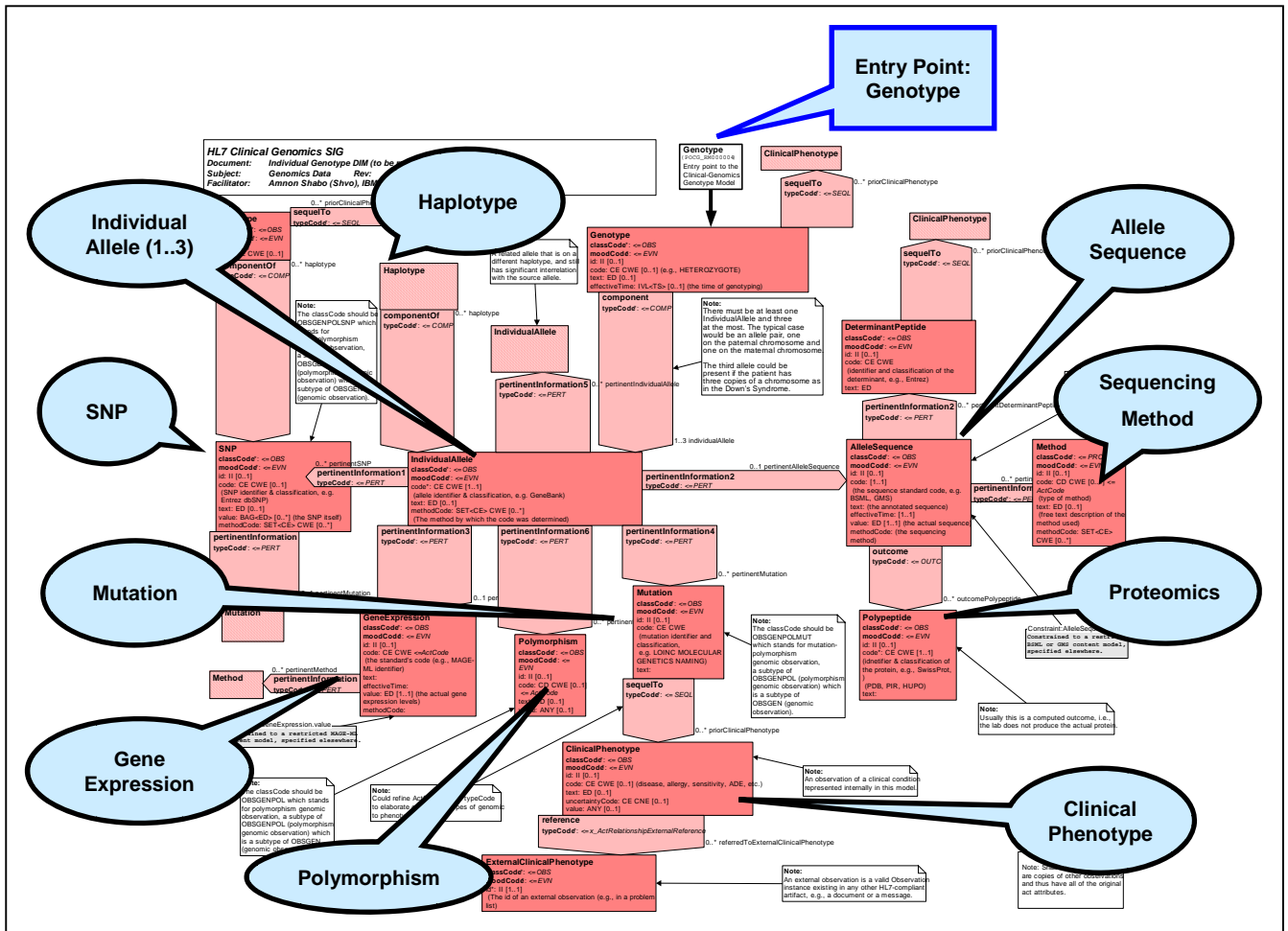
Obiettivo del Gruppo Bioinformatica è, quindi, normalizzare e standardizzare la semantica dei dati bioinformatici e le modalità di condivisione, seguendo i lavori del SIG Clinical Genomics di HL7 ed in generale, per supportare gli strumenti necessari alla cosiddetta medicina personalizzata (personalized healthcare).

Nell'ambito dello sviluppo di standard in bioinformatica, si stanno perseguendo i seguenti obiettivi:

- Creare un Domain Information Model (concetti, vocabolari, tipi di dato, etc.) per il dominio Clinical-Genomics;
- Derivare un R-MIM (modello concettuale informativo da cui costruire lo schema fisico del database della biobanca), ovvero un modello dal RIM (Reference Information Model) di HL7, un meta-modello astratto per rappresentare la semantica delle informazioni sanitarie;
- Specificare le tipologie di messaggi ed i documenti Clinical-Genomic (HL7 CDA di livello 3) per lo scambio e la condivisione di informazioni genetiche tra centri di ricerca;
- Rivisitare/utilizzare i seguenti standard emergenti in bio-informatica (XML) BSML (Bioinformatic Sequence Markup Language), MAGE-ML (Microarray and GeneExpression Markup Language)

L'adozione di tali metodologie consentirà una migliore integrazione fra i dati bioinformatici e i modelli adottati per scopi ospedalieri (dati medici, clinici, amministrativi, etc.);

La soluzione del SIG è estendere talune classi per comprendere i metadati coinvolti nella gnomica e nella proteomica, come illustrato nella figura seguente.



Le biotecnologie nella clinica quotidiana

Il più rivoluzionario orizzonte che la Post-Genomica aprirà nella diagnostica e terapia, e, quindi, nella pratica clinica quotidiana è una cura ed assistenza basata anche sulla identità genetica individuale, sulla conoscenza del rischio genomico, che permetterà di intervenire in modo più appropriato e mirato, con benefici non solo sulla cura ma anche sui sistemi sanitari.

Pertanto l'obiettivo più ambizioso e' un Fascicolo Sanitario Elettronico nella pratica assistenziale (Lifetime Electronic Medical Record), in grado di relazionare patologia e background genetico e supportare, così, protocolli diagnostici e test predittivi innovativi.

La Post-Genomica produrrà nella pratica clinica quotidiana una cura ed assistenza basata anche sulla identità genetica individuale, sulla conoscenza del rischio genomico, che permetterà di intervenire in modo più appropriato e mirato, con benefici non solo sulla cura ma anche sui sistemi sanitari.

Nel piano delle biotecnologie nella clinica quotidiana l'obiettivo è un Fascicolo Sanitario Personale (FaSP) nella pratica assistenziale, in grado di relazionare patologia e background genetico e supportare, così, protocolli diagnostici e test predittivi innovativi.

Il Fascicolo Sanitario Personale oggi rappresenta ogni soluzione capace di mettere a disposizione degli utenti autorizzati, ovunque ed in qualsiasi momento, una raccolta sistematica, tempestiva ed accurata di informazioni cliniche sugli eventi sanitari di cui il cittadino-paziente è stato protagonista. E' evidente che è importante che i dati siano acquisiti attraverso la connessione in rete di sistemi informativi eterogenei e forme efficaci di interoperabilità, ma soprattutto in un modo omogeneo e confrontabile.

L'orientamento è vedere il FaSP come una collezione virtuale di "unità documentali", che contengono informazioni analitiche o sintetiche su un evento sanitario (e.g. prestazione, contatto o episodio di malattia) o sulla storia clinica del cittadino; la cooperazione tra gli operatori sanitari deve necessariamente avvenire con standard generali, di conseguenza le unità documentali sono omogenee, condivise e confrontabili.

In Italia nel marzo del 2004 il Ministero dell'Innovazione, di concerto con il Ministero della Salute, ha annunciato uno studio di fattibilità sulla sanità elettronica. In tal modo si sono individuati alcuni "mattoni", importanti per la realizzazione di questa sanità. Il mattone relativo al Patient File, nell'ambito del Nuovo Sistema Informativo Sanitario (NSIS), ha tra l'altro il compito di analizzare le possibili soluzioni e di predisporre un quadro di riferimento per la diffusione equilibrata dell'e-Sanità in tutte le Regioni nel rispetto delle priorità e specifiche locali.

Il progetto Mattone "Realizzazione del Patient File" ha il compito di produrre nel corso del 2005 un insieme di linee guida concordate a livello sopra-regionale, per sostenere le attività di medio termine e le attività di lungo termine legate allo sviluppo del FaSP e delle tematiche collegate, in continuità con gli investimenti effettuati finora nel settore e con l'autonomia decisionale delle singole Regioni. Le soluzioni parziali verso il FaSP sono basate su principi ed esigenze che mettono l'accento su singoli aspetti del processo di integrazione.

Per raggiungere l'obiettivo di introdurre nel FaSP anche i dati bioinformatici occorre rappresentare tutte le numerose tipologie di dati per la definizione del profilo genomico di un individuo che richiede spesso l'analisi simultanea di molti geni e lo studio della loro attività in modo parallelo (e.g. nell'uomo numerosi geni possono avere decine o centinaia di mutazioni diverse responsabili della stessa malattia). Infatti, lo studio di queste mutazioni in parallelo sottolinea proprio la capacità di fornire informazioni su numerose alterazioni o stato di attività di geni nel medesimo istante. Una volta individuato il modello occorre definire l'unità documentale specifica per i dati genomici.

Questi sviluppi saranno importanti anche per le Reti oncologiche, le Reti di terapia, le Reti di collegamento con il territorio, per scambiare informazioni tipo:

- risultato di un dispositivo di testing DNA,
- report da una unità specialistica Annotazioni di risultati da un laboratorio sul DNA/Proteomica,
- collegamenti tra annotazioni genomiche e dati clinici.

Si creerà così una cooperazione tra strutture sanitarie, anche di eccellenza, il medico di base ed il paziente.

La formazione

L'aspetto dell'educazione in bioinformatica è un altro punto su cui occorre intervenire. Nel piano della formazione gli aspetti da considerare per sviluppare le competenze sono indirizzate verso:

- l'utente - accedere e utilizzare le risorse ed i servizi disponibili in internet,
- lo sviluppatore - contribuire a definire ed implementare nuove risorse e servizi bioinformatici,
- il ricercatore - sviluppare nuove metodologie di analisi bioinformatiche innovative.

Ancora troppo poche sono le possibilità offerte ai giovani per costruirsi una figura professionale nel campo. I gruppi di ricerca certamente educano giovani, attraverso tesi di laurea, dottorati, borse di formazione. Vi sono alcune lauree di secondo livello e Master di primo e secondo livello e corsi di Bioinformatica sono inclusi nella maggior parte dei curriculum di Biotecnologie e, in alcuni casi, di Scienze Biologiche e Scienze dell'Informazione, ma la copertura didattica di tali iniziative avviene attraverso sforzi talvolta eccessivamente gravosi da parte dei docenti che non possono ancora contare sull'assistenza di un numero sufficiente di giovani ma esperti bioinformatici.

Pertanto in Italia è del tutto assente una scuola ed una metodologia che formino bioinformatici e per tale motivo sarebbe utile creare un corso di laurea *ad hoc*; il grande numero di master nel settore, infatti, non sono in grado di rispondere a pieno alle esigenze del mondo della ricerca e dell'industria poiché i corsi forniscono agli studenti solamente le nozioni di base.

L'informatica italiana è per lo più teorica e ciò ostacola la nascita di validi sistemisti a differenza di quanto avviene all'estero, in cui la scuola di ingegneria ha un'impostazione pratica, di sperimentazione.

Un altro aspetto collegato alla formazione ed all'aggiornamento riguarda gli operatori sanitari che dovrebbero essere messi in condizione di utilizzare al meglio i risultati della ricerca nel settore della bioinformatica per migliorare le loro attività professionali quotidiane.

Pertanto l'attività formativa potrebbe essere collegata al sistema dell'ECM Educazione Continua in Medicina programma voluto dal Ministero della Salute.

Sarebbe opportuno realizzare un corso di Bioinformatica Medica per gli operatori sanitari, in particolare biologi e medici di medicina generale in modalità e-learning. I contenuti del corso potrebbero seguire una linea guida quale:

- Conoscenza della Genomica e della Bioinformatica
- Interazioni tra ricerca di base, ricerca applicata e farmacoterapia
- Applicazioni nella pratica clinica
- Bioinformatica e valutazioni economiche-sanitarie.

Conclusioni

Il settore Bioinformatica e le applicazioni nel settore biotecnologico e biomedicale sono in una fase di enorme crescita a livello internazionale. Al fine di pianificare le attività future in questo settore si propone di prevedere un progetto finanziario al fine di una pianificazione degli interventi specifici nel medio lungo periodo per il raggiungimento dei seguenti obiettivi:

- Stabilire un'effettiva cooperazione pubblico-privato fra la comunità di ricercatori in ambito nazionale e internazionale.
- Sviluppare un sistema integrato per le biobanche.
- Consentire un costante aggiornamento delle conoscenze in un settore altamente innovativo come quello della System Biology applicata all'analisi dei dati generati dalla Genomica e Proteomica di vari organismi.
- Stimolare la crescita di nuove imprese nel settore Biomedicale e Biotecnologico.
- Offrire una base strategica per agevolare la formazione interdisciplinare di nuovi profili, bioinformatici, biotecnologici e biomedico.
- Stimolare l'interazione e l'integrazione delle iniziative già finanziate con i laboratori pubblico-privati finanziati dal FIRB, FAR e dalle infrastrutture finanziate dal PON.
- Sviluppare infrastrutture di calcolo adeguate alle esigenze che si stanno profilando nel campo della Biologia, delle Biotecnologie, della Genetica e della Medicina.
- Finanziare l'adesione dell'Italia alle due iniziative internazionali.
- Favorire la omogeneizzazione degli accordi internazionali recentemente stipulati dall'Italia, al fine di agevolare lo scambio di ricercatori (Research HUB), e la creazione di laboratori congiunti (Joint Research LAB).
- Favorire l'interazione dei laboratori pubblico-privati con iniziative per lo sviluppo delle applicazioni di calcolo avanzato nell'ambito per applicazioni bioinformatiche e biomediche in Italia.

Bibliografia

R. Maceratini, F.L. Ricci. *Il Medico On-Line – Manuale di Informatica Medica*. Verduci Editore, Roma, 2000.

F. Martin-Sanchez, V. Maojo e G. Lopez-Campos - 2002 “Integrating Genomics into Health Information Systems”, *Methods of Information in Medicine*.

Progetto europeo di e-Infrastructure EGEE disponibile su www.eu-egee.org.

L:M: Arthur. *Introduzione alla Bioinformatica*. McGraw-Hill Italia, 2004