

BIOINFORMATICA E CALCOLO AD ALTE PRESTAZIONI

Con l'introduzione delle più moderne tecnologie di biologia molecolare si è cominciato a conoscere molto più in dettaglio la struttura del DNA, del RNA e delle proteine. Questa enorme mole di informazioni saranno necessarie per comprendere le malattie genetiche partendo dal Genoma Umano. La Bioinformatica e la *System Biology* contribuiranno alla creazione di nuove metodologie utili per la comprensione dei genomi dei diversi organismi. Su questa base verranno illustrati recenti progressi di alcune applicazioni di Bioinformatica basate sul calcolo ad alte prestazioni e il Grid computing.

1. INTRODUZIONE ALLE APPLICAZIONI DELLA BIOINFORMATICA

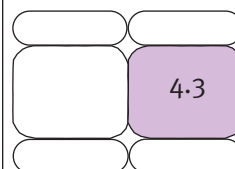
La biologia e la medicina sono la base per studiare l'organismo umano a tutti i livelli, sia morfologico che funzionale [1, 2]. Con i recenti progressi tecnologici nell'ambito delle biotecnologie e della biologia molecolare, è possibile conoscere in pochi giorni il genoma del singolo individuo. Quest'enorme mole di informazioni si sta accumulando in banche dati e, mediante l'utilizzo di appositi programmi di analisi, consentirà la comprensione dell'espressione dei geni, la loro regolazione e le mutazioni nel DNA che sono alla base di malattie genetiche [4]. Per gestire le sequenze genomiche e proteiche e le informazioni che ne descrivono la complessità funzionale dei geni, individuati nel corso dei progetti per il sequenziamento dei genomi di diversi organismi e specie [12, 21, 15], è stata creata una nuova disciplina, denominata "Bioinformatica", che si è recentemente estesa fino ad occuparsi del trattamento dell'informazione biologica e medica a tutti i livelli [5, 6].

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 e fisica allo scopo di migliorare la comprensione dei fenomeni biologici. Di seguito elenchiamo, a titolo esemplificativo, alcuni obiettivi della bioinformatica:

- sviluppo di strumenti per la generazione e il mantenimento dell'informazione proveniente dalle varie fonti: mappa fisica, mappa genetica, mappa cromosomica, mappa citogenetica, polimorfismi, e le informazioni relative alle sequenze genomiche e proteiche ecc.;
- raccolta e organizzazioni delle informazioni genetiche associate alle patologie mediche;
- sviluppo di programmi di calcolo per l'analisi delle sequenze;
- sviluppo di Interfacce grafiche in grado di visualizzare in maniera efficace l'informazione richiesta;



Luciano Milanesi



- sviluppo di metodi software che consentano di agevolare tutte le fasi dei progetti;
- sviluppo di strutture per database specializzate e integrate;
- sviluppo di strumenti informatici, includendo software, 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, dei brevetti e di 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.

Per far fronte a queste sfide, molti metodi sviluppati storicamente in altre discipline (matematica, scienze dell'informazione, fisica ecc.) sono stati applicati per risolvere queste problematiche. Fra le più significative metodologie impiegate per scoprire i principi biologici di base, partendo dai dati contenuti nelle diverse banche dati dalle sequenze genomiche fino alle strutture proteiche, si possono citare: teoria dell'informazione, statistica, intelligenza artificiale, logica, linguistica computazionale, metodi stocastici, teoria dei grafi, sistemi d'apprendimento basati su reti neurali, sistemi esperti ecc..

2. APPLICAZIONI DELLA BIOINFORMATICA PER LO STUDIO DI MALATTIE COMPLESSE

Le malattie complesse, quali le malattie cardiovascolari e le malattie neurodegenerative, sono dovute all'azione combinata di fattori genetici e ambientali. Il numero di geni implicati in queste malattie, il loro effetto e le loro possibili interazioni con i fattori ambientali non sono stati ancora determinati.

Negli ultimi anni la comunità scientifica ha rivolto una particolare attenzione allo studio

delle popolazioni isolate per identificare geni responsabili di malattie complesse. La scelta delle popolazioni isolate si basa sulla possibilità di semplificare il modello genetico alla base di tali malattie. Infatti, le caratteristiche di queste popolazioni quali il ridotto numero di fondatori e l'elevata consanguineità, causerebbero una riduzione del numero di geni implicati nello sviluppo di queste malattie. D'altra parte, l'uniformità dello stile di vita e dell'ambiente, permetterebbero di ridurre la variabilità dei fattori di rischio non genetici.

Recentemente sono stati sviluppati nuovi strumenti bioinformatici per l'identificazione e la ricerca nuovi geni e varianti geniche trovate nella popolazione e ritenute responsabili delle malattie complesse del sistema cardiovascolare e nervoso [14]. Il potere genetico del modello "isolati genetici" deriva dalla possibilità di costruire pedigree estesi che rapportano la popolazione vivente a pochi fondatori in un numero relativamente basso di generazioni. In aggiunta, le caratteristiche di queste popolazioni quali il ridotto numero di fondatori e l'elevata consanguineità, causerebbero una riduzione del numero di geni implicati nello sviluppo di queste malattie. D'altra parte, l'uniformità dello stile di vita e dell'ambiente, consente di ridurre l'interferenza dei fattori di rischio non genetici permettendo la dissezione dei tratti ereditari. Il CNR è particolarmente attivo in questo ambito, al fine di far progredire la ricerca nel settore dell'informatica applicata alla genetica di popolazioni mediante l'utilizzo di metodologie e algoritmi di *data mining* per il supporto alle decisioni e la predizione del rischio. In particolare, sono stati sviluppati nuovi software applicativi e nuovi algoritmi per la gestione, l'analisi e l'implementazione di banche dati di popolazioni, al fine di facilitare l'identificazione dei geni coinvolti nei tratti quantitativi e nelle malattie complesse.

L'analisi dei dati inerenti a questa tipologia di studi si basa su software di analisi quali (SOLAR, SimWalk2, MERLIN ecc.). La piattaforma sviluppata presso il CNR [16] ha consentito l'ottimizzazione degli algoritmi attualmente disponibili, riducendo così notevolmente il tempo di calcolo mediante l'uso dell'infrastruttura ad alte prestazioni HPC (*High Performance Computing*). Queste soluzioni hanno consentito un miglioramento significativo rispetto alle

tecnologie esistenti in termini di velocità di analisi, predisposizione all'automazione, affidabilità e riproducibilità, visualizzazione grafica di ampie genealogie e infine hanno permesso anche di ottenere una migliore accuratezza nel calcolo statistico-matematico.

3. SVILUPPI DELLE APPLICAZIONI PER IL CALCOLO INTENSIVO E IL GRID COMPUTING

Lo studio del DNA genomico abbinato allo sviluppo di strumentazione di sequenziamento massivo di interi genomi in pochi giorni, stanno rivoluzionando la conoscenza dei processi che sono alla base dei profili strutturali e funzionali di molte malattie genetiche e di alcune tipologie di tumori [17]. Infatti, la conoscenza dei processi di replicazione, trascrizione, traduzione e di localizzazione mediano e controllano l'espressione dell'informazione genetica. Lo scopo principale della biologia molecolare è quello di capire i meccanismi che stanno alla base delle capacità di regolazione di questi processi. Diversi metodi sono impiegati per lo studio e la predizione delle funzioni, basandosi sull'analisi dalle sequenze o delle strutture proteiche [18, 19]. Fra questi ricordiamo a titolo d'esempio: la predizione delle regioni codificanti per le proteine; la predizione delle strutture secondarie e terziarie partendo dalla sequenza genomica e/o amminoacidica; la dinamica molecolare; la produzione di modelli molecolare; la simulazione metabolica partendo dalle funzioni biochimiche dell'organismo umano ecc..

Dal punto di vista prettamente bioinformatico ognuno di questi processi è una possibile sorgente di studio e di applicazione di nuove metodologie d'indagine. Uno dei principali obiettivi della Bioinformatica basata sulle tecnologie messe a punto dalle tecniche Informatiche di Grid computing e ICT, è di dare 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 fatto-

ri 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 [20, 22].

4. BIOINFORMATICA IN SUPPORTO ALL'INNOVAZIONE E ALL'INDUSTRIA FARMACEUTICA IN AMBITO EUROPEO

L'industria farmaceutica, sempre di più utilizza i dati, le metodologie e le conoscenze prodotte nei progetti di ricerca nazionali e internazionali al fine di accelerare lo sviluppo di nuove medicine, proteine specifiche e terapie genetiche in grado di intervenire in maniera selettiva per risolvere le possibili cause di malattie.

Da questa intensificazione nella produzione di nuove sequenze, nasce l'esigenza di riconoscere nuove strutture e funzioni dei geni coinvolti nei vari processi metabolici. Per questo è necessario identificare i modelli quantitativi che stanno alla base di questi processi. È quindi diventato essenziale poter dedurre le informazioni utili partendo dall'analisi delle strutture e dalle funzioni a loro associate, utilizzando al meglio le informazioni contenute nelle varie banche dati genetiche, e sia dai metodi e programmi messi a disposizione dalla Bioinformatica.

Nell'ambito del forum europeo ESFRI "European Strategy Forum on Research Infrastructures" sono state promosse iniziative quali ELIXIR per la Bioinformatica e BMMRI per le Biobanche [23]. Questi progetti, in congiunzione con altre iniziative europee, si pongono come obiettivo la possibilità di sviluppare nuove applicazioni, per esempio nei seguenti settori:

- gestione di laboratori biomedici distribuiti;
- calcolo di espressione dei geni su larga scala;
- studio di malattie genetiche basate su dati genotipici e fenotipici;
- analisi filogenetica di diverse specie;
- accesso a database biologici distribuiti;
- analisi funzionale delle proteine e del proteoma di diverse specie;
- simulazioni di dinamica molecolare basate su tecnologie di HPC e Grid computing.

Mediante iniziative come EGEE, DEISA, BioGRID, EDGE, BMMRI [3, 4, 9, 11] e il progetto eu-

ropeo EGI (*European Grid Infrastructure*), per esempio, è stato possibile utilizzare con estrema efficacia i più avanzati centri di calcolo europei basandosi sull'avanzatissima infrastruttura per il calcolo HPC distribuito (DEISA) e su Grid Computing in collaborazione con i maggiori centri di supercalcolo.

In particolare sono state messe a punto nuove applicazioni di "Drug Discovery" per la ricerca di nuovi farmaci nel campo delle Malaria e dell'Influenza Aviaria [10], mediante il calcolo automatico per la ricerca basata sul *Virtual Screening* utilizzando migliaia di computer distribuiti in Europa. In questo progetto è stato totalmente automatizzato il calcolo di milioni di possibili molecole utilizzabili per la creazione di nuove medicine.

5. PROGETTI INTERNAZIONALI

Nel seguito verranno brevemente illustrate alcune iniziative a cui si è attivamente contribuito. La descrizione di altri progetti in ambito europeo è reperibile direttamente dal sito della Comunità Europea.

5.1. *Enabling Grids per E-science* (EGEE-III) - *European Grid Infrastructure* (EGI)

Il Grid computing è diventato indispensabile per una scienza basata su dati intensivi. Il concetto è stato dimostrato attraverso pro-

getti sostenibili come EGEE - *Enabling Grids per E-science* [11] il cui logo è rappresentato nella figura 1. EGEE si propone di sviluppare un *middleware* per una vasta gamma di comunità scientifiche, come l'astrofisica, la chimica computazionale, la terra e le scienze della vita, la fusione e la fisica delle particelle e la Bioinformatica, costruito sulla rete paneuropea GÉANT2. A seguito del progetto EGEE, la Comunità Europea ha promosso recentemente il progetto EGI (*European Grid Infrastructure*) il cui logo è rappresentato nella figura 2 con lo scopo di consentire ai ricercatori presenti in diversi Paesi di collaborare in ricerche comuni basate su infrastrutture Grid per perseguire i seguenti obiettivi:

- espandere e ottimizzare la più grande produzione di Grid in Europa con un continuo funzionamento delle infrastrutture, un sostegno a più comunità di utenti e l'aggiunta di ulteriori risorse di calcolo e dati;

- preparare la migrazione degli attuali Grid da un progetto base ad un modello sostenibile di infrastrutture federato basato su Iniziative di Grid Nazionale (NGI).

EGI fornirà una rete europea coerente e affidabile, assicurando che l'Europa sia in prima linea per la possibilità di utilizzare il calcolo scientifico.

5.2. *Biobanking e Infrastrutture di Ricerca di Risorse Biomolecolari* (BBMRI)

La conservazione e il trattamento di campioni e di dati biomedici sono essenziali per il mondo accademico e l'industria di ricerca al fine di curare e prevenire le malattie umane. La fase preparatoria si basa sul progetto BBMRI - *Biobanking and Biomolecular Resources Research Infrastructure* [9] con i seguenti scopi principali:

- preparare l'integrazione delle biobanche esistenti, mediante l'adozione di componenti innovativi e in linea con gli standard etici, giuridici e sociali a livello europeo secondo lo schema rappresentato nella figura 3;

- aumentare l'efficacia nelle ricerche delle scienze biomediche e nello sviluppo delle risorse biologiche nella sanità pubblica;

- ampliare e garantire la competitività della ricerca europea e dell'industria in un contesto globale;

- sviluppare un quadro finanziario sostenibile.

FIGURA 1

Logo del progetto *Enabling Grids per E-science* (EGEE)



FIGURA 2

Logo del progetto *European Grid Infrastructure* (EGI)



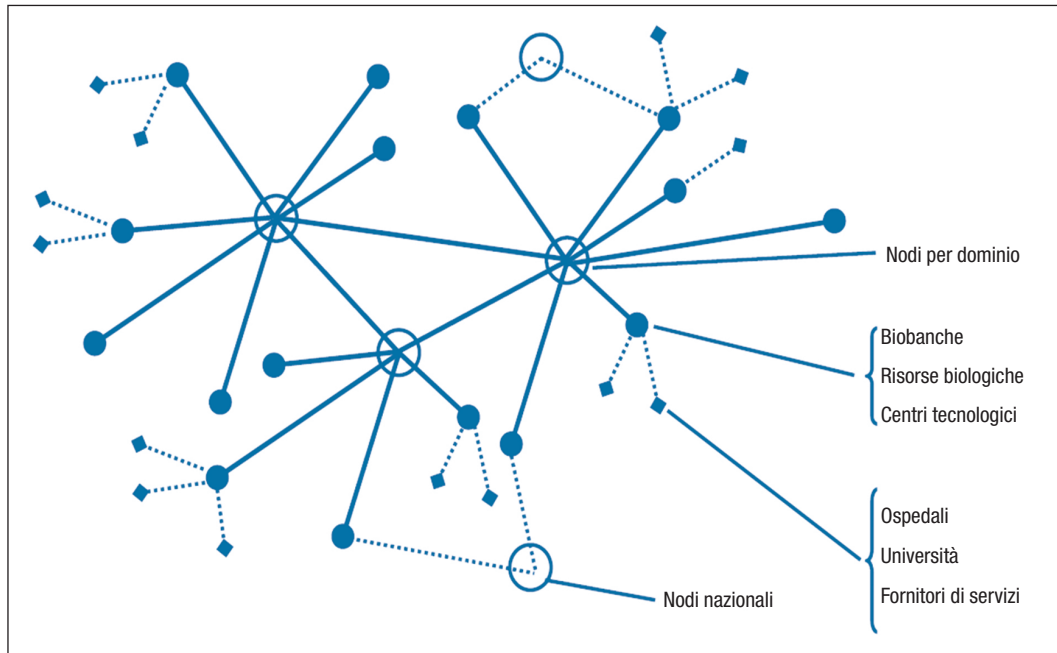


FIGURA 3
Schema di coordinamento e di collegamento del progetto BBMRI fra le Biobanche federate in Europa

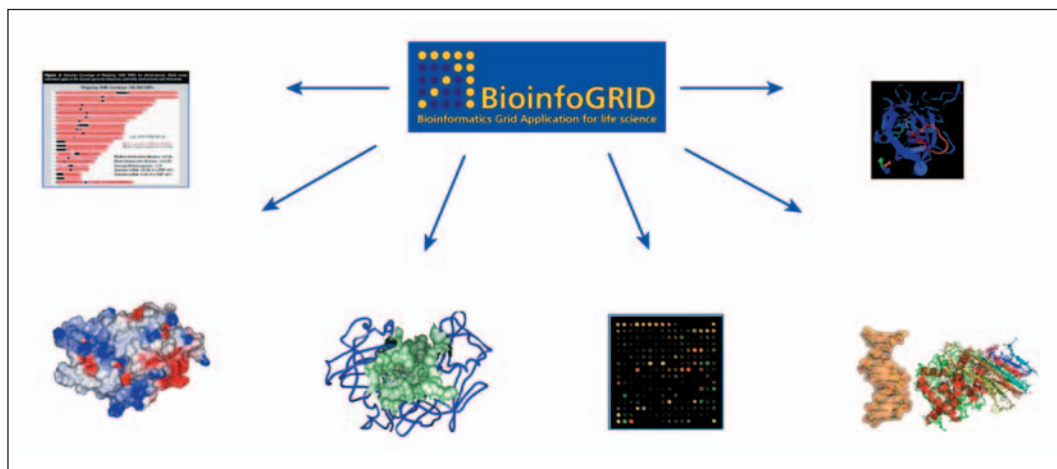


FIGURA 4
Ambiti applicativi del progetto europeo BioinfoGRID per applicazioni in ambito bioinformatico basate su Grid Computing

5.3. Applicazioni GRID di Bioinformatica per vita scientifica (BioinfoGRID)

Il progetto europeo BioinfoGRID (<http://www.bioinfoGRID.eu>) [4] ha contribuito a migliorare i servizi e le applicazioni bioinformatiche per gli utenti di biologia molecolare mediante l'utilizzo delle infrastrutture Grid, sviluppando applicazioni di genomica, proteomica, trascrittomica e dinamica molecolare. Nella figura 4 sono mostrati schematicamente i vari ambiti applicativi della Bioinformatica basata sulla tecnologia Grid.

6. PROGETTI NAZIONALI

L'Italia ha già promosso nell'ambito della Bioinformatica, progetti nazionali finanziati

dal MIUR come LITBIO [13], LIBI e ITALBIO-NET (*Italian Bioinformatics Network*). Queste iniziative hanno consentito di raggiungere risultati in diversi settori tecnologici e metodologici della Bioinformatica e della *System Biology* per lo studio dei sistemi biologici complessi [7, 17].

È importante sottolineare il forte legame internazionale delle iniziative dei progetti nazionali in Bioinformatica.

Come esempio di forte internazionalizzazione, il progetto interdipartimentale "CNR-Bioinformatics" [8] del CNR il cui logo è rappresentato nella figura 5, in congiunzione con il progetto europeo EGEE-III finanziato dalla Commissione Europea e coordinato dal CERN

di Ginevra, ha promosso un'iniziativa di carattere internazionale per realizzare analisi su larga scala finalizzata all'individuazione di nuove molecole a scopo terapeutico.

Mediante questi progetti e in collegamento con l'Accademia Sinica in Taiwan, CNRS-IN2P3 in Francia e il CNR-ITB in Italia, è stata intrapresa un'iniziativa per l'analisi di 300.000 possibili composti farmacologici per combattere il virus H5N1 per l'Influenza Aviaria. Lo scopo principale di questa iniziativa è consistito nell'identificare dei potenziali composti in grado di inibire le attività di un enzima sulla superficie del virus capaci di causare l'influenza, denominato Neuraminidase del sottotipo N1. La figura 6 mostra il posizionamento di un inibitore relativamente all'enzima neuraminidasi del virus dell'aviaria H5N1. Mediante l'uso dell'infrastruttura GRID composta da 50.000 computer messa a punto dal progetto europeo EGEE-III è stato possibile identificare le più promettenti molecole da sottomettere a test biologici al fine di accelerare il processo di sviluppo di farmaci contro il virus influenzale.

L'infrastruttura di calcolo messa a punto da questo progetto ha consentito la riduzione del numero dei possibili farmaci in modo da passare alla seconda fase di laboratorio, per con-

trollarne la tossicità e le altre caratteristiche biochimiche necessarie prima dello sviluppo del farmaco finale. Con questa tecnologia sarà possibile utilizzare con estrema efficacia i più avanzati centri di calcolo europei per la ricerca di nuovi potenziali molecole per la produzione di farmaci per combattere nuovi tipi di virus influenzali e per la ricerca di nuovi farmaci per malattie quali la malaria e l'influenza suina.

7. PROSPETTIVE FUTURE

Solo una piccola frazione del genoma umano è utilizzata per la costruzione delle proteine: si può stimare che il DNA totale codificante sia circa 10^8 basi vale a dire corrisponda al 3% circa del genoma totale. Per la ricerca e la predizione dei geni, sono stati utilizzati algoritmi basati su reti neurali e metodi statistici. Il metodo sperimentale per l'identificazione delle regioni codificanti di una regione del DNA, consiste nel confrontare la sequenza di cDNA e la corrispondente sequenza genomica, mediante un multi-allineamento fra sequenze. Nella maggioranza dei casi la sequenza di cDNA non è nota, risulta quindi necessario sviluppare una serie di metodi per il riconoscimento delle regioni del DNA in cui risiede il gene. Una parte importante di questo processo consiste nel riconoscere i siti di *splicing*, delle regioni codificanti e delle regioni promotrici. È stato recentemente finanziato a livello internazionale un progetto per il sequenziamento di 1000 genomi di persone appartenenti a popolazioni diverse. Questo processo in congiunzione con molti altri studi, metterà in luce le differenze genetiche e le mutazioni puntiformi (SNP *Single Nucleotide Polymorphisms*). Proprio la correlazione delle modifiche funzionali derivate da queste mutazioni caratterizzano il futuro della medicina personalizzata, cioè la possibilità di costruire farmaci specifici e cure per ogni persona che tengano conto delle modifiche del suo genoma. Questo processo appena iniziato vedrà importanti ricadute nel prossimo decennio con un notevole impatto sulla salute dell'uomo e sulla produzione e dosaggio di nuovi farmaci mirati. Attualmente le fasi per la ricerca di nuovi farmaci è estremamente costosa e richiede tempi lunghi. Quando il genoma sarà completamente disponibile,

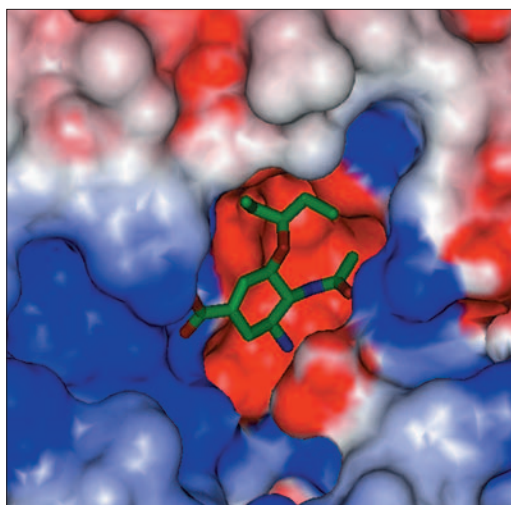
FIGURA 5

Logo del progetto interdipartimentale del CNR di Bioinformatica



FIGURA 6

Interazione tra l'inibitore e l'enzima neuraminidasi del virus dell'aviaria H5N1



sarà possibile studiare l'influenza sulla salute in base all'interazione dei geni fra loro e agli altri fattori ambientali e alimentari. Inoltre, con il progresso dell'ingegneria genetica e della medicina molecolare sarà possibile prevenire e curare alcune malattie ancora prima del loro manifestarsi. Tutto questo porterà lo sviluppo di nuovi strumenti Informatici per l'analisi di una enorme mole di dati, e lo sviluppo di nuove metodologie di indagine. La Bioinformatica e la *System Biology* forniranno la base metodologica alla biologia e alla medicina molecolare per il raggiungimento di tali obiettivi basandosi su piattaforme avanzate per il calcolo ad alte prestazioni HPC e il Grid Computing.

Riconoscimenti

Questo lavoro è basato sui seguenti progetti Europei: INBIOMEDvision, ELIXIR, EGEE-III, BioinfoGRID, BBMRI, EDGE; progetti nazionali: CNR-BIOINFORMATICS, LITBIO (RBLA0332RH), ITALBIONET (RB-PR05ZK2Z_001), GENPOP (RBIN064YAT_003).

Bibliografia

- [1] Alfieri R., Merelli I., Mosca E., Milanesi L. *The Cell Cycle DB: a systems biology approach to cell cycle analysis*. Nucleic Acids Res. 2008 Jan;36 (Database issue):D641-5.
- [2] Alfieri R., Barberis M., Gaglio D., Chiaradonna F., Milanesi L., Vanoni M., Klipp E., Alberghina L.: *Towards a systems biology approach to mammalian cell cycle: modeling the entrance into S phase of quiescent fibroblasts after serum stimulation*. BMC Bioinformatics. 2009 Oct 15;10 Suppl 12:S8.
- [3] Barbera R., Donvito G., Falzone A., La Rocca G., Milanesi L., Maggi G.P., Vicario S.: *The GENIUS Grid Portal and robot certificates: a new tool for e-Science*. BMC Bioinformatics. 2009 Jun 16;10 Suppl 6:S21.
- [4] BIOINFOGRID: *Bioinformatics Grid Application for Life Science*. <http://www.bioinfoGRID.eu>
- [5] Castellani G., Verondini E., Giampieri E., Milanesi L., Bersani F., Zironi I., Remondini D.: Large scale modelling of neuronal systems. *Il Nuovo Cimento*. Vol. 32 C, n. 2, March - April, 2009, p. 9-12.
- [6] Calabria A., Mosca E., Viti F., Merelli I., Milanesi L.: SNPranker: a tool for identification and scoring of SNPs associated to target genes. *J Integr Bioinform*. 2010 Mar 25, Vol. 7, n. 3, doi: 10.2390/biecoll-jib-2010-138.PMID: 20375450.
- [7] Chiappori F., D'Ursi P., Merelli I., Milanesi L., Rovida E.: *In silico saturation mutagenesis and docking screening for the analysis of protein ligand interaction: the Endothelial Protein C Receptor case study*. BMC Bioinformatics. 2009 Oct 15;10 Suppl 12:S8.
- [8] CNR-BIOINFORMATICS: *Progetto interdipartimentale CNR*. <http://www.cnr-bioinformatics.it>
- [9] BBMRI - Biobanking and Biomolecular Resources Research Infrastructure. <http://www.bbMRI.eu>
- [10] D'Ursi P., Chiappori F., Merelli I., Cozzi P., Rovida E., Milanesi L.: Virtual screening pipeline and ligand modelling for H5N1 neuraminidase. *Biochem Biophys Res Commun.*, Jun 12, Vol. 383, n. 4, 2009, p. 445-9. Epub 2009 Apr 14.
- [11] EGEE-III - Enabling Grids for E-science-III (INF-SO-RI-222667). <http://www.eu-egEE.org/> Francesconi M., Remondini D., Neretti N., Sedivy J.M., Cooper L.N., Verondini E., Milanesi L., Castellani G.: *Reconstructing networks of pathways via significance analysis of their intersections*. BMC Bioinformatics, 2008 Apr 25;9 Suppl 4:S9. 8.
- [12] Lazzari B., Caprera A., Cestaro A., Merelli I., Del Corvo M., Fontana P., Milanesi L., Velasco R., Stella A.: *Ontology-oriented retrieval of putative microRNAs in Vitis vinifera via GrapeMiRNA: a web database of de novo predicted grape microRNAs*. BMC Plant Biol. 2009 Jun 29;9:82.
- [13] LITBIO - *Laboratory of Interdisciplinary Technologies in Bioinformatics*. <http://www.litbio.it>
- [14] ITALBIONET - Italian Bioinformatics Network.
- [15] Merelli I., Guffanti A., Fabbri M., Cocito A., Furia L., Grazini U., Bonnal R.J., Milanesi L., McBlane F.: *RSSsite: a reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes*. Nucleic Acids Res. 2010 May 16. doi:10.1093/nar/gkq391.
- [16] Milanesi L., Calabria A., Di Pasquale D., Gnocchi M., Orro A., Trombetti G.: A HPC and Grid enabling framework for genetic linkage analysis of SNPs. *Il Nuovo Cimento*, Vol. 32 C, n. 2, March - April 2009, p. 249-253.
- [17] Milanesi L., Romano P., Castellani G., Remondini D., Lio' P.: *Trends in Biomedical Complex Systems*. BMC Bioinformatics. 2009 Oct 15;10 Suppl 12:S8.
- [18] Mosca E., Bertoli G., Piscitelli E., Reinbold R., Zucchi I., Milanesi L.: *A strategy for the identification of functionally related genes based on data mining and data integration*. BMC Bioinformatics. 2009 Oct 15;10 Suppl 12:S8.
- [19] Mosca E., Bertoli G., Piscitelli E., Reinbold R., Zucchi I., Milanesi L.: *Identification of functionally related genes using data mining and data integration: a breast cancer case study*. BMC Bioinformatics 2009, 10(Suppl 12):S8.

- 0
-
- [20] Orro A., Guffanti G., Salvi E., Macciardi F., Milanesi L.: *SNPLims: A Data Management System for Genome Wide Association Studies*. BMC Bioinformatics 2008 Mar 26;9 Suppl 2:S13.
- [21] Panzitta F., Caprera A., Merelli I., Luciano Milanesi, Williams J.L., Lazzari B., Stella A.: Mining the Bovine Genome with the "Bovine SNP Retriever". *Journal of Heredity*, Nov-Dec, Vol. 99, n. 6, 2008, p. 696-8.
- [22] Torri F., Akelai A., Lupoli S., Sironi M., Amann-Zalcenstein D., Fumagalli M., Dal Fiume C., Ben-Asher E., Kanyas K., Cagliani R., Cozzi P., Trombetti G., Lievers L.S., Salvi E., Orro A., Beckmann J. S., Lancet D., Kohn Y., Milanesi L., Ebstein R. B., Lerer B., Macciardi F.: *Fine mapping of AHI1 as a schizophrenia susceptibility gene: from association to evolutionary evidence*. FASEB J. 2010 Apr 6. PMID: 20371615.
- [23] Viti F., Merelli I., Caprera A., Lazzari B., Stella A., Milanesi L.: *Ontology-based, Tissue MicroArray oriented, image centred tissue bank*. BMC Bioinformatics. 2008 Apr 25;9 Suppl 4:S4.

1

0

1

0

LUCIANO MILANESI è ricercatore dal 1983 presso il CNR - Istituto di Tecnologie Biomediche. Dal 1988 ha partecipato a diversi progetti nazionali ed internazionali volti allo sviluppo di nuovi metodi Bioinformatici per l'analisi del Genoma Umano. Ha partecipato dal 2000 ad oggi a più di 20 progetti di cui 10 europei, fra cui ha coordinato il progetto LITBIO "Laboratorio Interdisciplinare di Bioinformatica" e BIOINFOGRID "Bioinformatics Grid Application for Life Science". È autore di più di 260 pubblicazioni ed è membro dell'editorial board di diverse riviste scientifiche nel campo della bioinformatica, System Biology ed Informatica Medica.
E-mail: luciano.milanesi@itb.cnr.it