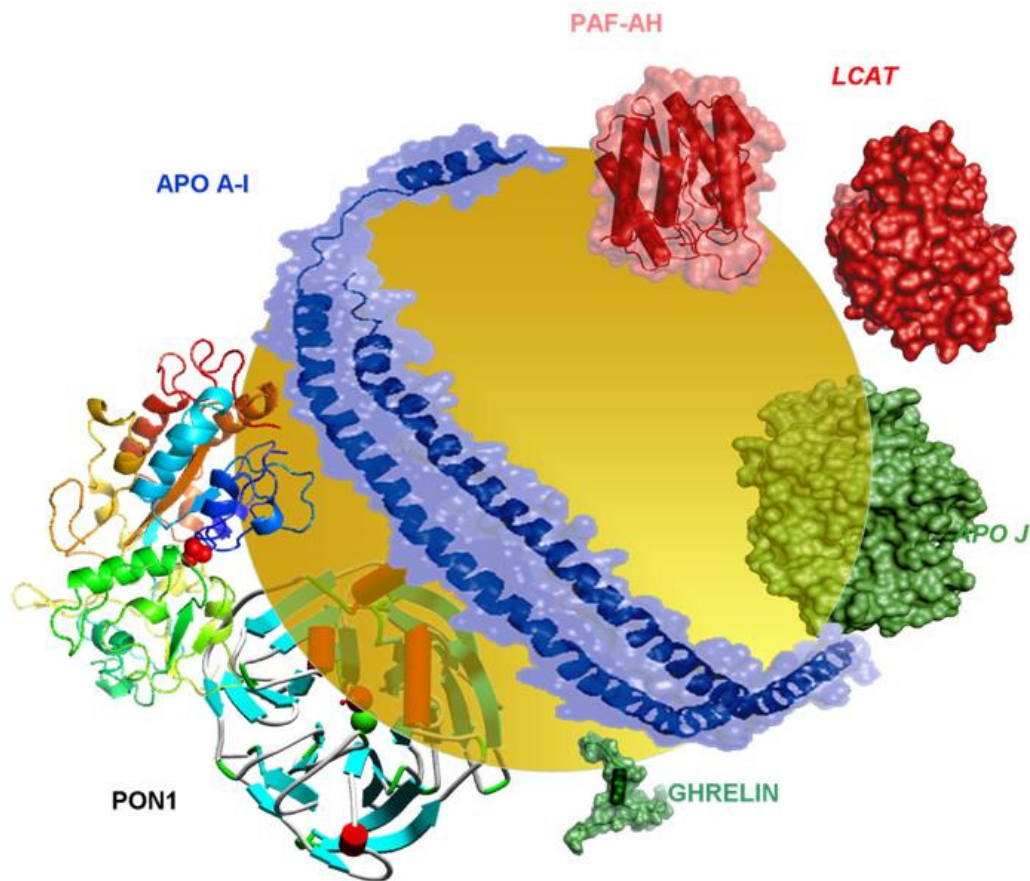


NATURALMENTE

Fatti e trame delle Scienze

I percorsi della scienza nel XXI secolo

Luciano Cozzi



NATURALMENTE

Raccolta degli articoli di Maria Arcà

Spedizione: Poste Italiane SpA - Spedizione in abbonamento postale - D. L. 353/2003 (conv. in L. 27/02/2004 n. 46) art. 1, comma 1, CB PISA

Iscrizione al ROC numero 16383

Direttore responsabile: Luciano Luciani

Segretario di redazione: Enrico Pappalettere
(e.pappalettere@alice.it) 3487934426

Redazione: Sandra Bocelli, Francesca Civile, Brunella Danesi, Fabio Fantini, Fabrizia Gianni, Vincenzo Terreni

Impaginazione: Vincenzo Terreni
(terreni@naturalmentescienza.it)

Edizione e stampa: ETS Piazza Carrara, 16-19 PISA - tel. 050 29544 - fax 050 20158

Proprietà: ANISN - Pisa c/o Museo di Storia naturale e del Territorio, Via Roma, 79 - 56011 Calci (Pi)

Abbonamenti:

Conto Corrente Postale n. 14721567

Banca Intesa - San Paolo

IBAN: IT 95 T 0306914020013958150114

Cassa Risparmio di Lucca, Pisa e Livorno

IBAN: IT 96 A 0620014011000000359148

Ordinario 20,00 euro; ordinario e raccolta TUTTO NATURALMENTE 30,00 euro; ordinario e TUTTO NATURALMENTE pdf 25,00 euro; sostenitore 35,00 euro; Scuole, Associazioni, Musei, Enti ecc. 27,00 euro; biennale 36,00 euro; estero 40,00 euro; singolo numero 8,00 euro; numeri arretrati 12,00 euro; copie saggio su richiesta.

Registrato il 25/02/1989 presso il Tribunale di Pisa al n. 6/89

Informazioni: www.naturalmentescienza.it

050/571060-7213020; fax: 06/233238204

Un ringraziamento particolare alle case editrici

ZANICHELLI e BOVOLENTA

per l'aiuto alla realizzazione di questo numero

Collaboratori

Maria Arcà Centro studi Ac. Nucleici CNR Roma

Maria Bellucci doc. St. Fil. Prato

Claudia Binelli doc. Sc. Nat. Torino

Marcello Buiatti doc. Genetica Università di Firenze

Luciana Bussotti doc. Sc. Nat. Livorno

Stefania Consigliere dip. Antropologia Università di Genova

Luciano Cozzi doc. Sc. Nat. Milano

Tomaso Di Fraia dip. Archeologia Università di Pisa

Elio Fabri doc. Astronomia Università di Pisa

Tiziano Gorini doc. Lettere Livorno

Alessandra Magistrelli doc. Sc. Nat. Roma

Pieggiacomo Pagano ENEA Bologna

Marco Piccolino doc. Fisiologia e Storia della Scienza
Università di Ferrara

Giorgio Porrotto cultore di politica scolastica Roma

Laura Sbrana doc. Lettere Pisa

Marco Tongiorgi doc. Stratigrafia Università di Pisa

Maria Turchetto Dipartimento Filosofia e Beni culturali
Università *Ca' Foscari* di Venezia

I percorsi della scienza nel XXI secolo

Luciano Cozzi

1. Il 2000 parte prima
5. Il 2001 parte seconda
9. Il 2002 parte terza
13. Il 2003 parte quarta
18. Il 2004 parte quinta
22. Il 2005 parte sesta
27. Il 2006 parte settima
31. Il 2007 parte ottava
36. Il 2008 parte nona
41. Il 2009 parte decima
46. Considerazioni conclusive

I percorsi della scienza nel XXI secolo

Il 2000 parte prima

LUCIANO COZZI

Con questo numero prende avvio una serie di rassegne il cui obiettivo è quello di dare un'idea delle tendenze principali delle ricerche nell'ambito delle scienze della vita nel XXI secolo. L'obiettivo dichiarato è piuttosto ambizioso, ma lo strumento per raggiungerlo ha in sé la levità del gioco, più che la seriosità della ricerca. Per ciascuno dei primi dieci anni del secolo ho selezionato gli articoli più citati e tra questi ne ho scelto una decina, cercando di individuare quelli potenzialmente più interessanti per i lettori di *NATURALMENTE*.

In questa prima puntata, per rendere l'idea, il primo articolo ha avuto oltre 8.000 citazioni in pubblicazioni successive, l'ultimo quasi 3.000.

Se la collezione appare più coerente di quanto potrebbe sembrare è forse perché dipende in parte dalle mie scelte, d'altro canto è interessante notare la carenza di pubblicazioni che riguardino alcuni dei temi di cui mi interessano più frequentemente, origine ed evoluzione della vita soprattutto. Forse si tratta di ambiti interessanti ma di nicchia. D'altro canto conforta constatare che un taglio evolutivo compare in modo diffuso in articoli che trattano altri argomenti, dalla medicina alla genomica.

1. Il marchio del cancro

Il primo articolo, il più citato tra quelli scelti, è una rassegna che illustra l'attuale stato delle cose nella ricerca sui meccanismi che consentono di identificare i processi di insorgenza dei tumori. Gli Autori offrono una disamina molto approfondita, affrontando il tema dal punto di vista biochimico, biologico molecolare, citologico. Essi raccolgono e riordinano i risultati di decenni di ricerche che hanno portato a focalizzare l'attenzione su un piccolo numero di tratti distintivi del processo neoplastico, condivisi da gran parte e forse da tutti i tipi di tumore. Quest'ultima caratteristica è forse la più interessante, poiché suggerisce una sostanziale unitarietà nel processo di insorgenza dei tumori.

Uno dei meccanismi studiati con maggiore attenzione è l'acquisizione dell'autosufficienza in termini di segnali di crescita, che rende le cellule cancerose in grado di auto sostenere la propria riproduzione, indipendentemente dai segnali provenienti dalle cellule sane del corpo. La perdita della sensibilità ai segnali di rallentamento della crescita è il naturale completamento di questa capacità, accoppiata alla mancata risposta ai segnali di stimolo dell'apoptosi, il cosiddetto "suicidio cellulare programmato", che dovrebbe scattare quando una cellula assume caratteri anomali. La crescita incontrollata richiede energia e, quindi, un altro pro-

cesso chiave nella crescita tumorale è lo stimolo dell'angiogenesi, che garantisce l'incremento dell'apporto di sangue e quindi di nutrimenti alla massa del tumore, oltre a costituire la premessa alla formazione di metastasi. Quest'ultimo processo è talvolta affiancato dall'invasione attiva dei tessuti circostanti, che contribuisce alla diffusione del tumore e all'aggravarsi del quadro clinico derivante.

Legate a tutte queste manifestazioni, forse alla loro base, vi sono le alterazioni geniche e cromosomiche che consentono alle cellule neoplastiche di differenziare le proprie strategie, percorrendo vie metaboliche alternative. Un cocktail esplosivo, come è ben noto, e che per questo è indispensabile conoscere al meglio.

D. Harahan, R. A. Weinberg *The hallmarks of cancer* CELL 100, 57-70 (2000).

2. L'ontologia dei geni

Il sottotitolo di questo lavoro è intrigante: *uno strumento per l'unificazione della biologia*. Gli autori fanno parte di un'organizzazione, il *Gene Ontology Consortium* (Consorzio per l'ontologia del gene), il cui obiettivo è il conseguimento di un vocabolario controllato e condiviso relativo alla descrizione dei geni, delle proteine e dei processi in cui essi sono coinvolti nelle cellule degli organismi eucariotici. L'obiettivo è fondato sul fatto che le ricerche fin qui condotte hanno evidenziato come molti geni e prodotti genici relativi a funzioni fondamentali degli organismi siano condivisi da tutti gli eucarioti o quasi, il che rende possibile l'adozione di un vocabolario comune per descrivere processi comuni. Il Consorzio, che è raggiungibile al sito web <http://www.geneontology.org/>, ha al momento attivato tre vocabolari, relativi (i) ai processi biologici, (ii) alle funzioni molecolari e (iii) ai componenti cellulari. I tre vocabolari nel complesso raccolgono attualmente 29.282 diversi termini, oltre ad averne individuati ben 1.426 obsoleti e da abbandonare.

È facile sottovalutare l'importanza di un lavoro di questo genere, che potrebbe sembrare una faccenda da compilatori e da informatici, poco legata alla vera ricerca scientifica, ma un tale giudizio non terrebbe in debito conto quanto sia difficoltoso per i ricercatori man
maz:
nel
pess
cont
ifor-
utto
. La
nere
> un
intraicio alla ricerca. ma non e solo per questo che il

...continua...

I percorsi della scienza nel XXI secolo

Il 2001

(parte seconda)

LUCIANO COZZI

Questo articolo è la seconda puntata dedicata a un'analisi degli indirizzi delle scienze biologiche nel primo decennio del nuovo secolo. Un'analisi né rigorosa né oggettiva, ma che si spera possa risultare interessante e divertente, almeno per coloro i quali, come me, sono irrimediabilmente affetti dall'amore per il campo di studi entro il quale lavorano. Nel presentare gli articoli ho cercato di evidenziarne soprattutto le possibili ragioni di interesse per i lettori di NATURALMENTE, senza alcuna pretesa di una trattazione esaustiva né di un impossibile equilibrio nell'estensione delle diverse presentazioni.

In alcuni casi, ho dovuto raggruppare due articoli in un'unica menzione, dato che si tratta di lavori paralleli relativi allo stesso ambito di ricerca. In questi casi, ho basato la classifica non sulla somma delle citazioni, ma soltanto sul più citato dei due lavori.

1. Dati quantitativi in tempo reale grazie alla PCR

La PCR consente di effettuare esperimenti che forniscono in tempo reale risultati quantitativi sul gene in esame.

La quantificazione dei risultati può essere *assoluta*, quando si basa sul numero di copie rilevate, mentre è *relativa*, se saggia la quantità in rapporto a quella di un'altra molecola assunta come riferimento.

Livak e Schmittgen descrivono nel loro articolo un metodo di quantificazione relativa, denominato 2^{-DDC_T} , discutendo gli assunti su cui si basa, le sue applicazioni e i suoi possibili adattamenti.

Al di là degli aspetti tecnici, il modello in esame si basa su un'elaborazione matematica che parte dall'equazione fondamentale che si utilizza per gestire la PCR:

$$X_n = X_0 \cdot (1 + E_X)^n$$

dove X_n è il numero di molecole presenti al ciclo x -esimo, X_0 è il numero di molecole iniziali, E_X è un parametro che indica l'efficienza del processo e n il numero di cicli. Il modello sfrutta un'equazione relativamente semplice, che consente di valutare la quantità di molecole X_0 in rapporto alla quantità della molecola di riferimento. I dettagli di questa elaborazione matematica esulano dal presente articolo e chi fosse interessato può trovarli nella pubblicazione originale, accessibile gratuitamente *online*.

L'approccio del lavoro di Pfaffl è del tutto analogo, ma esso è specificamente dedicato allo studio degli mRNA grazie all'utilizzo della PCR in associazione con la

transcrittasi inversa per ottenere un buon numero di tratti di DNA complementari all'mRNA in esame. Anche Pfaffl propone un metodo matematico che consente la determinazione quantitativa relativa, senza alcuna necessità di una curva di riferimento, come invece occorre nella quantificazione assoluta.

K. J. Livak e T. D. Schmittgen *Analysis of relative gene expression data using real-time quantitative PCR and the DDCT Method* METHODS 25, 402-408, 2001.

M. W. Pfaffl *A new mathematical model for relative quantification in real-time RT-PCR* NUCLEIC ACIDS RES; 29: e 45, 2001.

2. Il Progetto Genoma Umano

Il 2001 è stato l'anno dello HGP, *Human Genome Project*, il cui completamento era stato annunciato alla fine dell'anno nel corso di una celebre conferenza alla quale presero parte anche Bill Clinton e Tony Blair. La comunicazione ufficiale fu data proprio nei due articoli qui richiamati, pubblicati l'uno su NATURE, a cura dei *Natural Institutes of Health* (NIH), e l'altro su Science, a cura della *Celera Genomics*, guidata da Craig Venter.

Quel completamento costituiva soltanto un primo abbozzo, tuttavia: riguardava poco più dell'80% del genoma e conteneva ancora un tasso non trascurabile di incertezze residue. Una percentuale significativa del genoma umano, attorno all'8%, a tutt'oggi non è stata analizzata e difficilmente lo sarà in un prossimo futuro, data la scarsa importanza attribuita alle regioni comprese in questa porzione, per lo più centromeri e telomeri. Ciò nonostante sono ben noti i risultati ottenuti dallo HGP: la riduzione drastica del numero dei geni stimati, da quasi 100 000 a circa 24 000; il fatto che meno dell'1,5% del genoma codifica per proteine; la presenza di un numero di duplicazioni molto più alto di quello riscontrato negli altri mammiferi.

La discussa storia di questo completamento si può intuire confrontando il titolo scelto da Venter, *La sequenza del genoma umano*, con quello che l'NIH ha dato all'articolo di NATURE: *Il sequenziamento e l'analisi iniziali del genoma umano*. Una differenza non marginale, ma che riflette modalità operative radicalmente differenti scelte da:

Il ge
di bz
dere
così

...continua...

ppie
oce-
bps,
ziare

Il 2002 (parte terza)

LUCIANO COZZI

Eccoci al terzo appuntamento di questa rassegna sulle tendenze della ricerca biologica nel nuovo millennio. Siamo già al 2002, eppure per tanti versi pare di essere molto più indietro nel tempo. Alcuni degli articoli qui presentati risentono, infatti, degli anni trascorsi e appaiono superati dai progressi realizzati negli otto anni successivi.

Ciò è evidente per alcuni ambiti di ricerca di punta, come la RNAi o la riprogrammazione di cellule adulte per ottenere cellule con le caratteristiche delle staminali. A maggior ragione la considerazione vale per i lavori che presentano strumenti di ricerca, come gli archivi informatici legati ai siti web. Data la velocità di modificazione di quel che riguarda Internet, il primo articolo della presente rassegna non potrà essere utilizzato come guida per l'utilizzo del data base al quale è dedicato, pur mantenendo intatto il suo interesse per chi voglia comprendere non solo come funziona un dato strumento, ma come si sia arrivati a implementarlo.

La stessa considerazione vale per gran parte delle altre scelte, di questo e degli altri anni: da affezionato studioso dell'evoluzione, sono convinto che per comprendere la natura di una condizione sia di grande importanza, se non indispensabile, ricostruirne la storia. Anche quando si tratti di storia recente.

1. Un archivio web per le strutture delle biomolecole

Nelle precedenti puntate abbiamo già incontrato articoli dedicati alla presentazione e alla descrizione di software dedicati a specifici compiti, ma questo è il primo caso di un articolo dedicato a un sito web.

Il sito in questione è il *Protein Data Bank* (<http://www.rcsb.org/pdb/home/home.do>), il primo al mondo dedicato alla raccolta di dati riguardanti le strutture delle biomolecole. Attualmente esso è un portale che consente l'accesso a una vasta serie di informazioni riguardanti la biologia strutturale molecolare.

L'articolo chiarisce gli obiettivi del PDB, ne descrive il sistema di immagazzinamento dei dati e quello di accesso all'archivio e delinea i possibili sviluppi futuri di questo strumento. In accordo con tali previsioni, dal 2002 a oggi il portale è mutato parecchio, a cominciare dal fatto che è diventato parte di un'organizzazione più ampia, denominata wwPDB (*world wide Protein Data Bank*), raggiungibile all'indirizzo <http://www.wwpdb.org/>.

Questa grande organizzazione deriva da un piccolo archivio creato inizialmente ai *Brookhaven National Laboratories* (BNL), nel lontano 1971. La storia che conduce da quell'impresa iniziale all'attuale PDB è delineata dal continuo aggiornamento degli strumenti utilizzati per la sua implementazione, che hanno portato da poche manciate di dati, accessibili a un ridotto numero di specialisti, a una struttura accessibile per tutti, la cui architettura consente di distinguere i dati immagazzinati in base alle tecniche utilizzate e agli scopi prefissi, così da fornire strumenti diversi ai bioinformatici, agli strutturalisti, ai docenti e studenti di qualsiasi livello ai quali capita di dovere consultare il PDB, che a tutt'oggi (luglio 2010) include i dati che descrivono ben 66828 strutture relative a proteine e acidi nucleici.

Berman H., Westbrook J., Feng Z., Gilliland G., *et al.* *The Protein Data Bank* NUCLEIC ACIDS RES. 28 (2000).

2. L'aterosclerosi, patologia infiammatoria

La seconda pubblicazione è di stampo prettamente medico e riguarda l'aterosclerosi. Un tempo considerata una patologia derivante da una disfunzione nell'accumulo dei lipidi, oggi l'aterosclerosi è spiegata invece con l'attivazione di una risposta infiammatoria, la cui evoluzione è rilevante in ogni stadio della patologia aterosclerotica, dall'aterogenesi che ne costituisce la fase iniziale all'esito finale, spesso consistente in complicazioni che conducono alla trombosi.

Studi precedenti avevano già evidenziato come la formazione dell'ateroma, la placca la cui presenza indica che la patologia è in atto, sia associata a un marcato incremento dell'adesione dei leucociti alle pareti endoteliali e alla loro penetrazione nella tonaca intima.

Lo studio qui presentato ha rilevato in pazienti umani un alto grado di correlazione tra i livelli di proteina C reattiva, un tipico marcatore dei processi infiammatori, e il rischio di insorgenza di aterosclerosi.

Secondo l'ipotesi interpretativa seguita dagli autori della ricerca, l'innescò del processo patologico starebbe nell'ossidazione di alcuni specifici lipidi della tonaca intima, che comporterebbe modificazioni locali inter-

preta
patol
poss
Com
fatto

...continua...

orse
oni,
ro.
ere
rire

I percorsi della scienza nel XXI secolo

Il 2003 (parte quarta)

LUCIANO COZZI

Con l'approssimarsi dell'anno in esame alla data attuale, diviene più difficile individuare articoli che abbiano lasciato un segno profondo e, come già rilevato nella scorsa puntata, è più facile che le citazioni premino lavori specialistici, relativi a tecniche e strumenti di immediata utilità. Questo fatto induce a una riflessione: se sempre più articoli sono dedicati a presentare nuovi strumenti di ricerca, i risultati delle ricerche effettuati con questi strumenti dove sono? Va detto che in alcuni campi di studio gli esiti non mancano. Diversi articoli che incontreremo questa volta riguardano lo sviluppo e il controllo del sistema immunitario, un tema tra quelli più attivamente indagati dalla biologia del nuovo secolo, sia per le ricadute pratiche, in campo medico, sia per l'interesse teorico che esso riveste. In altri ambiti, come nella filogenesi, si ha la sensazione che i ricercatori faticino a tenere dietro ai progressi tecnologici, ma è possibile che si tratti di un errore di prospettiva, dovuto al fatto che le tecniche interessano tutta la comunità, mentre i risultati riguardano solo gli esperti di un settore specifico.

1. Cellule dendritiche e immunità

Questo articolo riprende una rassegna pubblicata diversi anni fa (1998) e la ripropone all'attenzione dei ricercatori. È forse ingiusto attribuirle il primo posto nella nostra classifica, dato che ha avuto diversi anni per accumulare citazioni, tuttavia si tratta di un lavoro di grande interesse medico e biologico e la prima posizione non è fuori luogo.

La rassegna è dedicata alle cellule dendritiche (DC), un tipo di cellule APC (*Antigen Presenting Cells*) note da decenni, ma studiate in modo proficuo soltanto a partire dagli anni '90. Esse sono molto meno note dei linfociti, ma non meno importanti.

La loro sottovalutazione si spiega con l'impossibilità, fino al 1992, di coltivarle *in vitro* e con l'incapacità di reperire marcatori specifici per identificarle *in vivo*. Ciò ha ritardato il riconoscimento del loro ruolo biologico. Un cambiamento decisivo è stato segnato dall'elaborazione di una metodologia affidabile per ottenere cellule dendritiche da monociti umani, così da poterle studiare con maggior agio.

In questo modo è stato possibile fare luce sul ruolo centrale delle DC nel sistema immunitario. Da tempo si sapeva che esse interagiscono con i linfociti T, ma quel che si è constatato è che esse sono indispensabili per l'attivazione di queste cellule, così come per l'attivazione dei linfociti B, seppure attraverso un meccani-

smo indiretto. Sono infatti le DC che catturano e rielaborano gli antigeni, producono molecole co-stimolatorie per i linfociti, e secernono le citochine che danno inizio alla risposta immunitaria.

I ricercatori sono riusciti a chiarire le ragioni, citologiche e molecolari, per cui le DC sono tanto efficaci nell'espore gli antigeni e questo ha consentito di chiarire il rapporto tra le DC cosiddette "mature" e quelle "immature", che costituiscono una riserva di pronto utilizzo proprio per l'esposizione di antigeni. Si è infine scoperto che le DC partecipano attivamente al controllo del processo di inattivazione della risposta dei linfociti T agli antigeni *self* e questo le rende un materiale di studio di particolare interesse per il campo dei trapianti.

Banchereau, J., Steinman, R.M. *Dendritic cells and the control of immunity*, NATURE 392:245-52 (1998).

2. Ancora Mr. Bayes

Mr. Bayes è un software per l'analisi bayesiana di alberi filogenetici. Ne abbiamo già parlato nella rassegna dedicata agli articoli del 2001, anno in cui era stata presentata la prima versione di questo programma. Il presente lavoro è dedicato a Mr. Bayes 3, versione opportunamente aggiornata e potenziata. La novità più rilevante è la possibilità di gestire e integrare dati eterogenei, sia morfologici sia molecolari e, in questo secondo caso, relativi a sequenze nucleotidiche o amminoacidiche o anche a interi genomi. L'aumento del numero di dati e la diversificazione delle fonti ha l'effetto di rallentare il procedere del programma, che si basa su un algoritmo ben noto nel campo, la catena di Markov Monte Carlo. Il rallentamento, tuttavia, non sembra essere di entità superiore a quello previsto dall'aumento di complessità delle schiere di dati.

Il breve articolo qui recensito presenta le caratteristiche di questa nuova versione del programma e finisce per assomigliare più a un opuscolo di presentazione che a un tradizionale lavoro scientifico, tuttavia le oltre 5500 citazioni che ha ricevuto indicano un evidente apprezzamento da parte di chi lavora nel campo.

Ronquist, F., Huelsenbeck, J.P. MrBayes. 3: *Bayesian phylogenetic inference under mixed models*, BIOINFORMATICS; 19:1572-1574 (2003).

3. Mf

Nelle
già in
Intern

...continua...

no
rse
na

Il 2004 (parte quinta)

LUCIANO COZZI

Siamo giunti alla quinta puntata di questa rassegna del primo decennio del nuovo secolo. Anche qui dominano gli strumenti matematici e informatici di analisi dei dati, confermando una tendenza già evidente nelle rassegne precedenti. Non mancano, tuttavia, articoli più tradizionali, come quelli dedicati allo studio dei recettori Toll-like, quello che si interessa della rivoluzione portata nel mondo della biologia molecolare dai micro-RNA e quello dedicato al controllo dell'apoptosi. Nell'insieme, anche il 2004 risulta un anno interessante.

1. *MEGA3*

Un altro software a dominare la classifica, con oltre 8.000 citazioni. Come si evince dal nome, non è la prima apparizione di questo programma, giunto alla terza versione, né, possiamo anticipare, sarà l'unica volta che lo incontreremo. *MEGA* è un software dedicato all'analisi e all'allineamento di sequenze molecolari per studi filogenetici. Il suo nome è infatti l'acronimo per *Molecular Evolutionary Genetics Analysis*, che dichiara esplicitamente la funzione primaria di questo programma, vale a dire quella di analizzare diverse sequenze molecolari per facilitarne la comparazione e per evidenziare eventuali modificazioni evolutive. L'articolo è dedicato in gran parte a illustrare le novità della versione presentata e in particolare i nuovi strumenti offerti ai ricercatori, quali la possibilità di costruire allineamenti automatici o manuali, la guida nel ricercare database sul web dai quali acquisire nuovi dati, strumenti di inferenza statistica e saggi per ipotesi filogenetiche.

Per chi abbia un minimo di familiarità con l'informatica è sorprendente come, anche nel volgere di pochi anni, i software specialistici come *MEGA3* abbiano finito per assumere interfacce facili da riconoscere e da gestire anche per chi specialista non sia affatto. La filosofia dell'*user friendly* dà notevoli risultati sull'aspetto delle schermate. Sulla sostanza il discorso è ovviamente diverso, ma questo esula dall'ambito del presente articolo.

S. Kumar, K. Tamura, M. Nei *MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment* BRIEF BIOINFORM. 5:150-16 (2004)

2. *Coot*: disegnare i cristalli

Anche il secondo articolo presenta un programma informatico, ma questa volta si tratta di uno strumento dedicato ai biologi strutturali. *Coot* è uno strumento dedicato all'elaborazione di mappature grafiche di

molecole, inserito in un pacchetto di software più ampio e potente, chiamato CCP4mg, che consente la risoluzione di analisi strutturali ai raggi X, la comparazione tra strutture diverse e anche la gestione di un'interfaccia grafica di alta qualità.

La cristallografia ha cominciato a utilizzare gli strumenti informatici in modo sempre più diffuso nel corso degli anni '80, ma solo di recente i pacchetti di software dedicati hanno cominciato a includere anche strumenti per la rappresentazione grafica a livello molecolare.

Data la sua filosofia, *Coot* è dotato di funzioni che di solito non si trovano in programmi di grafica molecolare, quali la gestione di coordinate di simmetria, le mappe di densità elettronica e anche strumenti per l'elaborazione di modelli per i rotameri, gli isomeri conformazionali di molecole con legami semplici.

Anche in questo caso, la parte principale dell'articolo è dedicata alla descrizione delle prestazioni del programma e in particolare all'analisi dei casi più interessanti. Interessanti, almeno, per coloro i quali si occupino di questo campo, ai quali lasciamo il compito di approfondire lo studio di *Coot*.

P. Emsley, K. Cowtan *Coot, Model-building tools for molecular graphics* ACTA CRYST. D60, 2126-2132 (2004)

3. L'invasione dei *micro-RNA*

La fine del secolo scorso è stata segnata da una vera e propria rivoluzione, associata alla revisione del ruolo degli RNA nella cellula. In osservanza al dogma centrale della biologia molecolare, la funzione degli RNA è stata sempre considerata quella di fare da mediatori nel passaggio dell'informazione tra il DNA e le proteine, ma la scoperta dei ribozimi, dell'interferenza dell'RNA e di una famiglia, ancora oggi in crescita, di RNA di dimensioni ridotte ha mostrato ai ricercatori che gli RNA erano più versatili e più importanti di quanto si fosse precedentemente sospettato, non solo nella sintesi proteica, ma nella regolazione dell'espressione del DNA e quindi nei processi di sviluppo e di differenziazione, nonché nei fenomeni epigenetici.

La bella rassegna che occupa la terza posizione della nostra rassegna è dedicata a un'analisi della situazione attuale e delle prospettive future. Per chi ne ha bisogno, il testo è disponibile in formato PDF. [...continua...](#)

I percorsi della scienza nel XXI secolo

Il 2005 (parte sesta)

LUCIANO COZZI

Siamo giunti alla metà di questa serie di articoli, il cui obiettivo è quello di individuare le tendenze della ricerca in campo biologico nel primo decennio del nuovo secolo. Lo strumento è quello di proporre una top ten di articoli per ciascun anno.

Il ritmo con cui esce NATURALMENTE è più frequente della cadenza annuale delle puntate, tuttavia, e questo genera un effetto di schiacciamento, poiché, mentre ci avviciniamo ai nostri giorni, le rassegne si fanno più ricche di articoli che accumulano citazioni rapidamente, anche se è possibile che finiscano per essere trascurate con il tempo. Passiamo, insomma, dalla storia recente all'attualità.

La collezione del 2005, comunque, mantiene una discreta coerenza con quelle degli anni precedenti e propone: 4 articoli che, in un modo o nell'altro, riguardano programmi informatici, 2 articoli sul ruolo dei miRNA, 3 articoli di interesse medico, tra i quali anche la rassegna annuale sul cancro in USA, già incontrata negli anni precedenti, e 1 articolo di biologia strutturale. Bioinformatica e RNA con funzioni regolative si confermano quindi tra le linee di ricerca più seguite, ma per fortuna c'è spazio anche per altri tipi di lavori, che contribuiscono a dare agli studi in campo biologico la loro caratteristica varietà. Che, si sa, è un pregio in termini evolutivi.

1. Un programma per l'analisi dell'aplotipo

Abbiamo già incontrato un articolo dedicato allo studio dell'aplotipo umano nella rassegna del 2002. Per aplotipo si intende la combinazione di specifici alleli su un cromosoma o su segmento cromosomico. Lo studio dell'aplotipo è strettamente associato all'individuazione di un disequilibrio di associazione (*Linkage Disequilibrium, LD*), che indica un'associazione statisticamente significativa tra specifici alleli nei genomi di una data popolazione. Il presente articolo, il più citato della rassegna annuale, offre la descrizione di un software dedicato all'individuazione di aplotipi e all'analisi statistica di *linkage disequilibria* a partire da dati genomici grezzi. Il programma, *Haploview*, si può scaricare all'indirizzo web: <http://www.broad.mit.edu/mpg/haploview/>. La sua importanza è legata all'attenzione sempre crescente che l'individuazione e lo studio di aplotipi sta ricevendo nel campo della ricerca medica.

L'articolo offre una breve introduzione e una veloce scorsa delle prestazioni e dell'implementazione del programma. Le ben 4536 citazioni ricevute in sei anni stanno a indicare il successo di *Haploview* tra i ricercatori.

Tale successo deriva dalla versatilità del software e dalla sua compatibilità con molti formati di dati differenti, dovuta al fatto che il programma è scritto in Java. Oltre a ciò, *Haploview* offre una serie di supporti grafici e statistici, che consentono di automatizzare analisi che vanno dal rilevamento di un disequilibrio secondo Hardy-Weinberg ai più comuni test di attendibilità dei risultati.

J. C. Barrett, B. Fry, J. Maller, M. J. Daly *Haploview: analysis and visualization of LD and haplotype maps* *BIOINFORMATICS* 21:263-265 (2005)

2. Il cancro negli USA, anno dopo anno

Per la terza volta consecutiva, il rapporto sullo stato delle malattie tumorali in USA, curato dalla *American Cancer Society*, entra in classifica e quest'anno è addirittura al secondo posto. È appena il caso di sottolineare che questo indica quanto sia importante questo tema per la ricerca, pura o applicata che sia.

I numeri offerti per quest'anno sono: 1 372 910 nuovi casi e 570 280 decessi previsti. Si parla di numeri attesi perché la rilevazione e le registrazioni delle cause di morte non in tutti gli stati sono abbastanza rapidi da consentire di elaborare le statistiche annuali; pertanto il numero dei nuovi casi e il numero dei morti vengono stimati statisticamente grazie all'analisi della correlazione con dati parziali più facilmente reperibili o vengono ricavati grazie all'elaborazione di una curva di relazione con l'andamento negli anni precedenti.

Circa il 25 dei decessi totali negli USA è dovuto al cancro e l'analisi statistica conferma che le malattie tumorali sono la maggior causa di decessi per le persone di età inferiore agli 85 anni, che la percentuale presso i maschi pare essersi stabilizzata, mentre quella presso le femmine è ancora in lieve, ma costante aumento. Le tre forme più comuni di tumore restano quella alla regione bronco-polmonare, quella al colon-retto e quella alla prostata per i maschi, quella al seno e quella colon-rettale per le femmine. Per tutte queste forme, la mortalità si conferma in costante diminuzione.

Resta marcata la differente incidenza tra bianchi e afro-americi, già evidenziata nel rapporto dell'anno precedente. Del resto, la rilevazione annuale, se da un canto consente di avere a disposizione dati sempre

aggiore
eventi
lunghe
tabelle
interessi

...continua...

I percorsi della scienza nel XXI secolo

Il 2006 (parte settima)

LUCIANO COZZI

Prosegue la nostra rassegna, anno per anno, degli articoli più citati di questo inizio di nuovo millennio. Il primo dato che balza all'occhio nell'antologia proposta per il 2006 è la predominanza di argomenti di interesse medico. Ai primi due posti abbiamo infatti due rapporti, uno dedicato all'obesità e l'altro, già noto, al cancro; ma quasi tutti gli articoli presentati qui hanno un evidente interesse in campo medico.

Cancro, malattie autoimmuni, staminali sono gli argomenti più gettonati e si affiancano a escursioni nel campo dell'evoluzione, dei miRNA e del mondo della comunicazione intracellulare (*signaling*).

Un'ultima considerazione: si conferma la tendenza a premiare con citazioni più numerose le rassegne, rispetto agli articoli originali di ricerca. È comodo per chiunque avere fonti che raccolgano tutte le informazioni relative a un determinato ambito di studio.

1. Troppo grassi

Il primo articolo è dedicato a un argomento che stupiva mancasse: il problema dell'obesità, tanto sentito negli Stati Uniti e non solo. Lo studio qui presentato si propone come obiettivo la valutazione del possibile sviluppo di questa patologia e delle sue implicazioni negli anni a venire.

È un lavoro impegnativo, che ha coinvolto un campione di oltre 8000 persone, circa una metà delle quali adulti e l'altra di individui tra i 2 e i 19 anni.

I risultati raccolti sono allarmanti. Nel 2003-2004 oltre il 17% dei bambini e adolescenti statunitensi era sovrappeso e oltre il 32% degli adulti era obeso. L'elaborazione statistica dei dati mostra una significativa tendenza all'aumento di queste percentuali, soprattutto tra i maschi e tra le etnie non ispaniche.

La stima della condizione (normale, a rischio di sovrappeso, sovrappeso, obeso, obeso grave) è stata fatta riferendosi alla tabelle per l'indice di massa corporea fornite dal Centers for Disease Control and Prevention. Per l'elaborazione dei dati sono stati utilizzati due diversi software analitici, i cui risultati sono stati poi incrociati.

L'articolo riporta e analizza una messe di dati dettagliati, la cui visione può interessare soltanto gli addetti ai lavori e che, comunque, portano al quadro generale illustrato.

C. L. Ogden *et al.* *Prevalence of overweight and obesity in the United States 1999-2004.* JAMA.; 295:1549 (2006)

2. Ultime sul cancro

Abbiamo già incontrato un paio di volte questa rassegna annuale. Quest'anno i numeri riportati sono: 1.399.790 casi di nuovi tumori e 564.830 decessi, circa il 25% dei decessi totali negli USA. Come l'anno scorso, tuttavia, si registra una diminuzione nel numero dei decessi, sia per le patologie tumorali in generale, sia per le tre forme più diffuse nei maschi (bronchi e polmoni, colon e retto, prostata) e nelle femmine (seno, colon e retto, considerati separatamente, in questo caso). L'unica eccezione è un lieve incremento dei decessi per tumori polmonari nelle donne, già rilevato negli anni precedenti e confermato in questo 2006.

Si ripresenta anche in questo 2006 la scandalosa differenza tra afroamericani e bianchi. Le donne afroamericane manifestano un'incidenza maggiore del 18% rispetto alle donne bianche, per quanto riguarda i decessi da tumore, i maschi addirittura del 40%. Confermata anche la differenza tra bianchi e altre etnie non afroamericane, che hanno tassi inferiori di mortalità per le forme di tumore riportato, ma che superano i bianchi per quanto riguarda i decessi causati da tumori allo stomaco, al fegato e, per le donne, alla cervice.

Una sezione del rapporto è dedicata alle differenze regionali. Secondo i relatori, le differenze sono dovute a tre fattori: maggiore o minore incidenza dei fattori di rischio, diagnosi e terapie più o meno precoci, diversa affidabilità nella documentazione tra i diversi stati.

La massima variabilità si ha per i tumori polmonari, per i quali, su 100.000 abitanti, si va da un massimo di 42,3 per gli uomini e 21,5 per le donne nello Utah a 138,2 per gli uomini e 72.3 per le donne nel Kentucky. In questo caso, la variazione dei fattori di rischio è quella che ha la massima incidenza: il Kentucky risulta essere uno degli stati con più fumatori e lo Utah quello con meno fumatori. Un'analisi a parte merita anche l'incidenza del cancro nei bambini (1-14 anni), nei quali esso è la seconda causa di morte. Com'era prevedibile, le leucemie sono i tumori con l'incidenza massima nei bambini.

Il rapporto si chiude, di prammatica, invitando a esercitare cautela nel formulare previsioni sugli andamenti futuri and liate erri- utto ione
End sull' toric
...continua...
per l dei d

I percorsi della scienza nel XXI secolo

Il 2007 (parte ottava)

LUCIANO COZZI

In questo 2007 incontreremo parecchi temi già trattati. È il caso di programmi informatici come MEGA4, la cui nuova versione si guadagna il primo posto in classifica. Ci sono poi i ritorni di rassegne annuali, come quella dedicata al cancro negli USA, e i ritorni di argomenti già visti, come la riprogrammazione delle staminali. Mi sono chiesto se non fosse il caso di non considerare di nuovo questo tipo di articoli ripetitivi, ma in ultima analisi mi pare che, se l'obiettivo della mia serie di rassegne è quello di fotografare lo stato delle cose, le ripetizioni debbono rimanere, altrimenti il quadro ne uscirebbe alterato, più vario e innovativo di quanto non sia nella realtà.

1. E MEGA fa 4

Abbiamo già incontrato il software MEGA due volte nel nostro cammino, nel 2001 e nel 2004. Evidentemente si tratta di un prodotto particolarmente apprezzato. Gli autori presentano questa quarta versione, dichiarando che essa amplia e potenzia le possibilità già offerte da quelle precedenti. *Editing* del DNA a partire da archivi presenti sul Web, allineamento manuale o automatico di sequenze, costruzione di alberi filogenetici attraverso la stima delle distanze evolutive sono tra le potenzialità più sfruttate di MEGA. La versione 4 consente di aggiungere didascalie e legende per facilitare la lettura dei modelli elaborati; è un dettaglio, ma di grande utilità pratica. Si tratta, infatti, di un passo importante nella direzione di una maggiore leggibilità e intuitività degli schemi prodotti da questo software, che risulta fruibile da un'utenza sempre più ampia e meno specializzata.

Più consistente è l'implementazione del metodo *Maximum Composite Likelihood* (MCL), che consente di stimare le distanze evolutive, tenendo conto anche dei vincoli relativi alle transizioni (purina - purina e pirimidina - pirimidina) e alle transversioni (purina - pirimidina o viceversa).

MEGA4 nasce come applicazione per Windows a 32 bit, ma è disponibile anche in versione per Linux, comunque gratuita e reperibile in rete.

K. Tamura, J. Dudley, M. Nei, S. Kumar *MEGA4: Molecular Evolutionary Genetics Analysis* (MEGA) Software Version 4.0. *MOL BIOL EVOL.* (2007)

2. *In vino veritas*

Il secondo articolo della rassegna del 2007 porta la firma di alcune decine di ricercatori, seguendo un uso assai diffuso di recente, ma offre al lettore italiano il

conforto di vedere che tra gli autori compaiono diversi connazionali, operanti presso centri di ricerca situati a San Michele all'Adige, Lodi e Padova. Forse non è un caso che la ricerca tratti del sequenziamento del genoma della vite, *Vitis vinifera*, che in Italia ha grande importanza tradizionale ed economica.

Come recita il titolo non si tratta propriamente di un sequenziamento, ma della costruzione di una sequenza consenso di elevata qualità per una varietà eterozigote. Nel linguaggio tecnico dei bioinformatici, una sequenza consenso è la sequenza di nucleotidi più frequente, laddove si confrontino tra loro campioni differenti. Si tratta di un'operazione ben più complessa del sequenziamento di un singolo genoma e al contempo assai più utile, avendo a che fare con una varietà eterozigote, come nel caso in questione, nel quale l'essenza in studio è il clone da cui si ottiene il Pinot nero. Il genoma di *V. vinifera* ha dimensione di 504,6 Mb; gli autori della ricerca l'hanno suddiviso in 2093 raggruppamenti (*metaconting*), connessi a loro volta a 19 gruppi di associazione. Si stima che i geni di *V. vinifera* siano 29.585 e nel presente lavoro ne è stato coinvolto oltre il 96%. L'obiettivo della ricerca era quello di individuare i geni più importanti dal punto di vista pratico. Tra questi vi sono i geni che possono influenzare la qualità del vino, attraverso reazioni metaboliche secondarie, per esempio a carico dei terpeni e dei fenoli, che contribuiscono al bouquet del vino. Non meno importanti sono i geni coinvolti nella resistenza ai patogeni.

Sono stati scoperti circa 2 milioni di polimorfismi da singolo nucleotide, dei quali oltre l'87% è stato mappato e legato a uno specifico gene. Lo studio ha anche evidenziato una serie di recenti duplicazioni cromosomiche, che hanno coinvolto almeno 10 cromosomi, sui 19 del genoma aploide della specie.

Pur fornendo anche alcune indicazioni dirette, come nel caso del metabolismo, il lavoro offre per lo più materiale per ricerche successive e questo, oltre al grande interesse economico della vite, spiega senza dubbio il grande numero di citazioni.

R. Velasco, A. Zharkikh, M. Troglio, D. A. Cartwright, A. Cestaro, D. Pruss, M. Pindo *et al.*, *A high quality draft consensus sequence of the genome of a heterozygous grapevine variety*, *PLoS ONE* 2: e1326, (2007)

3.

Il
all'
co

...continua...

critica
per la

I percorsi della scienza nel XXI secolo

Il 2008 (parte nona)

LUCIANO COZZI

1. Come va la lotta al cancro

Il cancro è la prima causa di morte negli USA, comportando circa il 25% dei decessi totali. Il suo impatto per persone sotto gli 85 anni di età è superiore anche alle patologie cardiache. È senza dubbio per questo che l'annuale rapporto riguardante l'andamento delle malattie neoplastiche si guadagna quasi sempre una buona posizione nelle nostre rassegne, quest'anno addirittura è al primo posto.

Per il 2008 le attese sono state di 1.437.180 nuovi casi e 565.650 decessi. Ambedue le stime sono inferiori a quelle formulate nel 2007. Rispetto al 2004, preso come raffronto, il decremento riguardo alla mortalità risulta del 18,4% negli uomini e del 10,5% nelle donne. Come sempre, il rapporto offre un'attenta descrizione dei metodi di raccolta e di elaborazione dei dati grezzi, nonché un'ampia serie di valutazioni per i singoli tipi di tumori, la cui casistica è suddivisa per età, per sesso, per regione di provenienza e per gruppo etnico di appartenenza. Le patologie tumorali più frequenti restano il cancro alla prostata per gli uomini e quello al seno per le donne; per ambedue i sessi, seguono bronchi e polmoni e poi colon e retto. Date le differenze nella prontezza della diagnosi e nell'efficacia delle terapie, le forme a carico di bronchi e polmoni sono le più letali in ambedue i sessi, 31% negli uomini, solo il 26% nelle donne, in pieno accordo con la minore diffusione del fumo tra le donne.

Come nei rapporti degli anni precedenti, la comparazione tra diversi gruppi etnici risulta impossibile a causa delle troppe fonti di variabilità, con l'unica, eclatante eccezione del gap tra bianchi e afro-americani, che permane scandalosamente alto, sia in termini di incidenza sia, ancor di più, in termini di sopravvivenza.

A. Jemal, R. Siegel, E. Ward, *et al.* *Cancer statistics 2008* CA CANCER J. CLIN.; 58:71-96 (2008)

2. Un software per comparare storie evolutive

È possibile comparare le storie evolutive di specie simili? Il lavoro qui presentato afferma di sì, sempre che si possiedano abbastanza dati.

La gestione di tali dati è il compito per il quale è stato creato Mesquite, un software disponibile in rete, all'indirizzo: <http://mesquiteproject.org>, dove si trovano anche parecchi materiali di supporto all'uso del programma. L'articolo qui presentato è di fatto un manuale per l'uso di Mesquite.

Se consideriamo specie che occupino nicchie simili e che abbiano capacità di dispersione comparabili, do-

vremmo attenderci che le loro storie evolutive siano state influenzate in modo simile da eventi storici che abbiano modificato il territorio.

Nella realtà le risposte che specie diverse mostrano a un certo tipo di modificazioni biogeografiche vanno da un estremo al quale tutte le specie mostrano la stessa risposta, all'altro, condizione in cui le risposte sono del tutto indipendenti. Tra queste due condizioni limite, vi è una gradazione continua di condizioni intermedie.

Evidenziare dove si colloca la storia parallela di due specie in certe condizioni evolutive può aiutare a rilevare cambiamenti nell'entità dell'impatto del clima in diverse epoche, sostengono gli autori di questo articolo e, per dimostrarlo, utilizzano il loro programma su un caso pratico.

La scelta è caduta su due specie di collemboli, saproxiliche (che si nutrono di legno in decomposizione), diffuse in una regione relativamente isolata della Grande Catena Divisoria, che percorre da nord a sud l'intero continente australiano, presso la costa orientale.

Come è prevedibile, i risultati ottenuti confermano la bontà del software.

W. P. Maddison, D. R. Maddison *Mesquite: a modular system for evolutionary analysis* Version 1.0. <http://mesquiteproject.org> (2008)

3. I biocombustibili, promessa o rischio?

Al terzo posto, ho scelto di mettere una coppia di articoli riguardanti lo stesso tema, quello dei biocombustibili, che vengono di solito identificati con l'etanolo ricavato dalla fermentazione del mais, in modo riduttivo, come vedremo.

Nati come promessa di una risposta eco sostenibile alla necessità di carburanti, hanno presto suscitato perplessità e poi aspre critiche, vuoi ecologiche, vuoi politico-economiche.

Searchinger e collaboratori, autori del primo dei due articoli, offrono un'analisi impietosa della situazione attuale. Molti studi, essi argomentano, hanno evidenziato l'impatto positivo che il passaggio all'uso dei biocombustibili avrebbe sull'effetto serra, ma questi studi hanno focalizzato la loro attenzione solo sull'ultima parte del processo globale, scordando di tenere conto di un fattore fondamentale: se i biocombustibili si diffonderanno

diventano un problema globale
...continua... un problema globale

I percorsi della scienza nel XXI secolo

Il 2009 (parte decima)

LUCIANO COZZI

Siamo quasi al termine di questa carrellata sulla ricerca in campo biologico e dintorni nel primo decennio del nuovo secolo. Anzi, a rigore la decima puntata dovrebbe sancire la fine della serie, ma il mio amore per le cifre tonde mi spinge oltre e seguirà un'undicesima puntata. Anche in questa occasione dominano i ritorni di lavori, o almeno di argomenti, già incontrati. Cancro, malattie cardiovascolari, genomica, RNA, bioinformatica. Fa eccezione la pubblicazione all'ottavo posto, che si occupa della pandemia di H1N1, la famosa influenza suina.

Anche in questa puntata mi concederò di essere più rapido sulle presenze costanti nella classifica annuale e di dedicare maggiore attenzione ai lavori più innovativi o dalle ricadute particolarmente importanti.

1. Ancora il cancro al primo posto

Abbiamo già incontrato diverse volte le previsioni annuali che la *American Cancer Society* fornisce riguardo al numero di nuovi casi e di morti attese per patologie tumorali negli USA, basandosi sulle stime di enti federali. Per il 2009 si parla di 1 479 350 nuovi casi, un dato che conferma la sostanziale stabilizzazione del fenomeno nell'ultimo decennio, soprattutto per gli uomini, -1,8% all'anno, contro il solo -0,6% annuo per le donne.

Gli organi più colpiti restano polmoni, prostata e intestino crasso (colon-retto) per gli uomini, seno e colon-retto per le donne.

I decessi attesi per il 2009 erano 562 340, con un significativo -19,2% dal 1990 in poi, sostenuto soprattutto dal miglioramento per quanto riguarda il tumore al polmone (-37%), laddove il colon-retto resta il più resistente al miglioramento degli interventi terapeutici (soltanto -17%).

Come nei rapporti precedenti, anche in questo caso le analisi scorporate per sesso, età e razza fanno emergere soprattutto una vistosa disparità tra bianchi e afroamericani. Tutte le differenze già emerse sono del resto confermate, per esempio quelle legate alla variazione di incidenze e di mortalità nei diversi stati.

A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, M. J. Thun *Cancer statistics 2009 CA CANCER J. CLIN.* 2009

2. Correlazioni tra individui e tra popolazioni

Quando in statistica si studia una correlazione, comunemente i soggetti sui quali si opera lo studio sono individui. Per esempio, in medicina gli studi su nuovi farmaci o le analisi epidemiologiche, come quelle offer-

te dalla precedente rassegna relativa al cancro, sono basati su dati inerenti a persone. In questi classici casi, i parametri che sono elaborati, media e frequenza per esempio, sono da intendersi come attribuiti alla singola persona.

Una correlazione si dice ecologica quando gli oggetti ai quali si applica sono gruppi di persone (in generale di individui, non necessariamente umani). In questi casi, i parametri statistici elaborati indicano proprietà medie delle diverse popolazioni e non degli individui.

Se prendiamo ancora come esempio la pubblicazione della *American Cancer Society*, una correlazione ecologica consentirebbe di studiare le variazioni da stato a stato delle differenze di incidenza tra bianchi e afroamericani, che sarebbero trattati come singoli soggetti, ma sono popolazioni (i bianchi del Kentucky o della California, gli afroamericani della Louisiana o dell'Ohio).

Pur essendo spesso usata con vantaggio in studi sociologici o epidemiologici, la correlazione ecologica non aveva mai ricevuto una definizione e una trattazione formale e colmare tale lacuna è proprio l'obiettivo del presente lavoro, che parte quindi da una trattazione basilare in termini matematici.

Operare sulle differenze tra gruppi o tra individui non dovrebbe generare grandi differenze, ma il modo in cui i dati sono elaborati per arrivare alla correlazione individuale o a quella ecologica differiscono sensibilmente. Rimando all'articolo originale chi sia interessato alla natura di tali differenze.

Quel che vale la pena sottolineare è che l'autore giunge a confutare l'intercambiabilità dei due parametri e a proporre quindi di rigettare ogni inferenza fatta sui singoli, partendo da studi basati sulla correlazione ecologica, e viceversa.

Nel suo ambito, si tratta di un'affermazione ricca di conseguenze, poiché costringe a rimettere in discussione una pratica piuttosto diffusa e a rivedere dati considerati acquisiti.

W. S. Robinson *Ecological Correlations and the Behavior of Individuals* INT. J. EPIDEMIOL. 38 (2): 337-341, 2009

3. Infarto e malattie cardiovascolari: come vanno le cose

Dopo i
morte r
cardiov:
esiste u
Associat

li
e
e
t
a

...continua...

I percorsi della scienza nel XXI secolo

Considerazioni conclusive

LUCIANO COZZI

Siamo giunti alla fine di questa lunga carrellata, quasi un piano sequenza, sulle tendenze delle ricerche in ambito biologico nel primo decennio degli anni 2000.

Avevo cominciando dichiarando che il mio lavoro voleva avere “la levità del gioco, più che la seriosità della ricerca”, ma alla fine mi sono lasciato prendere la mano e, pur mantenendo presupposti poco rigorosi, le mie rassegne hanno assunto un tono fin troppo professionale.

Rivediamo brevemente questi presupposti.

In primo luogo era lasciato al mio giudizio quali argomenti potessero essere considerati di pertinenza biologica e quali no. Si tratta di una scelta che non può avere alcuna pretesa di oggettività, anche se in molti casi appare ovvia. Certo il Progetto Genoma è faccenda dei biologi, ma dove sta il confine di separazione tra la biologia e la chimica? E tra la biologia e la medicina?

In secondo luogo, il numero di citazioni, che pure è un parametro molto usato per sostenere il prestigio di una pubblicazione, è anche molto discutibile e discusso. Se un autore di prestigio incappa in un passo falso e pubblica una sciocchezza, il numero delle citazioni sarà probabilmente molto alto, ma le citazioni saranno negative e non contribuiranno al prestigio dell'articolo né del suo autore. Un articolo del genere dovrebbe essere menzionato nelle mie classifiche, ma la cosa avrebbe poco senso, poiché pur avendo eco, è difficile sostenere che indichi una tendenza nella ricerca. Si tratta di un caso paradossale, ma nemmeno poi tanto.

Il saggio di Michael Behe *Darwin's black box: The biochemical challenge to evolution*, è considerata da molti un caposaldo dell'*Intelligent Design*. Dal 1996, quando è stato pubblicato, ha ricevuto quasi 1500 citazioni. Si può ammettere che ha sollevato una vigorosa discussione, ma questo non conduce a sostenere che l'*ID* sia una vivace tendenza della ricerca contemporanea.

L'ultima precisazione è relativa allo strumento. Il computo delle citazioni è quello consentito da Google Scholar (<http://scholar.google.it/>), uno strumento potente e versatile, ma non esente da pecche, a quanto pare.

Alcuni articoli che sono comparsi nelle mie classifiche annuali risultano avere nel consuntivo molte meno citazioni di quante ne avessero al tempo.

È il caso di quattro articoli, due rassegne annuali *Cancer Statistics*, quelle del 2003 e del 2004, un articolo sul sequenziamento del genoma di *Vitis vinifera*, del 2007, e una rassegna dello stesso anno che proponeva una

critica all'utilizzo del mtDNA per la costruzione di alberi filogeografici. Non sono in grado di stabilire che cosa sia accaduto, ma mi pare evidente che ci sia o ci sia stata una qualche irregolarità nel conteggio delle citazioni.

Date le premesse, è evidente che le considerazioni che si possono formulare a partire dall'elenco degli articoli fin qui citati non possono essere spacciate come un'analisi rigorosa e obiettiva. È tuttavia mia speranza che consentano di formulare qualche riflessione interessante.

I 10 Number Ones

Cominciamo con la lista dei 10 Numeri Uno, l'elenco degli articoli più citati anno per anno. In effetti i titoli sono 11, perché nel 2001 al primo posto si erano piazzati due articoli associati.

Uno dei due, da solo, ha oltre 8000 citazioni in più del secondo in classifica, quindi il primo posto è indubbiamente meritato. L'articolo associato non è andato così bene, ma è stato più citato di diversi numeri uno, quindi anch'esso si merita la sua menzione.

La Tabella 1 riporta l'elenco completo, dal quale si evince che due argomenti ritornano due volte ciascuno (Il software MEGA e le statistiche annuali sull'andamento del cancro negli USA).

Si può rilevare anche una spiccata preponderanza degli articoli relativi alla presentazione di metodi di analisi matematiche, statistiche, o di laboratorio.

Questa tendenza era emersa con grande evidenza già nel corso degli articoli annuali e merita una riflessione. Per quale ragione i lavori che descrivono metodologie sono spesso più citati di quelli che propongono i risultati di ricerche originali? Una possibilità è che le tecniche siano trasversali a diversi ambiti di ricerca e quindi siano più spesso citate di ciascuno dei lavori nei quali trovano impiego. È una prospettiva interessante e confortante. La mia idea è tuttavia un'altra. Gran parte dei lavori pubblicati dalle riviste specialistiche sono relativi ad applicazioni di indubbio interesse pratico, ma di relativo spessore teorico. Per questa ragione gli articoli relativi alle tecniche utilizzate trovano più spazio delle applicazioni stesse. In un certo senso, le

alle rassi
come le
Se si acc
Tabella

...continua...