

La letteratura scientifica e l'uso di PubMed

Molti dati e informazioni sulle conoscenze acquisite sulla biologia molecolare e cellulare, sui dati biomedici e sulla bioinformatica sono raccolti e organizzati in libri, per esempio nei libri di testo universitari.

Scrivere un libro spesso richiede tempi lunghi ed è difficile che nei libri si trattino anche le ultime scoperte o che vi siano informazioni sulle tecniche più recenti. Quindi, ogni volta che si vuole affrontare un determinato argomento nell'ambito della ricerca biologica, un buon testo è sempre necessario, ma costituisce solo un punto di partenza. Un secondo passo per approfondire l'argomento di interesse consiste nel cercare degli articoli *di rassegna* (anche detti *review*). Le review sono in genere scritte da scienziati e ricercatori di grande esperienza e spessore che lavorano in quel campo specifico e che ne conoscono approfonditamente la letteratura passata e recente. Le review offrono una panoramica ampia del campo di interesse, e sono in genere ricche delle necessarie *referenze*, ovvero delle informazioni per accedere ai singoli lavori scientifici che, col loro contributo, hanno portato a costruire la comprensione e le conoscenze riportate nella review (**Figura 1**). Il livello successivo di aggiornamento e di dettaglio, comunque, si raggiunge attraverso l'accesso ai singoli lavori scientifici sull'argomento di interesse (**Figura 2**), che possono essere stati pubblicati dopo la review.

Le informazioni necessarie per identificare questo lavoro sono quelle che si ritrovano nelle referenze dei libri, delle review e nei lavori stessi, e in genere si presentano in questa forma:

Wu CC, Beird HC, Zhang J, Futreal PA. (2018) FusionPathway: Prediction of pathways and therapeutic targets associated with gene fusions in cancer. PLoS Comput Biol. 14(7)

Cominciano con l'elenco degli autori (cognome e iniziali del nome), e riportano poi l'anno di pubblicazione, il titolo, la rivista (numero del fascicolo e pagina di inizio del lavoro). Per un ulteriore livello di aggiornamento può essere utile consultare anche il portale bioRxiv, che colleziona *preprint*, ovvero manoscritti non ancora accettati per la pubblicazione.



libri di testo

articoli di rassegna

Figura 1

I libri di testo riportano le informazioni generali e fondamentali, a livelli che dipendono dalle scelte degli autori e degli editori. In genere le conoscenze riportate sono aggiornate a qualche mese prima della pubblicazione del libro. I libri di testo in genere si rivolgono a un ben preciso tipo di lettore (per esempio lo studente universitario o di scuola superiore) e non riportano dati sulle scoperte più recenti. Gli articoli di rassegna, invece, in genere si rivolgono a lettori professionalmente coinvolti nella ricerca e quindi possono essere usati per approfondimenti e come aggiornamenti.

The image shows a screenshot of a scientific article page from PLOS Computational Biology. Red boxes and arrows highlight specific parts of the page:

- nome della rivista**: Points to the PLOS Computational Biology logo at the top left.
- titolo del lavoro**: Points to the article title: "FusionPathway: Prediction of pathways and therapeutic targets associated with gene fusions in cancer".
- autori del lavoro**: Points to the author list: "Chia-Chin Wu*, Hannah C. Beird, Jianhua Zhang, P. Andrew Futreal".
- affiliazione degli autori**: Points to the affiliation: "Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America".
- abstract**: Points to the abstract text, which begins with "Numerous gene fusions have been uncovered across multiple cancer types. Although the ability to target several of these fusions has led to the development of some successful anti-cancer drugs, most of them are not druggable..."

Other visible elements on the page include a "Check for updates" button, an "OPEN ACCESS" section with citation information, and an "Author summary" section.

Figura 2

Un ottimo livello di dettaglio si raggiunge nei singoli lavori scientifici. Qui vediamo la prima pagina di un lavoro che è stato pubblicato sulla rivista *PLoS Computational Biology*. È intitolato "FusionPathway: Prediction of pathways and therapeutic targets associated with gene fusions in cancer", gli autori sono Chia-Chin Wu e altri, ed è stato pubblicato il 24 luglio 2018.

Banche dati di letteratura scientifica biomedica

PubMed è la principale risorsa per la letteratura di interesse biomedico. È una banca dati sviluppata e mantenuta all'NCBI (*National Center for Biotechnology Information*), presso l'NIH (*National Institute of Health*) negli USA. Comprende più di 28 milioni di citazioni di letteratura biomedica, oltre a diversi libri scientifici scaricabili gratuitamente. PubMed comprende le referenze dettagliate e l'*abstract* di ogni lavoro, ma non contiene gli articoli scientifici completi. Per ognuna delle referenze è disponibile il *link* al lavoro completo sul sito della rivista che lo ha pubblicato. Non tutte le riviste sono incluse in PubMed: esiste un processo di valutazione e selezione delle riviste che garantiscono una migliore qualità della pubblicazione, che di conseguenza si dicono *indicizzate* in PubMed. A ogni lavoro che viene pubblicato nelle riviste indicizzate in PubMed viene assegnato un codice identificativo unico, detto PMID (PubMed Identifier). Una parte di PubMed, detta **PubMed Central**, è dedicata agli articoli scientifici disponibili gratuitamente, cioè pubblicati con la politica dell'*open access* (vedi oltre). **Europe PMC** è un deposito europeo di letteratura biomedica e altri dati. Contiene circa 35 milioni di *abstract* (compresi i circa 28 milioni di PubMed) e quasi 5 milioni di articoli completi.

The screenshot shows the PubMed search interface. At the top, the search bar contains the text "sh2 domain". Below the search bar, there are options for "Format: Summary", "Sort by: Most Recent", and "Per page: 20". A red box highlights the search term "sh2 domain" in the search bar, and another red box highlights the word "parole-chiave" in the "Sort by" dropdown menu. The main content area displays "Best matches for sh2 domain:" followed by a list of search results. The first result is "SH2-catalytic domain linker heterogeneity influences allosteric coupling across the SFK family, Register AC et al. Biochemistry. (2014)". Other results include "Preparation of crystals for characterizing the Grb7 SH2 domain before and after complex formation with a bicyclic peptide antagonist" and "The SH2 domain is crucial for function of Fyn in neuronal migration and cortical lamination". The "Search results" section shows "Items: 1 to 20 of 9171". On the right side, there are filters for "Sort by:" (Best match, Most recent), "Results by year" (a bar chart), and "PMC Images search for sh2 domain" (a grid of image thumbnails).

Figura 3 - Ricerca generica in PubMed

Ecco il risultato di una ricerca effettuata con le parole-chiave «SH2 domain» (= dominio SH2). Viene offerta la possibilità di scegliere tra risultati elencati in ordine di pubblicazione (dal più recente al meno recente) oppure secondo un criterio di importanza (dal più al meno importante). Nel riquadro in alto vengono messi in evidenza alcuni lavori particolarmente importanti sull'argomento richiesto. Il motore di ricerca ha identificato 9171 lavori in cui nel titolo o nell'abstract si trova la parola-chiave cercata.

Ricerca generica in PubMed

Come in altre banche dati (per esempio UniProt, vedi paragrafo 12.3 del libro), le ricerche possono essere effettuate in modo generico o avanzato, specificando i campi in cui ricercare le parole-chiave specificate.

Per usare la ricerca generica, si possono scrivere le parole-chiave nell'apposito campo di ricerca (**Figura 3**).

Di solito questo tipo di ricerche produce un elenco troppo lungo di lavori, che vengono elencati venti alla volta con le informazioni essenziali (titolo, autori, rivista e data; **Figura 4**). Di *default*, i lavori vengono elencati in ordine di data di pubblicazione a partire dal più recente. È anche possibile chiedere al motore di ricerca di visualizzare i lavori che costituiscono la *best match* con la parola-chiave cercata: in questo caso, il motore di ricerca usa un metodo di *machine learning* per elencare i lavori sulla base di una valutazione della loro **rilevanza**. Nelle ricerche generiche, il numero di lavori che vengono selezionati decresce con l'aumentare del numero di parole-chiave che si usano (oltre che ovviamente col diminuire della loro frequenza nel *database*).

Search results
 Items: 1 to 20 of 9171
 << First < Prev Page 1 of 459 Next > Last >>

[Exploring Proteomic Drug Targets, Therapeutic Strategies and Protein-Protein Interactions in Cancer: Mechanistic View.](#)
 1. Dar KB, Bhat AH, Amin S, Anjum S, Reshi BA, Zargar MA, Masood A, Ganie SA.
 Curr Cancer Drug Targets. 2018 Aug 2. doi: 10.2174/1568009618666180803104631. [Epub ahead of print]
 PMID: 30073927
[Similar articles](#)

[A natural compound derivative P-13 inhibits STAT3 signaling by covalently inhibiting Janus kinase 2.](#)
 2. Huang H, Niu J, Wang F, Hu L, Yu Q.
 Invest New Drugs. 2018 Aug 3. doi: 10.1007/s10637-018-0637-2. [Epub ahead of print]
 PMID: 30073465
[Similar articles](#)

[Clinical and genetic profiles of patients with X-linked agammaglobulinemia from southeast Turkey: Novel mutations in BTK gene.](#)
 3. Doğruel D, Serbes M, Şaşıhüseyinoğlu AŞ, Yılmaz M, Altıntaş
 Allergol Immunopathol (Madr). 2018 Jul 30. pii: S0301-0546(18)30074-0. doi: 10.1016/j.aller.2018.03.004. [Epub ahead of print]
 PMID: 30072168
[Similar articles](#)

Annotations: **titolo** (points to title of item 1), **autori** (points to authors of item 2), **estremi referenza** (points to journal info of item 3), **codice PMID** (points to PMID of item 3).

Figura 4

Per ognuno dei lavori identificati, sono mostrati il titolo, gli autori, gli estremi della referenza e il PMID. Il titolo di ogni lavoro può essere usato come link per accedere a ulteriori dati, quali per esempio l'abstract e il link al lavoro completo (vedi Figura 5).

Ricerca avanzata in PubMed

La ricerca avanzata è molto più efficace, e consente di usare le parole-chiave di interesse per cercarle solo in campi specifici. Per esempio, se cerchiamo un lavoro con autore Friedreich, sarà molto utile usare la ricerca del nome Friedreich nel solo campo *author* (autore), visto che effettuando una ricerca generica in PubMed col nome Friedreich verranno elencati migliaia di lavori sulla Friedreich *ataxia* (l'ataxia di Friedreich), una patologia neurologica abbastanza studiata. Restrungendo la ricerca del nome al solo campo *author*, il motore di ricerca restituisce un elenco di pochissimi lavori! Per usare la ricerca "per campi", bisogna selezionare l'opzione *advanced* che si trova sotto all'area in cui si possono elencare le parole-chiave.

Si visualizza quindi il menu di **Figura 5A**. È possibile selezionare il campo di ricerca di proprio interesse in un menu a tendina (*fields*, **Figura 5B**) e scrivere nell'apposita area di testo la parola-chiave corrispondente. Più criteri possono essere combinati usando gli operatori *booleani* (AND, OR oppure NOT; vedi anche **Box 2.1** nel libro).

In **Figura 5C** è riportata la ricerca combinata per ottenere l'elenco di tutti i lavori pubblicati dall'autore Eisenberg D sulla rivista *Nature*. Sono stati selezionati i campi *author* e *journal* e sono state scritte le parole-chiave scelte. In seguito alla scelta di uno specifico campo e mentre si procede all'inserimento delle parole-chiave, il motore di ricerca suggerisce possibili parole-chiave che possono essere selezionate da chi effettua la ricerca per risparmiare tempo ed evitare errori di battitura. Quando si è completata la formulazione dei criteri (eventualmente combinati con gli opportuni operatori *booleani*) si può lanciare la ricerca premendo sul bottone *search*. La ricerca effettuata e i suoi risultati vengono immediatamente mostrati e la stringa viene conservata nella *history* dell'utente per qualche giorno. Le ricerche nella *history* possono venire a loro volta riutilizzate e composte usando gli operatori *booleani*. Effettuando ricerche più complesse (come per esempio per cercare tutti i lavori di un autore su due riviste) è molto importante controllare la sintassi della stringa usata per la ricerca e controllare che le precedenze logiche siano garantite da un corretto uso delle parentesi

A PubMed Advanced Search Builder

Filters activated: Humans. [Clear all](#)

History deleted.

Use the builder below to create your search

[Edit](#) [Clear](#)

Builder

All Fields [Show index list](#)

AND All Fields [Show index list](#)

[Search](#) or [Add to history](#)

B

Affiliation

All Fields

Author

Author - Corporate

Author - First

Author - Full

Author - Identifier

Author - Last

Book

Conflict of Interest Statements

Date - Completion

Date - Create

Date - Entrez

Date - MeSH

Date - Modification

Date - Publication

EC/RN Number

Editor

Filter

Grant Number

ISBN

Investigator

Investigator - Full

Issue

Journal

Language

Location ID

MeSH Major Topic

MeSH Subheading

MeSH Terms

Other Term

Pagination

Pharmacological Action

Publication Type

Publisher

Secondary Source ID

Subject - Personal Name

Supplementary Concept

Text Word

Title

Title/Abstract

Transliterated Title

Volume

C

(Eisenberg D[Author] AND "Nature"[Journal])

[Edit](#)

Builder

Author Eisenberg D [Show index list](#)

AND Journal Nature [Show index list](#)

AND All Fields [Show index list](#)

[Search](#) or [Add to history](#)

Figura 5 - Ricerca avanzata in PubMed

È possibile selezionare il campo di ricerca di proprio interesse in un menu a tendina (*fields*) e scrivere nell'apposita area di testo la parola-chiave corrispondente (A). Più criteri possono essere composti usando gli operatori *booleani* (AND, OR oppure NOT, vedi anche Box 2.1 nel libro). La scelta dei campi è riportata in B. In C si può vedere la ricerca combinata per ottenere l'elenco di tutti i lavori pubblicati da Eisenberg D sulla rivista *Nature*. Sono stati selezionati i campi *author* e *journal* e sono state scritte le parole-chiave scelte. Mentre si procede alla selezione dei campi e alla compilazione della ricerca, il motore di ricerca scrive una stringa corrispondente, che può anche essere modificata da chi effettua la ricerca in caso di necessità.

Procediamo ora con un esempio di ricerca avanzata complessa per mettere in evidenza alcuni errori in cui è possibile incorrere e dare alcuni suggerimenti su come evitarli.

Vogliamo ottenere l'elenco di tutti i lavori che il noto bioinformatico David Eisenberg ha pubblicato sulle riviste *Science* e *Nature*, usando la composizione dei criteri nella ricerca avanzata.

Prima di cominciare, effettuiamo separatamente le due ricerche, facendoci elencare i lavori di Eisenberg su *Nature* e i lavori di Eisenberg su *Science*, come descritto nella **Figura 5C**. Eisenberg ha pubblicato 18 lavori su *Nature* e 23 lavori su *Science*. Ora sappiamo che la risposta giusta alla nostra domanda è 41 (23 + 18; i dati sono aggiornati a novembre 2018).

Proviamo ora a comporre il tutto in un'unica ricerca.

- 1) Dal menu a tendina selezioniamo *author* e cominciamo a scrivere il cognome che ci interessa (Eisenberg). Bisogna ricordare che **si comincia col cognome**. Il motore di ricerca usa i nomi noti, per aiutarci nella compilazione e scrive dei caratteri di autocompilazione. Non sempre è una buona idea affidarsi all'autocompilazione, che propone i nomi più frequenti che non necessariamente sono quelli che cerchiamo. Una volta scritto il cognome, il motore di ricerca ci propone delle iniziali (il cognome Eisenberg è abbastanza comune e ci sono diverse iniziali possibili) e anche dei nomi. Per esempio, ci propone Eisenberg David. È importante ricordare che per identificare tutti i lavori di un autore, è meglio non usare il nome per intero oltre al cognome, in quanto prima del 2000 PubMed annotava solo le iniziali dei nomi, e pertanto cercando l'autore col nome intero (Eisenberg David) si perderebbero tutti i lavori di quell'autore pubblicati prima del 2000 (Eisenberg D).
- 2) Proseguiamo poi selezionando il campo *journal* nel menu a tendina e nell'apposito spazio cominciamo a scrivere il nome della rivista. In questo caso, è davvero una buona idea affidarsi all'autocompilazione, scegliendo la rivista che ci interessa tra quelle proposte. Se non scegliamo il

nome appropriato per la rivista di nostro interesse, è possibile che il risultato della nostra ricerca sia sbagliato.

- 3) Selezioniamo ancora il capo *journal* per aggiungere la rivista *Science*, che il motore di ricerca identifica come “Science (New York, N.Y.)”.
- 4) Ora siamo pronti per lavorare sulla composizione di questi 3 criteri di ricerca usando gli operatori *booleani*. La ricerca che vogliamo effettuare deve avere Eisenberg come autore e le riviste possono essere *Nature* oppure *Science*. Quindi vogliamo che l’autore Eisenberg sia obbligatoriamente presente e useremo l’operatore AND tra il nome dell’autore e il nome delle due riviste, che possono essere presenti in alternativa non disgiuntiva, quindi separate dall’operatore OR. Se scriviamo la ricerca come mostrato in **Figura 6**, il motore di ricerca interpreta:

((Eisenberg D[Author]) AND "Nature"[Journal]) OR "Science (New York, N.Y.)"[Journal]

Questa ricerca ci propone un elenco di più di 170 000 lavori! Cosa c’è di sbagliato?

Esaminando con attenzione i criteri di ricerca usati, vediamo che c’è un AND tra l’autore e la rivista *Nature* (quindi avremo 18 lavori), e poi un OR sulla rivista *Science* (il che significa che avremo tutti i lavori pubblicati sulla rivista *Science*!). È sbagliato l’uso delle parentesi, che possiamo modificare fino a ottenere quello che davvero vogliamo, ovvero:

Eisenberg D[Author] AND ("Nature"[Journal] OR "Science (New York, N.Y.)"[Journal])

Questa ricerca darà come risultato il 41 corretto, confermando che serve attenzione nell’uso delle parentesi quando si effettuano ricerche complesse.

Una volta che abbiamo ottenuto il riferimento a un lavoro che ci interessa e che vogliamo leggere per intero, si può seguire il *link* del titolo del lavoro stesso (come già evidenziato in Figura 4) e accedere quindi all’abstract del lavoro (**Figura 7**). Dalla stessa pagina è possibile arrivare all’intero lavoro, disponibile sul sito della rivista su cui è stato pubblicato. Se il lavoro è stato pubblicato con la politica dell’*open access* (vedi oltre), l’accesso è gratuito e il lavoro può essere letto e scaricato (come html o come pdf), altrimenti è necessario pagare per scaricare il lavoro (la cifra si aggira di solito sui 30 \$), a meno che non sia in essere un abbonamento con l’istituzione di appartenenza.

((Eisenberg D[Author]) AND "Nature"[Journal]) OR "Science (New York, N.Y.)"[Journal]

[Edit](#)

Builder

	Author	⇩	Eisenberg D	-	Show index list			
	AND	⇩	Journal	⇩	Nature	-	Show index list	
	OR	⇩	Journal	⇩	Science (New York, N.Y.)	-	Show index list	
	AND	⇩	All Fields	⇩		-	+	Show index list

or [Add to history](#)

Figura 6 - Ricerca avanzata in PubMed

Format: Abstract ▾

Send to ▾

Invest New Drugs. 2018 Aug 3. doi: 10.1007/s10637-018-0637-2. [Epub ahead of print]

A natural compound derivative P-13 inhibits STAT3 signaling by covalently inhibiting Janus kinase 2.Huang H^{1,2}, Niu J^{1,2}, Wang F^{1,2}, Hu L³, Yu Q^{4,5}.

⊕ Author information

Abstract

We investigated the function and molecular mechanisms of 2-deoxy-4β-propylcarbamate-pulchellin (P-13), a sesquiterpene lactone derivative of 2-deoxy-4-epi-pulchellin from the traditional Chinese medicinal herb *Carpesium abrotanoides* L., in regulating STAT3 signaling and cancer cell growth. We found that P-13 inhibited the IL-6-induced, as well as the constitutive, STAT3 activation in a dose and time-dependent manner. In vitro kinase activity analyses demonstrated that P-13 directly inhibited JAK2 kinase activity. The inhibitory effects of P-13 on JAK2/STAT3 signaling could be blocked by reducing agents dithiothreitol (DTT) or glutathione (GSH), indicating an involvement of the thiol-reactive α-β unsaturated carbonyl group in P-13. Further analyses with mass spectrograph, as well as molecular docking, revealed that P-13 covalently bound with the C452 in the SH2 domain of JAK2. Furthermore, P-13 inhibited growth and induced death of many cancer cell lines, particularly those expressing constitutively activated STAT3. It also inhibited in vivo growth of human cancer cell xenografts. Taken together, these findings revealed P-13 as a novel covalent inhibitor of JAK2, uncovered a new mechanism to inhibit JAK2, and provided a promising anti-cancer drug candidate.

KEYWORDS: Cancer; Covalent interaction; JAK2/STAT3; Natural compound; Signal transduction

PMID: 30073465 DOI: 10.1007/s10637-018-0637-2

Full text linksSpringerLink
FULL-TEXT ARTICLE**Save items**

★ Add to Favorites ▾

Similar articles

Bigelovin inhibits STAT3 signaling by inactivating JAK2 and induces apoptosis [Acta Pharmacol Sin. 2015]

Parthenolide Inhibits STAT3 Signaling by Covalently Targeting Janus Kinase [Molecules. 2018]

2-Methoxystyrylpyridone inhibits signal transducer and activator of transcription 3 [Cancer Sci. 2014]

17-hydroxy-jolkinolide B inhibits signal transducers and activators of transcription 3 [Cancer Res. 2009]

Eriocalyxin B Inhibits STAT3 Signaling by Covalently Targeting STAT3 [PLoS One. 2015]

See reviews...

Figura 7

Seguendo il link collegato al titolo del lavoro (vedi Figura 4), si accede a una pagina contenente l'abstract del lavoro e il link all'intero contenuto del lavoro, disponibile (in questo caso a pagamento) sul sito della rivista scientifica, qui evidenziato dalla freccia rossa.

Struttura di un lavoro scientifico

La maggior parte degli articoli scientifici ha una struttura comune. Ogni lavoro è preceduto da un breve testo, detto **abstract** (Figure 2 e 7), che riassume le motivazioni del lavoro e i risultati conseguiti e descritti. Segue in genere un'introduzione, in cui si riportano gli elementi essenziali a inquadrare il lavoro pubblicato nel contesto dei lavori che lo hanno preceduto sullo stesso argomento. L'introduzione in genere è ricca di **citazioni** ai lavori precedenti, che sono poi elencati nelle referenze alla fine del lavoro stesso. Dopo l'introduzione spesso si passa ai **risultati**, che costituiscono il cuore dell'articolo e che vengono spiegati e descritti, accompagnati da figure (foto di dati sperimentali o grafici e diagrammi). Le parti tecniche e metodologiche, che sono indispensabili per poter eventualmente riprodurre i risultati riportati nell'articolo, vengono descritte nei **metodi**. A seguire c'è poi la **discussione**, una parte di solito molto interessante, in cui i risultati vengono inquadrati nel contesto delle ricerche precedenti e se ne evidenzia la validità, la novità e gli eventuali futuri sviluppi. In chiusura è riportato l'elenco delle **referenze**, ovvero di tutti i lavori che si è ritenuto utile e importante citare nell'articolo, per inquadrare gli argomenti, per usare dati o conclusioni già ottenuti in altri lavori, per consentire al lettore di accedere agli stessi dati, metodi o conclusioni usati nella stesura del lavoro. Le referenze sono riportate in un formato compatto (solitamente autori, titolo, anno ed estremi del lavoro), deciso dalla rivista in cui il lavoro viene pubblicato.

Le referenze nei lavori scientifici: uso e consumo

Nel testo del lavoro pubblicato, quindi, si fa continuamente riferimento ad altri lavori, spesso identificati col numero relativo all'elenco finale delle referenze (vedi paragrafo precedente e **Figura 8**).

Le citazioni sono una parte importante del lavoro scientifico e costituiscono inoltre un riconoscimento e una testimonianza dell'utilità dei lavori citati. Un lavoro che non viene citato non è stato utile nel cammino della ricerca, mentre invece un lavoro molto citato (ovvero citato in più lavori) è evidentemente importante e riconosciuto come tale dalla comunità scientifica. Da ciò consegue che le citazioni vengono spesso usate come un indicatore della qualità di un lavoro. Non sono un indicatore ideale, perché dipendono da vari fattori, tra cui il tempo (le citazioni aumentano col tempo, quindi non si possono confrontare lavori pubblicati in anni diversi), il momento (ci sono lavori ottimi e importanti che non vengono subito riconosciuti come tali dalla comunità scientifica), la notorietà degli autori sulla base della carriera scientifica precedente e anche la particolare comunità scientifica interessata a quel tipo di lavoro. Per esempio, in genere lavori di interesse clinico sono più citati rispetto a lavori di ricerca di base.

tion. Motifs can be detected (with relative ease) at the primary sequence level, but they almost always have a structural meaning, being clusters of spatially close residues working in concert to achieve a given function. The bioinformatics field of motif finding in proteins and DNA is well developed, providing several tools, approaches and databases (1–3), while fewer resources are available for structural motif finding in RNAs. Such tools can be particularly useful in helping the functional characterization of non-coding RNAs (ncRNAs), for which information about the involved specific sequences and structures is still scarce. ncRNAs are involved in a wide range of biological functions through diverse molecular mechanisms often involving the interaction with one or more RNA binding protein (RBP) partners, with other RNAs or with the genomic DNA (4,5). Experimental and computational tech-

REFERENCES

1. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
2. Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P. and Bateman, A. (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.*, **41**, D226–D232.
3. Sonnhammer, E.L.L., Eddy, S.R. and Durbin, R. (1997) Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins Struct. Funct. Genet.*, **28**, 405–420.
4. Fallmann, J., Sedlyarov, V., Tanzer, A., Kovarik, P. and Hofacker, I.L. (2015) AREsite2: An enhanced database for the comprehensive investigation of AU/GU/U-rich elements. *Nucleic Acids Res.*, **44**, D90–D95.
5. Leibovich, L. and Yakhini, Z. (2012) Efficient motif search in ranked lists and applications to variable gap motifs. *Nucleic Acids Res.*, **40**, 5832–5847.
6. Lukong, K.E., Chang, K.W., Khandjian, E.W. and Richard, S. (2008) RNA-binding proteins in human genetic disease. *Cell*, **24**, 416–425.
7. Kishore, S., Luber, S. and Zavolan, M. (2010) Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. *Brief. Funct. Genomics*, **9**, 391–404.

Figura 8

A sinistra è riportata parte dell'introduzione di un lavoro. Con un rettangolino rosso sono evidenziati i riferimenti (1-3) a tre articoli che sono riportati nella parte finale del lavoro, e cioè nelle *referenze*, mostrate a destra.

Valutazione

Valutazione delle riviste scientifiche: l'impact factor

Le citazioni vengono anche usate per valutare le riviste scientifiche, ognuna delle quali viene associata a un fattore di impatto, detto IF o *impact factor*, che si calcola come il numero medio di citazioni per articolo per anno. Il calcolo viene mediato su due anni consecutivi, e il calcolo comincia due anni prima. Per esempio, nel 2018 non si tiene conto dei lavori pubblicati nel 2018 o nel 2017, perché si considera che sia passato troppo poco tempo perché i lavori possano essere stati elaborati dalla comunità scientifica e abbiano potuto produrre delle citazioni in lavori a loro volta pubblicati. Per cui, nel 2018 escono gli IF relativi al 2017, calcolati sugli anni 2015-2016 (Figura 9).

Naturalmente, quanto più alto è l'impact factor tanto maggiore è il prestigio della rivista. Per esempio, le riviste *Nature* e *Science* prima citate hanno un IF 2017 pari a circa 41. Ciò non toglie che anche la rivista *Bioinformatics*, con un IF 2017 pari a 5,48, non sia ugualmente prestigiosa nel suo campo (vedi Nota). I valori di IF vengono normalmente riportati sul sito web delle riviste corrispondenti oppure possono essere consultati da banche dati commerciali come Journal of Citation Reports di Clarivate Analytics.

Nota: Ogni rivista è classificata in base all'ambito scientifico dei lavori che pubblica. Per esempio, le riviste *Science* e *Nature* sono classificate come multidisciplinari, dato che pubblicano lavori nei vari ambiti scientifici (fisica, storia, biologia, medicina ecc.), mentre *Bioinformatics* è catalogata tra matematica e biologia computazionale. L'IF è quindi un importante indicatore ma occorre considerarlo anche in relazione alla categoria specifica della rivista. *Science*, *Nature* e *Bioinformatics*, nonostante abbiano degli IF molto diversi, sono le riviste con più alto IF nelle rispettive categorie.

Valutazione dei singoli lavori: Scopus, WoS, Google Scholar

Come spiegato, l'IF si riferisce a una rivista scientifica, ma spesso si vuole sapere se un determinato lavoro è più o meno citato. I singoli lavori pubblicati nella stessa rivista possono avere un numero di citazioni molto diverso, a seconda dell'effettivo impatto del singolo lavoro sulla comunità scientifica. È chiaro che in generale un lavoro pubblicato su una rivista ad alto IF sarà migliore di un altro pubblicato su una rivista a basso IF, ma è davvero sorprendente scoprire quanti lavori molto citati e importanti siano stati pubblicati su riviste di importanza relativa.

	Full Journal Title	Total Cites	Journal Impact Factor
1	NATURE MEDICINE	75,461	32.621
2	CELL	230,625	31.398
3	Annual Review of Biochemistry	19,873	20.154
4	TRENDS IN BIOCHEMICAL SCIENCES	16,944	15.678
5	MOLECULAR CELL	61,604	14.248
6	Nature Chemical Biology	19,562	13.843
7	NATURE STRUCTURAL & MOLECULAR BIOLOGY	27,547	13.333
8	TRENDS IN MICROBIOLOGY	11,344	11.776
9	MOLECULAR PSYCHIATRY	18,460	11.640
10	NUCLEIC ACIDS RESEARCH	168,962	11.561

Journal Impact Factor Calculation

$$\text{2017 Journal Impact Factor} = \frac{29,492}{2,551} = 11.561$$

How is Journal Impact Factor Calculated?

$$\text{JIF} = \frac{\text{Citations in 2017 to items published in 2015 (16200) + 2016 (13292)}{29,492}}{\text{Number of citable items in 2015 (1281) + 2016 (1270)}{2,551}} = \frac{29,492}{2,551}$$

Figura 9

A sinistra si può osservare l'elenco delle migliori riviste di ambito biochimico e biomolecolare in ordine di IF 2017 decrescente, col numero totale di citazioni e l'*impact factor* calcolato sulla base delle citazioni medie degli articoli pubblicati in due anni consecutivi. A destra è riportato il calcolo per la rivista *Nucleic Acids Research*.

Le citazioni dei singoli lavori possono essere trovate su banche dati specializzate (Scopus o Web of Science, detto anche WoS, disponibili per l'università o gli istituti di ricerca) o su Scholar, una risorsa sviluppata da Google, disponibile gratuitamente online. Scopus e WoS calcolano le citazioni elaborando i dati degli articoli pubblicati; i loro dati sono controllati e affidabili. Scholar lavora indicizzando il web e quindi tratta come citazioni anche i richiami che non sono pubblicati su riviste scientifiche, ma semplicemente sul web. Le citazioni su Scholar sono quindi sempre superiori o uguali a quelle di Scopus e WoS, ma sono comunque rapide da ottenere, consentono di vedere velocemente sia chi ha citato un certo articolo sia su quale rivista è comparsa la citazione e sono accessibili gratuitamente.

L'analisi delle citazioni è anche molto utile per avere un quadro aggiornato dello stato dell'arte in un particolare ambito di ricerca. Se analizziamo le citazioni della letteratura a noi nota su un particolare ambito di ricerca, magari rilevante per il nostro progetto, avremo un quadro completo, aggiornato e anche parzialmente valutato degli sviluppi più recenti.

Valutazione degli scienziati: l'h-index

Ci sono molte metriche anche per valutare la qualità dei singoli ricercatori, autori di lavori scientifici, tra cui il numero dei lavori e l'IF medio. Un ulteriore tipo di valutazione viene fatto attraverso il calcolo dell'indice di Hirsch, anche detto *h-index*. Per calcolare l'*h-index* di uno scienziato, bisogna ordinare i suoi lavori in base al numero decrescente delle citazioni ricevute e contare quanti di essi hanno un numero di citazioni superiori o pari al proprio numero d'ordine (Figure 10 e 11). I valori di *h-index* per un certo autore vengono forniti dalle banche dati citate in precedenza, come Scopus, WoS o Google Scholar. È comunque importante ricordare che nei lavori pubblicati in ambito biomedico, l'ordine dei nomi degli autori segue regole precise. In generale, il primo autore corrisponde al ricercatore che ha svolto la maggior parte del lavoro, mentre l'ultimo autore e/o l'autore corrispondente, ovvero quello che deve essere contattato per fornire chiarimenti o ulteriori informazioni sull'articolo, corrispondono generalmente al coordinatore del progetto di ricerca. In altri settori, invece, (matematica, fisica, statistica, per fare degli esempi) l'ordine dei nomi segue quello alfabetico e il contributo di ogni singolo coautore è riportato separatamente. L'*h-index* è calcolato in modo

indipendente dall'ordine degli autori, penalizzando quindi primi e ultimi autori a vantaggio di quelli che hanno dato contributi di minore importanza. Esistono anche altre metriche più sofisticate e corrette, che tengono conto sia delle citazioni sia delle posizioni dei nomi degli autori nella valutazione dei singoli scienziati.

Documents	Citations	Title
1	112	Deciphering a global network of fu...
2	94	In situ structural analysis of the hu...
3	61	PTMcode: A database of known an...
4	30	PTMcode v2: A resource for functio...
5	20	Spatiotemporal variation of mamm...
6	9	Histone deacetylase inhibitors (HD...
7	9	Phosphate binding sites identifi...
8	7	Phosfinder: A web server for the id...
9	5	Identification of Nucleotide-Binding...
10	4	Nucleos: a web server for the ident...
11	2	Spatial tissue proteomics quantifi...
12	1	Systematic identification of phosph...

This author's *h*-index is 7

The *h*-index is based upon the number of documents and number of citations.

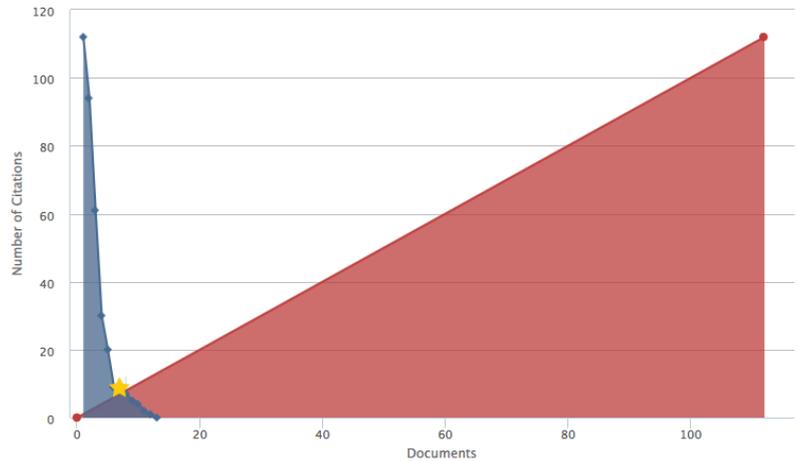


Figura 10

Sulla sinistra è riportato l'elenco dei lavori di un ricercatore, ordinato sulla base del numero decrescente delle citazioni che ogni lavoro ha ricevuto. L'*h*-index in questo caso è 7. L'*h*-index potrà salire a 8 quando uno dei lavori riportati sotto a quello evidenziato in giallo raggiungerà le 8 citazioni.

Einstein, Albert

Institute for Advanced Studies, Princeton, United States
 Author ID: 22988279600
 Other name formats: EINSTEIN, A. Einstein, A.

Subject area: Physics and Astronomy Mathematics Environmental Science
Biochemistry, Genetics and Molecular Biology Multidisciplinary Engineering Medicine

Document and citation trends:

[Follow this Author](#)

h-index: 40 [View *h*-graph](#)

Documents by author: 102 [Analyze author output](#)

Total citations: 21315 by 18655 documents [View citation overview](#)

Citata da [VISUALIZZA TUTTO](#)

	Tutte	Dal 2013
Citazioni	120302	36954
Indice H	111	66
i10-index	369	216

Figura 11

Lo scienziato Albert Einstein, noto per i suoi fondamentali contributi nella fisica teorica di base, su Scopus risulta avere un *h*-index pari a 40. Su Scholar, Einstein viene valutato con un *h*-index stratosferico, pari a 195, sicuramente più vicino al comune sentire.

Politiche per la pubblicazione: *open access* e *close access*

Perché un lavoro possa essere pubblicato, deve essere sottomesso alla rivista scelta. In genere il manoscritto viene affidato a un *editor*, un ricercatore esperto nel settore. L'*editor* lo manda in revisione ad almeno 2 *reviewer*, che lo valutano con attenzione e producono una critica scritta, eventualmente chiedendo aggiunte, modifiche o correzioni. Questo processo di revisione da parte di *pari* (ovvero ricercatori, come gli autori del manoscritto in revisione) è chiamato *peer review*. In generale, i *reviewer* sono anonimi per gli autori. Una volta che il manoscritto ha superato il processo di revisione può essere pubblicato e reso disponibile alla comunità scientifica. Se invece il manoscritto non supera il processo di revisione, si dice *rigettato* e gli autori possono eventualmente sottometterlo a un'altra rivista.

Esistono due politiche principali per la pubblicazione di un lavoro: l'*open access* e il *close access*.

Con la politica dell'*open access*, gli autori pagano una quota alla rivista (qualche migliaio di dollari o euro) e il loro lavoro viene reso disponibile *online* come html e come pdf. Chiunque può leggerlo e avvantaggiarsene senza pagare, a tutto vantaggio della diffusione del lavoro e della sua possibilità di essere citato.

Con la politica del *close access*, l'unica disponibile fino all'anno 2000 circa, ogni lavoro veniva reso disponibile a *pagamento*. Le riviste vendevano abbonamenti o singoli lavori e con questo si sostenevano.

Esiste una terza possibilità, intermedia, che consiste nel mantenere un lavoro disponibile a pagamento per il primo anno (o per i primi sei mesi) dal momento della pubblicazione per poi renderlo disponibile gratuitamente.

L'argomento è complesso, ma sia l'NIH sia la Comunità Europea si sono schierati a favore della politica dell'*open access*, per favorire l'accesso alla letteratura scientifica anche ai ricercatori delle comunità scientifiche meno ricche.

I lavori che sono disponibili gratuitamente, perché pubblicati con la politica dell'*open access* o perché è ormai trascorso il tempo di cessione a pagamento, sono disponibili in PubMed Central e in Europe PMC. In PubMed, l'accesso al testo completo avviene come mostrato in **Figura 12**.

The screenshot shows a PubMed article page. On the left, the article title is "Homeobox oncogene activation by pan-cancer DNA hypermethylation." Below the title, there is an "Abstract" section with "BACKGROUND", "RESULTS", and "CONCLUSIONS" paragraphs. On the right side of the page, there are several interactive sections: "Send to", "Full text links" (with a red arrow pointing to a "Read free full text at BMC" button), "Save items" (with an "Add to Favorites" button), and "Similar articles" (listing related research papers). At the bottom right, there are links for "See reviews..." and "See all...".

Figura 12

Ecco un lavoro pubblicato con la politica dell'*open access*. Il contenuto del lavoro è disponibile gratuitamente seguendo il link indicato con la freccia rossa.

Sitografia	
PubMed	https://www.ncbi.nlm.nih.gov/pubmed
Europe PMC	https://europepmc.org
Google scholar	http://scholar.google.it/
Scopus	https://www.scopus.com/
WoS	http://apps.webofknowledge.com
BioRxiv	https://www.biorxiv.org/

Esercizi

Ricerca generica e per campi in PubMed

- 1) Lanciare una ricerca in PubMed con parola-chiave Eisenberg.
- 2) Identificare il numero del totale degli articoli elencati e delle review e prenderne un appunto.
- 3) Ripetere la ricerca usando l'opzione "advanced search", e specificando che Eisenberg è un autore.
- 4) Prendere un appunto con il numero di articoli elencati e ripetere la ricerca usando anche l'iniziale del nome dell'autore (il server usa i caratteri inseriti per proporre come continuare: selezionare l'autore Eisenberg il cui nome comincia con "D", David; per far partire la ricerca premere "Search").
- 5) Confrontare i numeri relativi al risultato di questa ricerca con quelli ottenuti nelle ricerche precedenti.
- 6) Tornare all'opzione "advanced search" e porre condizioni sulla rivista, selezionando tutti gli articoli pubblicati su *Science*.
- 7) Componendo le ricerche effettuate ai punti precedenti, identificare i lavori dell'autore Eisenberg sulla rivista *Science*.
- 8) Identificare i lavori dell'autore Eisenberg D sulla rivista *Science* e prendere nota dei risultati.
- 9) Effettuare una nuova ricerca, selezionando con un'unica *query* tutti i lavori pubblicati da Eisenberg D sulla rivista *Nature*.
- 10) Usando una qualsiasi delle possibili metodologie dell'advanced search, identificare i lavori di Eisenberg D sulle riviste *Science* e *Nature*.
- 11) Verificare che il numero dei lavori ottenuto al punto 10 sia uguale alla somma dei lavori ottenuti ai punti 8 e 9.
- 12) Identificare il numero di articoli e il numero di review pubblicate su *Science* e *Nature* (di qualsiasi autore).
- 13) Scegliere un titolo di proprio interesse e seguire il *link* all'*abstract* del lavoro.
- 14) Verificare la disponibilità del lavoro completo.
- 15) Identificare il numero di articoli pubblicati sulla rivista *PLoS Biology*, scegliere un lavoro di proprio interesse e seguire il *link* all'*abstract*.
- 16) Accedere al lavoro scelto ed esaminarne testo e figure.
- 17) Cercare tutti i lavori dell'autore Baker D.
- 18) Cercare tutti i lavori pubblicati sulle riviste *PNAS* (Proc Natl. Acad. SCI USA) e *Bioinformatics* (indipendentemente dall'autore).
- 19) Usando il campo "Date – Publication", cercare tutti i lavori pubblicati negli ultimi 10 anni.
- 20) Seguire ora l'opzione "search history" (nell'advanced search) e comporre con l'operatore booleano AND le ricerche:
 - a) tutti i lavori con nome autore Baker D
 - b) tutti i lavori pubblicati su *PNAS* e *Bioinformatics*
 - c) tutti i lavori pubblicati negli ultimi 10 anni.
- 21) Annotare il numero di lavori elencati nell'output della ricerca e i PMID delle review ottenute.

Scholar

- 22) Collegarsi a scholar.google.com ed effettuare una ricerca col nome di Michael Levitt.
- 23) Individuare il “profilo” di questo autore e accedere all’elenco dei suoi lavori, di solito ordinato in base al numero decrescente delle citazioni.
- 24) Determinare l’*h-index* di Michael Levitt (Premio Nobel per la Chimica nel 2013) e confrontarlo con quello calcolato da Scholar.