

L'impatto dei *Big-Data* sulla ricerca biomedica

Federico Boem

Università di Milano

The End of Science

The quest for
knowledge used
to begin with
grand theories.
Now it begins
with massive
amounts of data.
Welcome to the
Petabyte Age.



Wired, 2007

Riassumendo

- **La scienza moderna come attività ipotetico-deduttivo-sperimentale sta finendo**
- **Abbandonare la nozione di causa (non servono per predire i fenomeni)**
- **Una scienza basata su grandi studi correlativi (“correlation is enough”)**

Glossario minimo (filosofico)

- **Fenomeni**
- **Sistemi sperimentali**
- **Stili di ragionamento**

Fenomeni

“I fenomeni non sono là fuori ad aspettare che gli scienziati li scoprano. È l’impresa dello sperimentare attivamente che genera le condizioni per cui i fenomeni diventano evidenti. Di fatto i fenomeni sono osservabili solo in certe specifiche condizioni”

Hacking, 1983



In altre parole

“In natura esiste solo la complessità, che siamo incredibilmente abili nell’analizzare. Lo facciamo [...] attraverso la presentazione in laboratorio di fenomeni puri ed isolati”

(Hacking 1983)

Sistemi sperimentali

“Nell’analizzare un problema il biologo è vincolato ad un frammento di realtà, un pezzo di universo che egli isola arbitrariamente al fine di definire certi suoi parametri. In biologia ogni studio [...] inizia con la scelta di un sistema. Tutto dipende da questa scelta: la portata entro la quale lo sperimentatore può muoversi, il tipo di domande che egli è in grado di porre e spesso anche le risposte che può fornire”

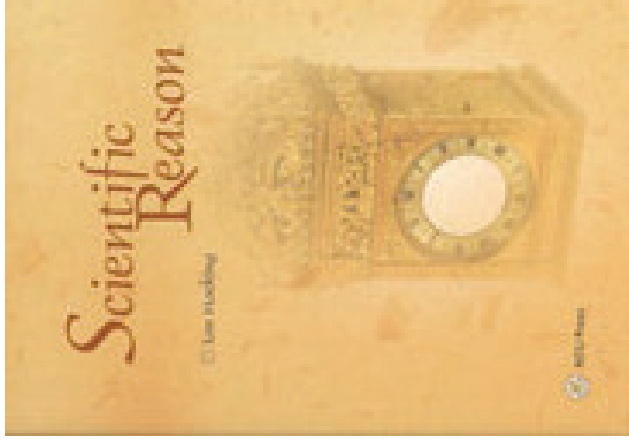
(Jacob 1988)



François Jacob, Nobel 1965

Stili di ragionamento

- Stili di ragionamento (Crombie 1994; Hacking 2009)
- Modi di sapere/fare (Pickstone 2000)



Che cos'è uno *stile*?

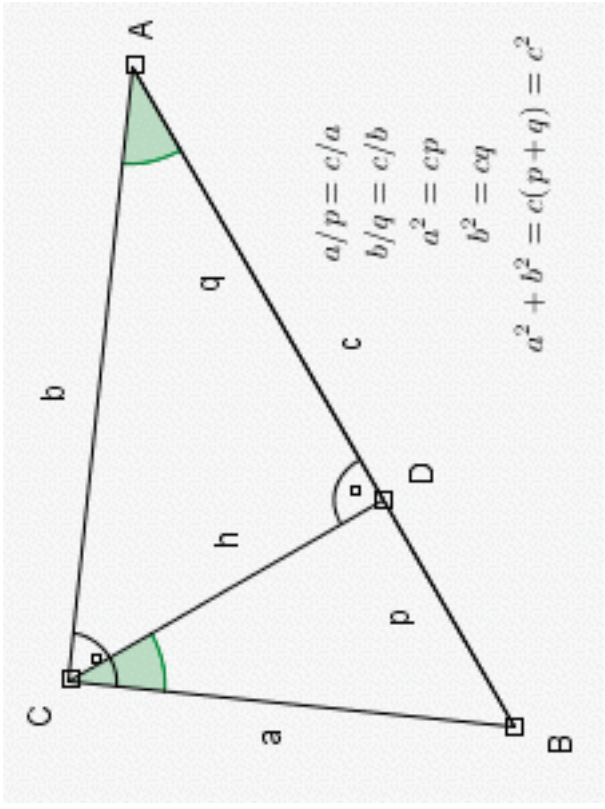
“Introduce nuovi oggetti della ricerca scientifica e nuove spiegazioni, nuovi tipi di prova, nuovi criteri nel determinare cosa conti come soluzione di un dato problema”

(Crombie, 1994)

Una categoria epistemica

Hacking sostiene che la peculiarità epistemica degli stili sia da ricercare non in merito a quali fatti siano rappresentati come veri o falsi ma sul come (e cioè secondo che “modo di pensare”) **tale conclusione circa la loro verità o falsità è ottenuta**

Esempi di stile



Glossario minimo (tecnico)

- **Omiche**
- **Big Data**
- **Data-Driven**

“Omiche”

- **Omiche:** discipline il cui fine è analizzare le interazioni e le funzioni tra entità e informazioni biologiche a diversi strati -ome (cluster) di organizzazione
- Genomica, epigenomica, proteomica, metabolomica ecc.

Big Data

La **Big Data Science** consiste certamente nella:

- **quantità** di dati (e.g. petabytes),
- nella **velocità** con cui questi dati vengono processati e ottenuti,
- nella **varietà** con cui sono ordinati e mostrati,
- nell'obiettivo esplicativo **globale** che essi promettono,
- nella **standardizzazione** delle procedure che riguardano la loro identificazione
- e nel **formato relazionale** con quale possono essere accresciuti o estesi in grandezza

Data-Driven

- **DD:** approcci che denotano una impostazione metodologica che privilegia la raccolta dei dati e il riconoscimento di pattern piuttosto che la formulazione delle ipotesi

LA BIOLOGIA?



La Torre di Babele, di Marten van Valckenborch, tardo XVII secolo

Perché questa immagine?

- Un insieme di discipline differenti dove si parlano linguaggi differenti
- **Biologia molecolare, biologia dello sviluppo, biologia evuzionistica ecc.**

Un passo indietro

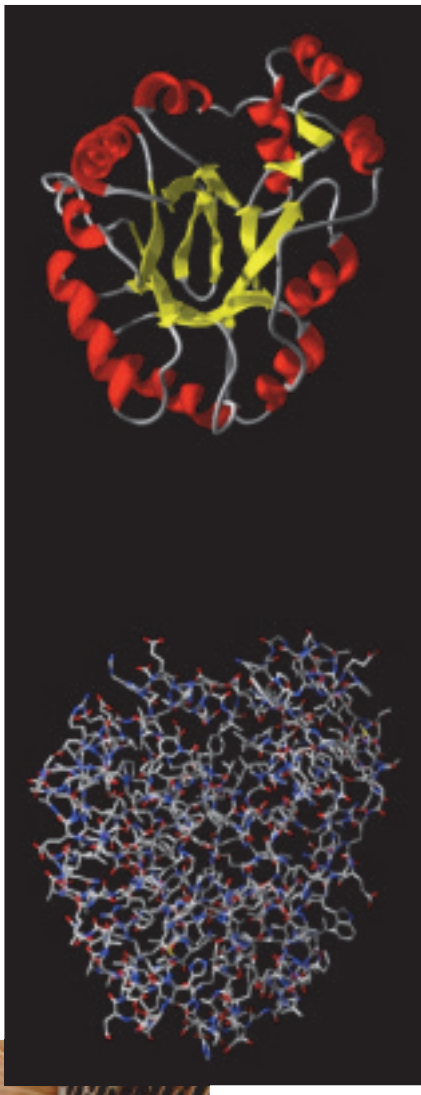
- Le molte origini della biologia
- Edmund Wilson nel suo “Aims and method of study in natural history”, (Science 1901) parla dei mestieri nel mondo della vita nei termini di:
 - *bug-hunters*,
 - *worm-slicers*
 - *egg-shakers*



Divisioni anche all'interno della biologia molecolare

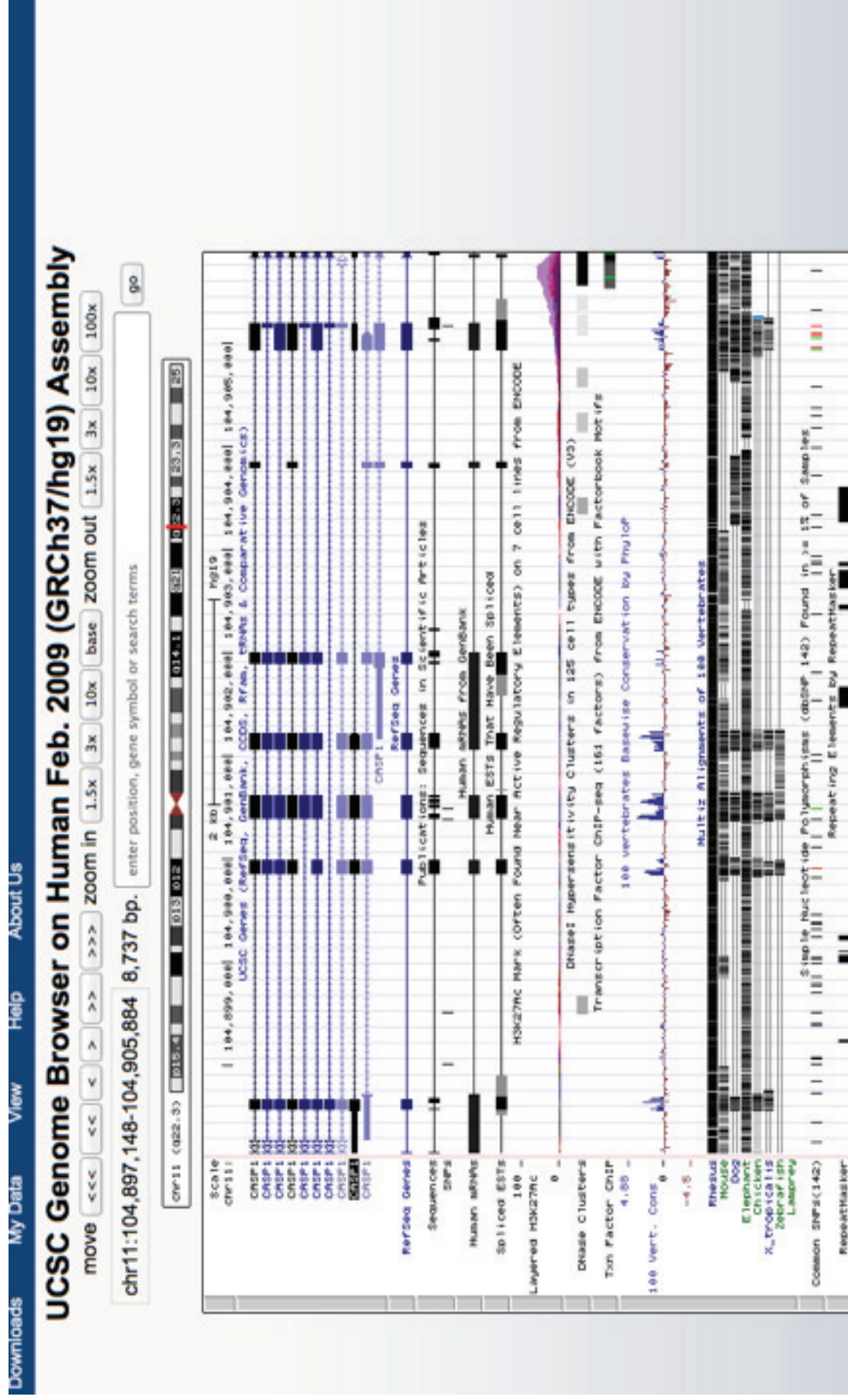


Biologia cellulare



Biologia strutturale

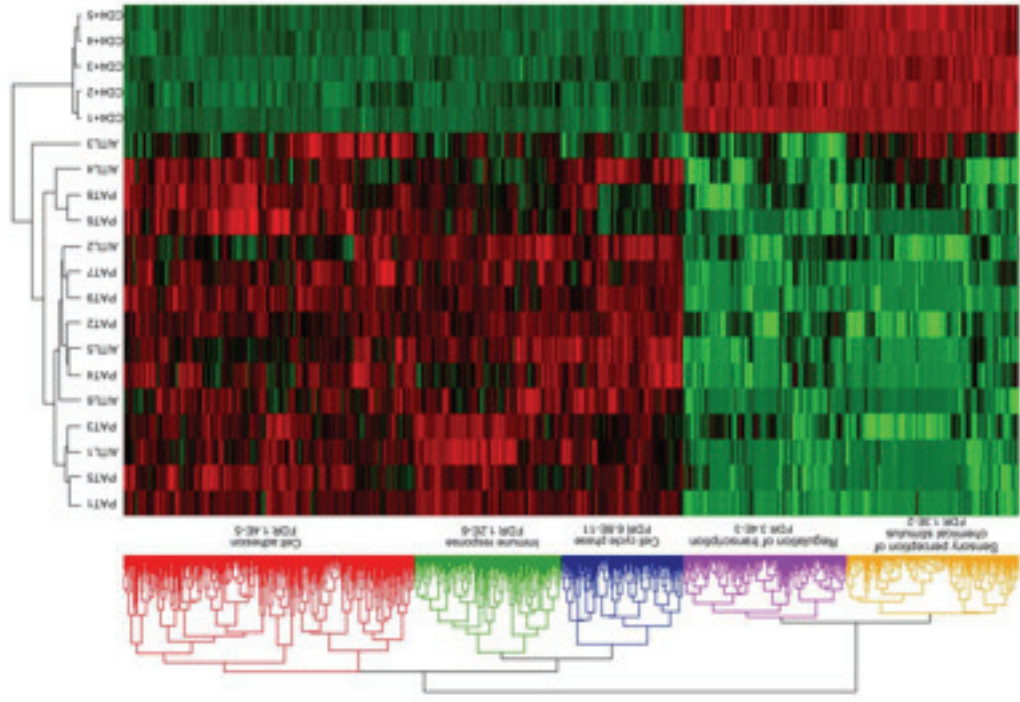
Genomica computazionale



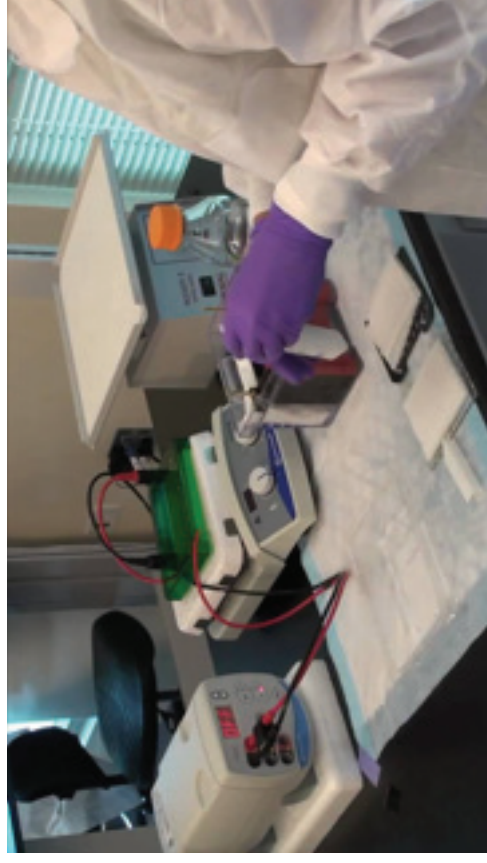
Una breve storia della biologia molecolare

- **La svolta molecolare** (anni 40' e 50'). Culminata nella scoperta della struttura del DNA da parte di Watson e Crick (1940s - 1950s)
- **La svolta biotecnologica** (anni 70'). Manipolazioni di oggetti biologici al fine produrre un certo risultato
- **L'avvento delle Omiche. Una terza svolta?**

Cosa significa fare biologia oggi?



vs



L'ascesa della visione molecolare

- **Esperimenti** come la via maestra per spiegare i fenomeni biologici
- **Metodi** importati dalla chimica, dalla fisica (più la metafora *informazionale*)
- **Un insieme di tecniche** più che una cornice teorica unificata, unificante e onnicomprensiva

Lo “stile molecolare”

- Trionfo dello “sperimentalismo”
- Legato anche a elementi di ragionamento analogico e **probabilistico** (Fisher e la bio-statistica)
- La vulgata: “La biologia molecolare è finalmente una scienza sperimentale, una **scienza matura** mentre la storia naturale era semplicemente **descrittiva**”

Meccanismi

“I meccanismi sono entità e/o attività organizzate in modo che siano produttive di cambiamenti regolari a partire da condizioni iniziali fino a condizioni finali”

(Machamer, Darden e Craver, 2000)

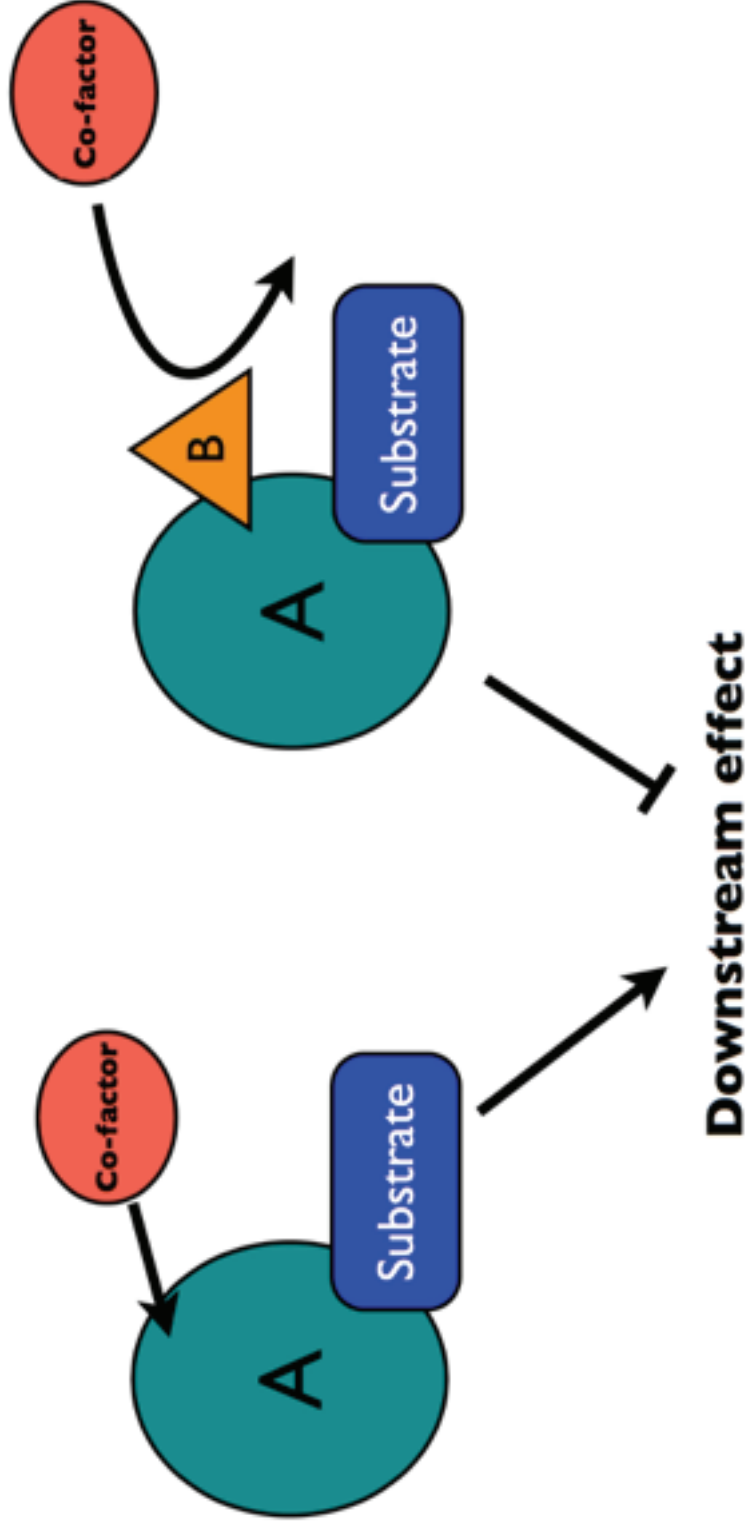
Biologia molecolare e meccanismi

- I biologi molecolari cercano meccanismi che producano, sottostiano o mantengano un fenomeno
- I meccanismi permettono spiegazione, predizione e **manipolazione/controllo**
- L'idea è che i meccanismi explicitino relazioni di natura **causale**
- **Modelli meccanicistici**

Scire per causas

Fin da Aristotele riteniamo che conoscere (nel senso di “saper spiegare”) un qualcosa significati essere in grado di **fornire le cause/ragioni**

Un modello molecolare: il meccanismo

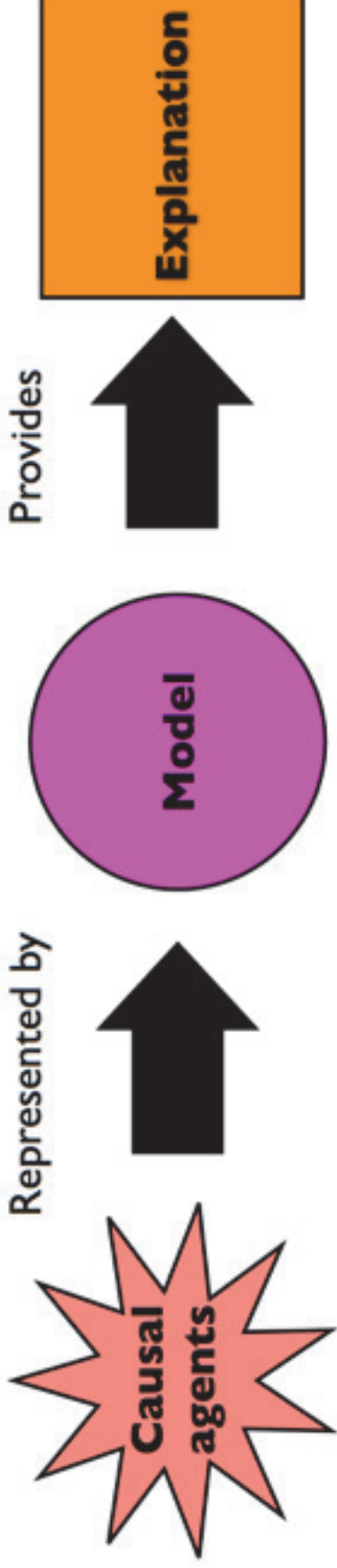


Cos'è un modello?

- Un modello è una **astrazione** di certe caratteristiche selezionate circa un dato fenomeno che permette agli scienziati di **evidenziare le connessioni causali e gli agenti causali** coinvolti in quel fenomeno
- Un modello **distorce** il suo oggetto di riferimento poiché esso non lo rispecchia fedelmente ma piuttosto **idealizza certe proprietà con il fine di far risaltare la sua (presunta) struttura causale**

(Cartwright 1983, Hacking 1983, Frigg & Hartmann 2009, Weisberg 2013)

Tagliando con l'accetta



- Questo è possibile anche perché il modello permette di manipolare l'oggetto di studio al fine di compiere simulazioni e predizioni
- Il modello “spiega” in virtù del suo essere “usabile”

La storia naturale

- Osservare
- Collezionare
- Confrontare



Collezionare non è scienza



“All science is either physics or stamp collecting”

citato in *Rutherford at Manchester* (1962) da J. B. Birks

The End of Science

The quest for
knowledge used
to begin with
grand theories.
Now it begins
with massive
amounts of data.
Welcome to the
Petabyte Age.



Wired, 2007

Un nuovo paradigma?



The FOURTH PARADIGM

DATA-INTENSIVE SCIENTIFIC DISCOVERY

EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

Ricerca Data-Driven

Due metodologie o due scienze?

OPINION HUMAN GENOME AT TEN

Point: Hypotheses first

There is little to show for all the time and money invested in genomic studies of cancer, says **Robert Weinberg** — and the approach is undermining tried-and-tested ways of doing, and of building, science.

NATURE | Vol 464 | 1 April 2010



Counterpoint: Data first

Large, unbiased genomic surveys are taking cancer therapeutics in directions that could never have been predicted by traditional molecular biology, says **Todd Golub**.

Nature 2010

Il punto di Weinberg

- “So have the old ways of doing business – of testing hypotheses become anachronisms? I think not”
- “Although aesthetically pleasing, [these approaches] have yielded few conceptual insights into how and why cells and tissues behave the way they do.”
- “Some feel that a thorough understanding of individual signal-processing components is an essential prerequisite to predicting the behaviour of entire signalling circuits”

Robert Weinberg, *Nature*, 2010

Reazioni

Leading Edge
Review

Cell

Lessons from the Cancer Genome

Levi A. Garraway^{1,2,4} and Eric S. Lander^{3,4,5,*}

¹Department of Medical Oncology and Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA 02215, USA

²Department of Medicine, Brigham and Women's Hospital

³Department of Systems Biology

Harvard Medical School, Boston, MA 02115, USA

⁴The Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

⁵Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

*Correspondence: lander@broadinstitute.org

<http://dx.doi.org/10.1016/j.cell.2013.03.002>

“These studies have revealed scores of new cancer genes, including many in processes not previously known to be causal targets in cancer”

Garraway&Lander, Cell 2013

La stessa cornice esplicativa

Big Data Biology: Between Eliminative
Inferences and Exploratory Experiments

Emanuele Ratti*†

- Le differenze non stanno a livello teorico
- Le euristiche della biologia computazionale compatibili con quelle della biologia molecolare

In breve

- Alcuni studi permettono alcune procedure inferenziali eliminative al fine di stabilire una **priorità nello sviluppo di ipotesi meccanicistico-sperimentali**
- Altri studi forniscono nuovi **strumenti esplorativi** nel fornire **utili generalizzazioni** che possano fare da complemento alle procedure di stabilire una priorità delle ipotesi
- Metodi per **“filtrare le ipotesi”**

Per chi ha frequentato un po' la filosofia...

- I dati non “parlano da soli”
- “Lo scienziato deve fare ordine: la scienza è composta di fatti così come una casa è composta di mattoni; ma un cumulo di fatti non è scienza più di quanto un mucchio di mattoni non sia una casa”

Henri Poincaré (1902)

Biologia molecolare vs biologica computazionale: uno scontro tra stili?

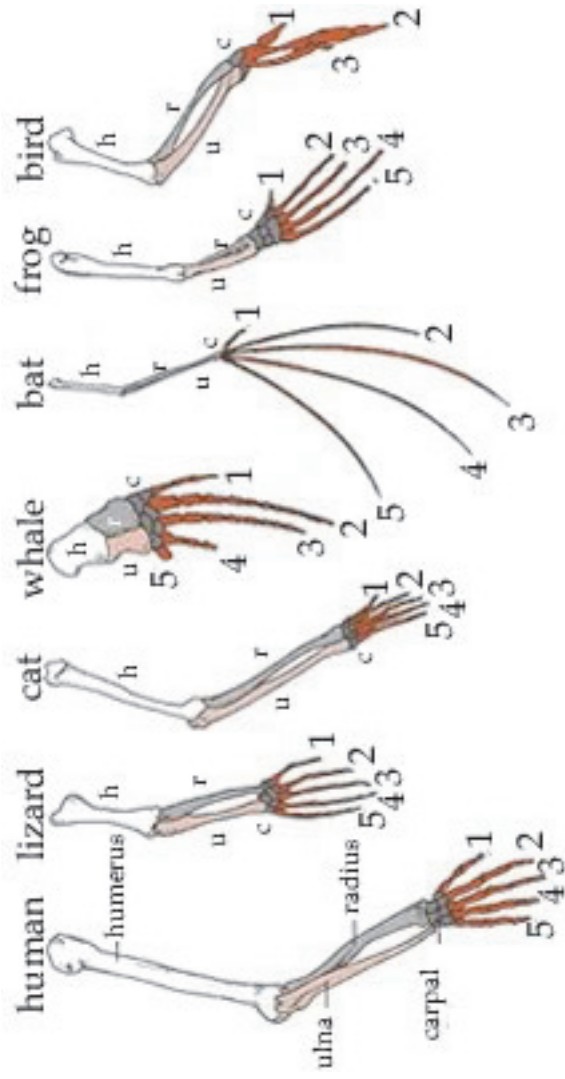
Hist. Sci., xlix (2011)

**THE COMPARATIVE AND THE EXEMPLARY:
REVISITING THE EARLY HISTORY OF MOLECULAR
BIOLOGY**

Bruno J. Strasser
Yale University

Soraya de Chadarevian
UCLA

The comparative...



Generalizzazione inferita attraverso la comparazione tra differenti specie ed organismi

...and the Exemplary



- “What is true for *E. coli* must be true also for an elephant” - Jacques Monod
- Generalizzazione inferita a partire da un caso esemplare

Storia Naturale nel XXI secolo

Science 24 October 2008:
Vol. 322 no. 5901 pp. 537–538
DOI: 10.1126/science.1163399


PERSPECTIVE

GENETICS

GenBank – Natural History in the 21st Century?

Bruno J. Strasser

[< Prev](#) | [Table of Contents](#) | [Next >](#)



**Studies in History and Philosophy of Science
Part C: Studies in History and Philosophy of
Biological and Biomedical Sciences**

Volume 43, Issue 1, March 2012, Pages 85–87

Data-Driven Research in the Biological and Biomedical Sciences
On Nature and Normativity: Normativity, Teleology, and Mechanism in
Biological Explanation

Discussion

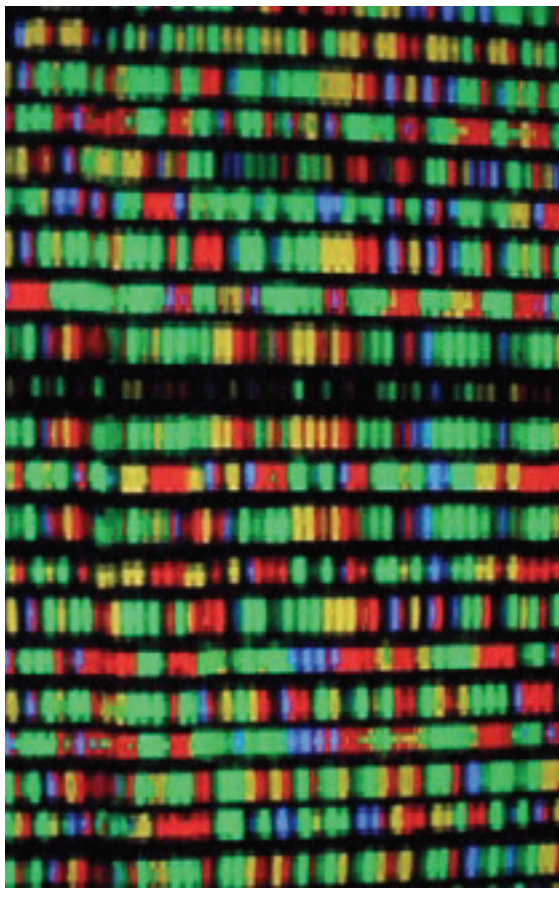
**Data-driven sciences: From wonder cabinets to electronic
databases**

Bruno J. Strasser ✓

Tipi di collezioni



Butterfly collection at NHM in London



Digital collection representing the human genome

Cos'è la svolta computazionale?

- **Biologia computazionale** ≠ **computer + laboratorio**
- **Vecchi approcci e metodi ma ri-usati secondo un nuovo e differente modo di pensare/fare**

Un nuovo stile di pensare: la mappa

Perspective

A Turning Point in Cancer Research: Sequencing the Human Genome

RENATO DULBECCO

Science 1986



1991

“Map-thinking” e la Big Science



Una metafora



Una visione complementare



“30,000-foot view” (Vogelstein et al. 2013)

Il ruolo dei database

- Database e strumenti computazionali sviluppati per organizzare e ricercare dati diventano uno strumento di ricerca legittimo al pari dell'approccio sperimentale
- Non è un mutamento di paradigma
- La visione molecolare permane

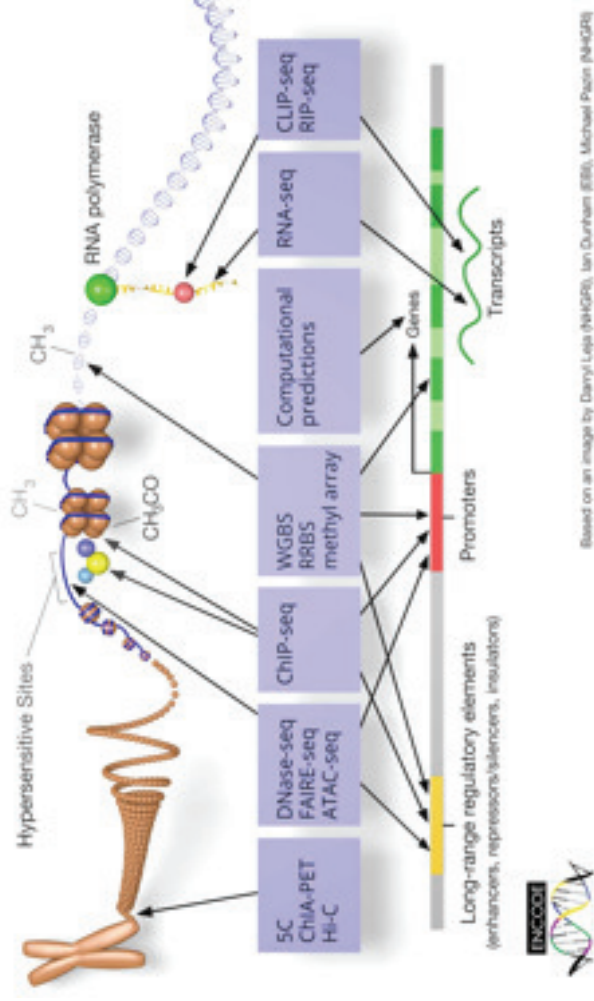
Studiare i database come mappe

- Una mappa può essere vista come un **peculiare modello**, con certe caratteristiche:
- **Ordinare/riordinare (vs manipolazione)**
- **Collocare/comparare (vs distorsione relazionale)**

Fare scienza usando modelli

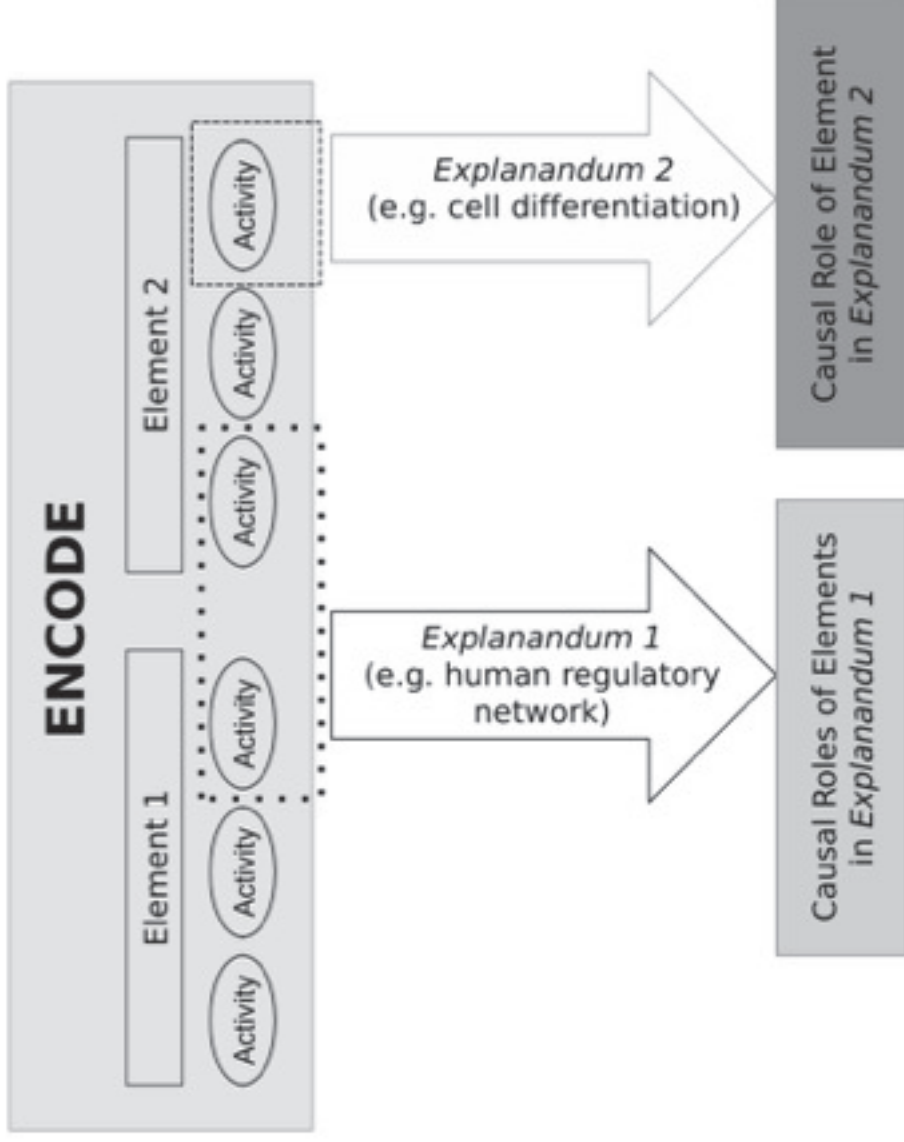
- I database sono modelli/mappa
- Si possono fare ricerche “sperimentali” in *silico*, interagendo su tali modelli
- Cambia il significato di cosa vuol dire “sperimentare”

Esempi di ricerca I



Un'enciclopedia degli elementi funzionali del genoma

Strategia epistemica di ENCODE



Germain, Ratti and Boem, *Biology&Philosophy*, 2013

Esempi di ricerca II

Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes

Tae-Min Kim¹, Ruibin Xi^{1,2}, Lovelace J. Luquette¹, Richard W. Park¹,

Mark D. Johnson³ and Peter J. Park^{1,4,5,6}

- Gli scienziati cercano, nei database, associazioni tra differenti metadati al fine di scoprire macro-regolarità
- Le macro-regolarità sono, di fatto, predizioni nel senso che forniscono un'aspettativa su cosa potrà essere più probabilmente osservato in contesti simili

Cosa cambia?

**Il map-thinking non è un rifiuto della ricerca connessioni
causali**

**Esso tuttavia cambia il ruolo che la dimensione
sperimentale ha nella ricerca**

**L'esperimento, da esplorativo diventa sempre più
confermativo (*pace Popper*)**

CONCLUSIONI 1

- Non esiste una vera opposizione epistemica tra **data-driven** e **hypotheses-driven**
- Le **differenze metodologiche** esistono ma possono essere ricondotte alla **stessa cornice teorica**
- La **biologia computazionale** è uno strumento per rendere più **efficiente** e **preciso il processo di scoperta** in seno alla **biologia molecolare**
- **Complementarità**

Uno sguardo sul futuro...



Bruce Alberts is Editor-in-Chief of Science.

EDITORIAL

The End of “Small Science”?

I AM PROMPTED TO WRITE THIS EDITORIAL BY THE RELEASE OF 30 PAPERS THIS MONTH FROM THE ENCODE Project Consortium. This decade-long project involved an international team of 442 scientists who have compiled what is being called an “encyclopedia of DNA elements,” a comprehensive list of functional elements in the human genome. The detailed overview is expected to spur further research on the fundamentals of life, health, and disease. ENCODE exemplifies

Conclusione 2

- L'imposizione di certe metodologie potrebbe avere un effetto su come la scienza è organizzata
- Certi studi possono essere fatti unicamente utilizzando approcci computazionali
- Inoltre l'unificazione del metodo (contro Feyerabend 1975 e Dupré 1993) costituisce una minaccia al pluralismo epistemico
- Un criterio di demarcazione “dal basso”?

Una vecchia provocazione?

“All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced sciences such as gravitational astronomy, the word "cause" never occurs. [...] To me it seems that philosophy ought not to assume such legislative functions, and that the reason why physics has ceased to look for causes is that, in fact, there are no such things. The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do not harm”

Bertrand Russell, 1912

Ringraziamenti

- I miei vecchi colleghi al FOLSATEC
(Pierre-Luc Germain e Emanuele Ratti)
- Prof.ssa Raffaella Campaner e il Prof.
Roberto Brigati
- Voi tutti



"That's all Folks!"