



# VENTIMILA GENI SOTTO I MARI DEI TROPICI

Fiorisce un nuovo approccio alla biologia per cercare di capire come dalle interazioni molecolari si originano le proprietà degli organismi

DANIEL SEGRÈ

**L**A GRANDE MAPPA DELLA BIOLOGIA è popolata da un' enorme quantità di punti di riferimento, i geni e i genomi dei sistemi viventi, nonché da un numero ancora maggiore di reazioni e interazioni che li congiungono, determinandone la dinamica su scala fisiologica ed evolutiva. Quest'immensa quantità di informazione è il pane quotidiano di molti «system biologists», una nuova stirpe di scienziati interdisciplinari che cercano di capire, spesso con l'aiuto di modelli al computer, come migliaia di singole interazioni molecolari possano dare origine collettivamente alle proprietà macroscopiche delle cellule viventi. Le applicazioni di questi approcci globali allo studio dei sistemi biologici sono molteplici, dalla capacità di individuare su larga scala geni responsabili di disfunzioni della crescita cellulare, all'ingegnerizzazione di microrganismi per scopi medici o ambientali. Per molti, peraltro, l'ambizione di comprendere i processi fondamentali della cellula e l'evoluzione dei sistemi viventi, costituisce di per sé una motivazione sufficiente.

A tutt'oggi duecentosessantuno genomi sono disponibili in rete<sup>1</sup>. La maggior parte sono genomi di minuscoli procarioti, che comprendono noti organismi patogeni, quale il *Mycobacterium tuberculosis*, ma anche organismi meno famosi, individuati di recente, quali il *Prochlorococcus marinus*, che permea gli oceani tropicali e subtropicali, contribuendo con la sua fotosintesi a cicli molecolari di carbonio e ossigeno su scala terrestre<sup>2</sup>. Ai genomi di procarioti già decodificati se ne aggiungono circa settecento in procinto di essere sequenziati. Seppure notevoli, questi numeri rimangono ridicolmente piccoli rispetto ai miliardi di specie di microrganismi che si stima esistano sul nostro pianeta. I microbi sono capaci di adattarsi efficacemente a condizioni estreme di temperatura, acidità, composizione chimica. Ciò fa sì che essi popolino ogni angolo della Terra, partecipando tanto a simbiosi stupefacenti all'interno di altri organismi, quanto al destino geochimico del nostro pianeta.

Sbrogliare la complicata matassa di reazioni e interazioni anche per uno solo di questi organismi, per non parlare di un ecosistema, costituisce una grossa sfida, che probabilmente terrà occupati biologi, chimici, fisici, ingegneri e matematici per diversi anni. Non a caso, il Dipartimento dell'Energia (DOE) degli Stati Uniti ha recentemente dedicato allo studio dei microorganismi di importanza ecologica un grosso programma di ricerca e finanziamenti, denominato Genomes to Life (GtL), «dai genomi alla vita»<sup>3</sup>. L'aspettativa è che la ricerca in questa direzione abbia un impatto notevole sui metodi di produzione di energia, sulla lotta contro i danni ambientali, e sul controllo di cambiamenti climatici su scala globale. E per raggiungere tali obiettivi, molti strumenti della Systems Biology potrebbero risultare essenziali.



MODIS/NASA

Una gigantesca massa di fitoplancton nel Golfo di Biscaglia fotografata dal satellite Aqua della Nasa il 16 maggio del 2004.

### Cacciatori di meta-microbi

«La Terra è un grande vivente», annotava cinquecento anni fa Leonardo Da Vinci nei suoi *Pensieri*. Non si riferiva certo a ecosistemi microbici, ma l'espressione vi si addice perfettamente. Le innumerevoli specie di microrganismi presenti sul nostro pianeta, secondo alcune stime, costituiscono più del 60% della materia vivente. All'efficiente metabolismo autotrofo di alcuni di questi microrganismi e dei loro antenati, è dovuta la presenza di ossigeno molecolare nella nostra atmosfera, e dalla loro ingegnosa sintesi di molecole organiche dipende l'esistenza dell'intera catena alimentare. La possibilità di isolare e coltivare singole specie di microrganismi è di fondamentale importanza per molti aspetti della biologia e della medicina. Ma la maggior parte dei microbi sono difficili da coltivare in laboratorio, proprio perchè la loro sopravvivenza dipende in modo complesso dall'interazione con altri microrganismi e con l'ambiente. Craig Venter, pioniere del shotgun sequencing, il metodo che ha permesso di completare la lettura del genoma umano, ha messo recentemente in evidenza in modo dettagliato la complessità delle comunità microbiche marine <sup>4</sup>. Un gruppo da lui coordinato ha applicato lo shotgun sequencing a popolazioni di microrganismi raccolti filtrando grosse quantità d'acqua nel Mar dei Sargassi. Nonostante l'impossibilità di ricomporre in dettaglio i singoli genomi delle migliaia di specie presenti in ogni campione, l'informazione ottenuta fornisce un quadro affascinante sulla diversità e complessità degli ecosistemi microbici. I campioni hanno rivelato almeno 1.800 specie di microrganismi, contenenti più di un milione di geni, di cui circa settantamila completamente nuovi. L'esperimento del Mar dei Sargassi è solamente la prima tappa di un ambizioso progetto che produrrà biblioteche genomiche di microrganismi marini per un gran numero di località attorno al mondo. Oltre a Venter, altri cacciatori di meta-microbi, stanno caratterizzando il contenuto genetico e le proprietà fisiologiche di varie comunità microbiche <sup>5,6</sup>. La quantità di dati disponibili è dunque destinata a crescere rapidamente. Resta da capire se la nostra capacità di utilizzare fruttuosamente questi dati si svilupperà altrettanto in fretta.

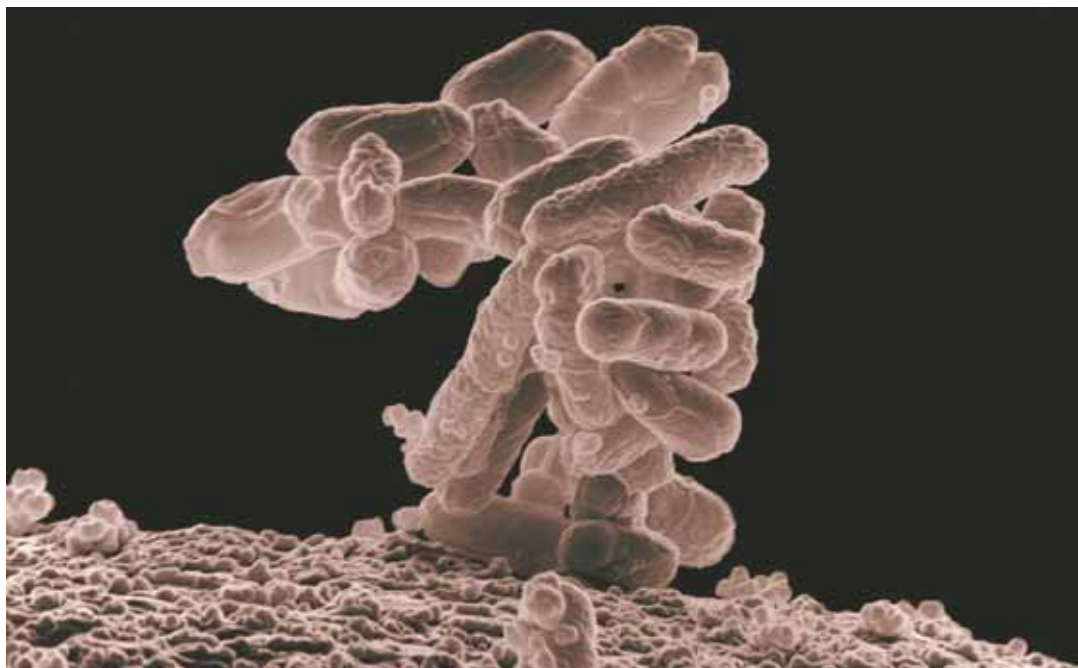
### Moduli e gerarchie

L'integrazione di metodi teorici e sperimentali mirati a trattare l'incredibile complessità dei sistemi biologici ha portato negli ultimi anni all'inizio di una vera e propria rivoluzione scientifica. La portata di tale rivoluzione è dimostrata dal fiorire di nuovi istituti, dipartimenti e riviste esplicitamente dedicati a quella che viene considerata quasi una nuova disciplina, la Systems Biology, ovvero la biologia dei sistemi <sup>7</sup>. La biologia molecolare permette di capire nei dettagli il funzionamento di singole componenti, come gli enzimi e i geni che li codificano. Ma per poter spiegare e

manipolare le proprietà di un intero organismo, c'è bisogno di formalizzare la conoscenza biologica, e di passare da descrizioni qualitative ad analisi e modelli quantitativi. Una delle caratteristiche della systems biology è l'uso di metodi sperimentali su larga scala («high throughput»). Oltre alle sequenze di DNA e alle mappe metaboliche, grossi database collezionano enormi quantità di dati relativi a livelli di mRNA e peptidi, all'interazione fra proteine, e fra proteine e DNA, nonché immagini al microscopio e sequenze video. Come scrive George Church in un editoriale della rivista appena inaugurata *Molecular Systems Biology* <sup>8</sup>, è fondamentale che dati sperimentali e modelli matematici utilizzino rappresentazioni standard, così da permettere la navigazione e ricerca computerizzata nei database, e quindi la costruzione modulare di modelli sempre più completi. Uno standard che si sta diffondendo rapidamente è il linguaggio Sbm1, o systems biology markup language, il corrispondente biologico dell'Html in cui sono scritte le pagine di Internet <sup>9</sup>. Proprio come in Internet, diversi gruppi sono in grado di scambiarsi modelli di vari organismi, e utilizzarli ognuno nel proprio simulatore preferito. Il programma BioSpice, ispirato al celebre software Spice per la simulazione di circuiti integrati, è un prototipo di questa nuova tecnologia, frutto di una collaborazione fra diversi centri di ricerca, finanziati dalla stessa Darpa che a suo tempo creò Internet <sup>10</sup>.

È ovvio domandarsi se questa struttura modulare delle reti biologiche sia semplicemente frutto della nostra necessità di schematizzare e capire, o sia una proprietà inerente dei sistemi viventi. Non sempre è facile cogliere la differenza. Sofisticati metodi di inferenza Bayesiana <sup>11</sup> e di clustering <sup>12</sup> vengono utilizzati per analizzare dati complessi, quali i livelli di espressione di mRNA. È spesso possibile giungere a rappresentazioni gerarchiche e modulari, che riducono un grande numero di geni a un ristretto insieme di classi funzionali direttamente legate al comportamento macroscopico della cellula <sup>13</sup>. Moduli di geni sono solitamente accomunati dai fattori di trascrizione che li regolano, a loro volta controllati da «master regulators» che rispondono a segnali esterni.

L'analisi sistematica di moduli funzionali in reti biologiche ha portato inoltre alla scoperta dell'esistenza di circuiti che ricorrono molto più frequentemente di quanto ci si aspetterebbe in una rete casuale <sup>14</sup>. In analogia con circuiti elettronici, si è visto che tali micro-circuiti svolgono funzioni simili a componenti elettroniche modulari, come filtri, amplificatori di segnale, integratori. Oltre a permettere una comprensione molto più profonda della logica della cellula, la scoperta di questi motivi pone affascinanti domande sull'evoluzione delle reti biologiche, e sulla straordinaria capacità della selezione naturale di giungere a soluzioni ingegneristicamente impeccabili.



ERIC RIBE / USDA/ARS

Una colonia di *Escherichia coli* fotografata al microscopio elettronico.

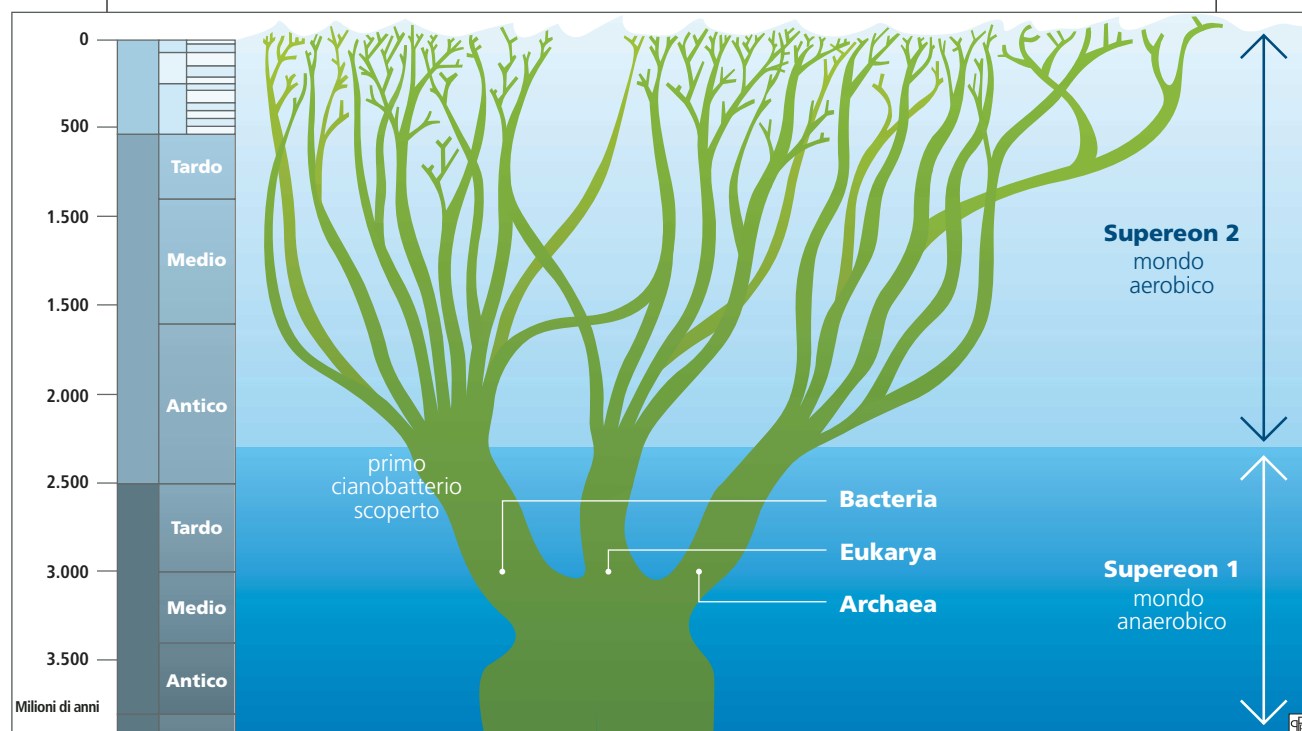
### Reti metaboliche

Lo studio di reti metaboliche su scala cellulare è un'altra delle innovazioni della systems biology. Nello studio di microrganismi di rilevanza ecologica, è particolarmente importante valutare il bilancio di massa ed energia, sia all'interno della cellula, che negli scambi con l'ambiente esterno. L'applicazione di modelli tradizionali di cinetica chimica allo studio della dinamica di reti biochimiche su scala cellulare richiederebbe l'uso di centinaia di equazioni differenziali, e migliaia di costanti cinetiche e altri parametri, attualmente ignoti e difficili da misurare.

Per ovviare a questo problema sono stati recentemente sviluppati dei metodi, basati su tecniche di programmazione matematica, originariamente utilizzate nell'ambito dell'ingegneria chimica e metabolica con lo scopo di ottimizzare processi biochimici industriali. L'idea fondamentale di uno di questi metodi, chiamato Flux Balance Analysis<sup>15</sup>, è che si può trattare una cellula come una minuscola fabbrica, che riceve materiale dall'esterno, ed esporta un prodotto finale. I processi intermedi tramite cui i materiali vengono trasformati sono le reazioni metaboliche. In una popolazione di cellule mantenute in stato stazionario, ogni reazione può essere associata con un flusso, corrispondente alla velocità della reazione enzimatica. Questo passo costituisce un'approssimazione drastica, tramite la quale si semplifica il problema, rinunciando ad analizzare direttamente gli aspetti cinetici del metabolismo. Stato stazionario, peraltro, non significa equilibrio: si tratta sempre di un sistema lontano dall'equilibrio, a cui vie-

ne fornita energia con flusso costante. Come le correnti in un circuito elettrico, le varie velocità sono vincolate da leggi di conservazione lineari, che definiscono uno spazio di stati possibili per la cellula.

All'interno di tale spazio di regimi metabolici permessi, è possibile cercare stati che hanno la proprietà di massimizzare una certa «funzione obiettivo». In una tipica applicazione di ingegneria metabolica la funzione obiettivo da massimizzare potrebbe essere la produzione di un antibiotico o di un aminoacido per l'industria alimentare. Il modello, in questo caso, può fornire informazioni preziose su quali enzimi debbano essere aggiunti, rimossi o modificati per aumentare la produzione<sup>16</sup>. Nello studio di microrganismi naturali si cerca invece di capire quale possa essere considerata la funzione obiettivo intrinseca della cellula, che la selezione naturale ha ciecamente massimizzato o minimizzato. In altre parole, ci si chiede se sia possibile rappresentare efficacemente in una singola espressione matematica, una combinazione lineare di flussi, il fitness darwiniano di una cellula. Una funzione obiettivo utilizzata di frequente nello studio di microrganismi è la massimizzazione di un flusso che ne descrive la crescita, ovvero della velocità di assemblaggio del prodotto più importante della fabbrica cellulare, cioè un'altra cellula. L'ipotesi di massimizzazione della velocità di produzione di biomassa riflette la competizione che in alcuni sistemi di batteri potrebbe favorire quello capace di crescere più velocemente. In generale, diverse cellule possono richiedere l'uso di combinazioni dinamiche di diverse funzioni obiettivo, e la pro-



La varietà microbica negli oceani è frutto di oltre 3,8 miliardi di anni di evoluzione. Nei primi 1,5 miliardi di anni si sarebbero selezionati i processi metabolici fondamentali, culminati con la fotosintesi nei cianobatteri che ha portato all'ossidazione di oceani e atmosfera. Da allora, per proteggere i processi cruciali, i microbi hanno evoluto innumerevoli adattamenti generando l'immensa varietà molecolare odierna.

babilità di sopravvivenza può dipendere da una miriade di funzioni non metaboliche. Ma la capacità di rappresentare in uno modello al computer l'intera rete metabolica di una cellula, se pure in modo approssimato, apre nuove interessanti prospettive.

### Ingegneria genetica virtuale

Una delle applicazioni più utili dei modelli al computer di flussi metabolici è la possibilità di fare esperimenti virtuali di ingegneria genetica senza allontanarsi dalla scrivania. Perturbazioni come la delezione di un gene sono facilmente simulate aggiungendo un opportuno vincolo che fissa a zero la velocità del corrispondente flusso metabolico <sup>17</sup>. Algoritmi standard di programmazione lineare possono calcolare in una frazione di secondo l'effetto di tali mutazioni sulla crescita della cellula, nonché sui flussi di tutte le reazioni metaboliche del sistema. In analogia con una fabbrica, nuovamente, questo equivale a chiedersi come cambierebbe la velocità di produzione se un certo reparto smettesse di funzionare. In questo modo è possibile chiedersi quali geni siano essenziali per la sopravvivenza di un organismo in un determinato ambiente. Il confronto di queste previsioni con dati sperimentali permette di capire se nella ricostruzione al computer dell'organismo

manchi qualche reazione, o ci siano reazioni superflue, o ancora se la funzione obiettivo utilizzata sia sbagliata. In media, per organismi molto studiati come l'*Escherichia coli* e il *Saccharomyces cerevisiae* – un lievito, eucariote unicellulare – si è in grado di predire correttamente l'effetto fenotipico della rimozione di un gene circa nell'ottanta per cento dei casi <sup>18</sup>. L'accordo fra modello ed esperimenti migliora se anziché ipotizzare che la rete metabolica perturbata sia in grado di adattarsi alla mutazione in modo ottimale, si assume uno spostamento minimo, nello spazio dei flussi, rispetto allo stato della cellula non perturbata <sup>19</sup>. In altre parole, quando la nostra fabbrica metaforica viene privata di un reparto, non è detto che tutta la catena di produzione si possa riarrangiare in tempi brevi per compensare in modo ottimale questa perdita. Più in generale, confrontando dati empirici e simulazioni, si è visto che l'ipotesi di crescita massima è ragionevole in alcuni casi, ma non in altri. Si sa per esempio che in assenza di nutrienti, la strategia migliore può essere quella di minimizzare il consumo di energia, e di sospendere l'attività di replicazione. Nei modelli al computer, la rimozione di singoli geni può essere facilmente estesa allo studio di doppie mutazioni <sup>20</sup>. Supponiamo di rimuovere, in due esperimenti diversi, due singoli reparti del-

la fabbrica cellulare. Ogni rimozione riduce di un po' la produzione. Ora in un terzo esperimento possiamo rimuovere i due reparti simultaneamente. Dall'effetto combinato, ovvero dal livello di epistasi, è possibile determinare quanto simili siano le funzioni dei due reparti. Un caso estremo è il collasso del sistema, se i due reparti sono il *back-up* l'uno dell'altro.

È interessante chiedersi se modelli di reti metaboliche come quelli sviluppati recentemente per singoli organismi possano essere utili nello studio di comunità microbiche. Uno dei primi passi, ancora in fase di sviluppo, è quello di poter costruire in modo automatizzato la rete metabolica per qualsiasi organismo di cui si possiede un genoma con annotazione delle funzioni enzimatiche<sup>21,22</sup>. Ma uno degli aspetti più affascinanti delle comunità microbiche, è che i loro «abitanti» devono essere caratterizzati da obiettivi metabolici diversi da quelli di singoli organismi, che tenuto conto della cooperazione fra le varie specie. Massimizzare la propria velocità di crescita a spese di altri non è una strategia vincente per un organismo che deve vivere in simbiosi con organismi

circostanti. Il problema della cooperazione metabolica fra diverse cellule è complesso, ma può essere affrontato anche a livello più astratto<sup>23</sup>. Il fatto che non esista apparentemente alcun motivo per cui una strategia di uso efficiente di risorse comuni debba evolversi spontaneamente e vincere su una tattica egoista, è un dilemma ben conosciuto nella teoria dei giochi (tragedy of the commons). Modelli matematici semplificati dimostrano come, in linea di principio, popolazioni di cellule dal metabolismo veloce ma inefficiente possano essere soppiantate da popolazioni di cellule metabolicamente lente ma parsimoniose, in condizioni opportune di diffusione e distribuzione spaziale. Come i microorganismi che convivono negli oceani, anche noi dobbiamo continuare a cooperare se vogliamo sperare di raggiungere una comprensione adeguata dei sistemi biologici complessi. La strada è ancora lunga, ma la posta in gioco potrebbe includere la salute del nostro pianeta.

Daniel Segrè, Boston University, [dsegre@bu.edu](mailto:dsegre@bu.edu)  
<http://prelude.bu.edu/darwin2005/>

#### Note

- Bernal, A., Ear, U. & Kyrpides, N. Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Res* 29, 126-7 (2001).
- Rocap, G. et al. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424, 1042-7 (2003).
- Frazier, M. E., Johnson, G. M., Thomassen, D. G., Oliver, C. E. & Patrinos, A. Realizing the potential of the genome revolution: the genomes to life program. *Science* 300, 290-3 (2003).
- Venter, J. C. et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66-74 (2004).
- Tyson, G. W. et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37-43 (2004).
- Hallam, S. J. et al. Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* 305, 1457-62 (2004).
- Westerhoff, H. V. & Palsson, B. O. The evolution of molecular biology into systems biology. *Nat Biotechnol* 22, 1249-52 (2004).
- Church, G. M. From systems biology to synthetic biology. *Molecular Systems Biology* [www.molecularsystemsbiology.com](http://www.molecularsystemsbiology.com) (2005).
- Hucka, M. et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524-31. (2003).
- Garvey, T. D., Lincoln, P., Pedersen, C. J., Martin, D. & Johnson, M. BioSPICE: access to the most current computational tools for biologists. *Omic* 7, 411-20 (2003).
- Segal, E. et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34, 166-76 (2003).
- Ihmels, J., Levy, R. & Barkai, N. Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*.

- Nat Biotechnol* 22, 86-92 (2004).
- Grosu, P., Townsend, J. P., Hartl, D. L. & Cavalieri, D. Pathway Processor: a tool for integrating whole-genome expression results into metabolic networks. *Genome Res* 12, 1121-6 (2002).
- Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* 298, 824-7. (2002).
- Schilling, C. H., Edwards, J. S., Letscher, D. & Palsson, B. O. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* 71, 286-306. (2000).
- Stephanopoulos, G., Alper, H. & Moxley, J. Exploiting biological complexity for strain improvement through systems biology. *Nat Biotechnol* 22, 1261-7 (2004).
- Edwards, J. S. & Palsson, B. O. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* 97, 5528-33. (2000).
- Famili, I., Forster, J., Nielsen, J. & Palsson, B. O. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc Natl Acad Sci USA* 100, 13134-9 (2003).
- Segrè, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* 99, 15112-15117 (2002).
- Segrè, D., Deluna, A., Church, G. M. & Kishony, R. Modular epistasis in yeast metabolism. *Nat Genet* 37, 77-83 (2005).
- Covert, M. W. et al. Metabolic modeling of microbial strains in silico. *Trends Biochem. Sci.* 26, 179-86. (2001).
- Segrè, D. et al. From annotated genomes to metabolic flux models and kinetic parameter fitting. *Omic* 7, 301-16 (2003).
- Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* 292, 504-7 (2001).