

ENTENDRE EL GENOMA: LES XARXES BOOLEANES COM A MODEL

Escrit per:

Albert Cardona i Torrens

Doctorand del Dept. de Genètica
Universitat de Barcelona

El genoma humà ha estat seqüenciat. Aquest fet, tot i l'extraordinari esforç desplegat per a la seva consecució, només representa una mera recopilació de gairebé tota la informació necessària per fer un ésser humà. Fins al moment es coneixen amb diferent grau de detall uns pocs centenars dels més de 30.000 gens detectats. Certament aquesta recopilació accelerarà el procés de caracterització de cadascun d'ells, que es traduirà en aplicacions fabuloses a la medicina. Però per altra banda, la comprensió del fenomen general de la vida i en particular del funcionament del genoma no haurà progressat. Aquesta comprensió a un nivell més elevat no és només interessant a nivell teòric o filosòfic, sinó que resulta imprescindible per a la comprensió dels fenòmens multigènics com són el desenvolupament i les malalties multifactorials. Per a tal objectiu es fa necessari un nou plantejament. Es demana dels investigadors l'abandó de l'enfocament reduccionista per emprendre el pas de considerar la unitat mínima d'estudi el genoma en si, un sistema complex amb propietats emergents resultants de la interacció de les seves parts.

L'anàlisi detallada de les regions de regulació - els promotors- de cada gen del genoma humà proveirà informació fonamental per conèixer les

interaccions entre gens respecte a la regulació de l'expressió (de la mateixa manera, l'anàlisi de la proteïna o ARN codificat pel gen indicarà quines interaccions amb altres productes gènics afectaran l'execució de la funció del gen). Tanmateix, examinar un sol genoma no és només ja una tasca de proporcions gegantines, sinó que per poder interpretar les dades es fa necessari conèixer abans quines propietats tenen, genèricament, els genomes. L'única via de què disposen els investigadors en el present és la modelització matemàtica mitjançant medis informàtics. Quina mena de model pot concebre's per explorar les propietats d'un sistema tal com un genoma? El model ha de considerar, fonamentalment, els gens i les interaccions entre ells mateixos. Els gens es caracteritzen per codificar un RNA missatger, el qual codifica, a la vegada, un producte gènic que consisteix en RNA o proteïna. De manera simplificada, l'estat d'expressió del gen serà actiu si s'estan formant els productes gènics pels quals codifica i inactiu en cas contrari. Aquest estat d'expressió és regulat per un o més productes gènics d'altres gens, els quals a la vegada tenen regulada l'expressió per altres gens, i així successivament. El model necessita, doncs, assignar a cada gen una variable que reculli l'estat d'expressió, i la manera més simple

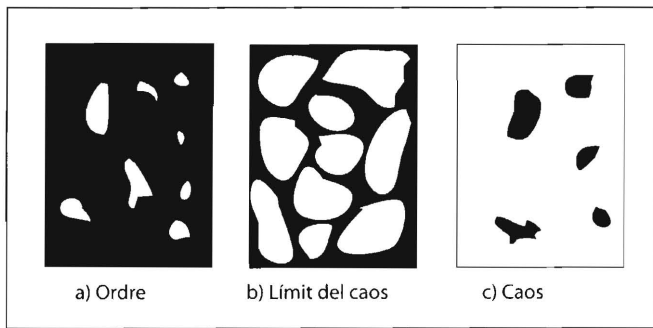


Fig. 1. Representació bidimensional d'una xarxa booleana. Cada píxel representa un node, que connecta amb els nodes immediatament adjacents. A: en l'estat ordenat, la major part dels nodes tenen fixat el seu estat d'activació (negre); uns pocs grups de nodes (illes blanques) mostren fluctuacions dels seus estats. B: al límit del caos, els grups de nodes fluctuants creixen sense arribar a contactar; l'efecte d'una alteració sobre un node qualsevol tindria amb gran probabilitat un efecte local. C: caos. Un mar de nodes amb l'estat d'activació fluctuant envolta algunes illes de nodes estables. Qualsevol alteració, fins i tot sobre un sol node, es transmetrà fàcilment per tota la xarxa.

d'acomplir-ho és fer aquesta variable booleana: actiu o inactiu. Si el gen és actiu afectarà activament l'estat d'expressió d'altres gens; si és inactiu ho farà passivament, per absència. Aquesta consideració és important perquè un gen pot o bé activar o bé inhibir l'expressió d'un altre gen. El model pot recollir aquesta interdependència de l'estat d'expressió gènica si es construeix com una xarxa que consti de tants nodes com gens, on l'estat d'activació de cada node depengui de l'estat d'activació d'altres nodes determinats. Cada node es caracteritza per unes entrades -l'estat d'activació d'altres nodes- i una sortida -l'estat d'activació propi- computada a partir de les entrades mitjançant una funció.

El pioner en modelitzacions d'aquesta mena fou i segueix essent Stuart Kauffman, investigador del Santa Fe Institute. El seu primer model constava d'una xarxa construïda amb nodes que obeeixen funcions booleanes, de manera que l'estat de cada node pot ser tan sols actiu o inactiu. L'estat dels nodes es calcula sincrònicament a partir de les entrades de cada node en cada instant de temps. Tots els nodes tenen el mateix nombre d'entrades (K), assignades a l'atzar d'entre totes les possibles, una funció booleana assignada a l'atzar i un estat inicial de cada node assignat també a l'atzar. La dinàmica del sistema es caracteritza per la precessió dels diferents estats d'activació de la xarxa (definitos com l'estat d'activació de tots els nodes en un instant) a través del temps. Kauffman va trobar que les xarxes booleanes podien caure en dinàmiques extraordinàriament ordenades, al límit del caos o bé caòtiques (Fig. 1), depenent del valor de K , és a dir, del nombre d'entrades dels nodes i del tipus de funcions booleanes.

Una xarxa construïda amb $K=2$ entrades per node, un cop iniciada, transita uns quants estats (run-in) fins a caure ràpidament en un atractor: un cicle d'estats que es manté indefinidament en el temps. És el règim ordenat. En aquest règim ordenat, l'alteració de l'estat d'un node a l'atzar produeix allaus (reaccions en cadena) de canvis en l'estat d'activació dels nodes, la magnitud de les quals segueix una llei potencial (moltes allaus molt petites i poques de grans). Per exemple, alterant a l'atzar l'estat d'activació d'un node cada vegada, la

majoria de vegades la dinàmica no sortirà de l'atractor on roman i unes poques en sortirà breument per tornar-hi a caure tot seguit; algun cop, però, saltarà a un atractor diferent (Fig. 2). El nombre i la llargada dels atractors són propietats mesurables i característiques de cada xarxa. És certament molt sorprenent que una xarxa de per exemple 10.000 nodes amb $K=2$, generada a l'atzar, orbiti entorn de només uns 100 estats diferents dels $2^{10.000}$ possibles! En paraules de Kauffman, autèntic ordre gratuït. Força sorprenent és també que la gran majoria de nodes quedi en estat "congelat", és a dir, el seu estat quedi fixat en actiu o inactiu, cosa que Kauffman proposa que correspon a gens de "housekeeping". Els nodes no congelats, que sofreixen canvis dins l'atractor, s'agrupen en diferents "illes" envoltades de nodes congelats. Cada illa pot presentar diferents atractors i per tant, per combinatòria, es generen molts atractors generals a partir d'uns pocs de locals.

Per valors de K superiors a 2 la xarxa booleana resideix en règim caòtic: flueix per estats diferents que se segueixen indefinidament sense cap patró. En règim caòtic, l'alteració de l'estat d'activació d'un sol node produeix efectes molt variats; existeix una gran sensibilitat a les condicions inicials

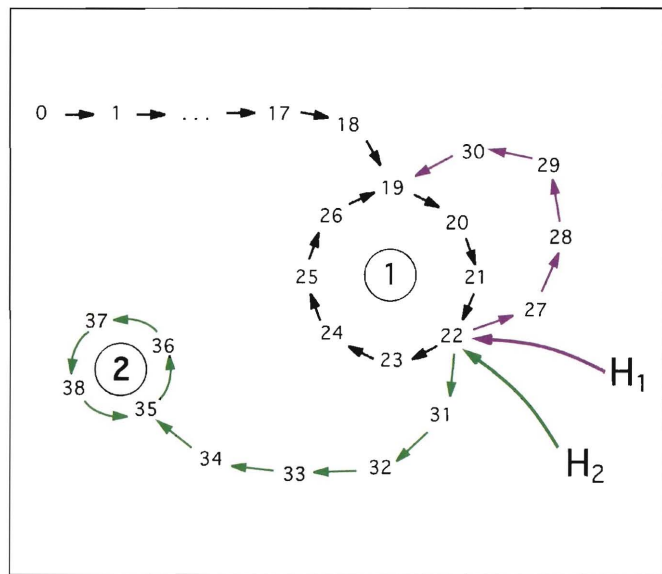


Fig. 2. Representació esquemàtica del flux temporal d'una xarxa que roman en règim ordenat. Cada nombre representa un estat d'activació diferent de la xarxa. Després de fluir per 18 estats (run-in), la xarxa cau en l'atractor 1, on roman indefinidament. L'alteració H1 fa sortir momentàniament la xarxa de l'atractor, però no resulta suficient per a escapar-ne. L'alteració H2, més potent, fa sortir la xarxa de l'atractor 1 i, després de fluir per quatre estats nous, cau en un segon atractor, on romandrà fins la pròxima alteració. Dues alteracions diferents sobre un mateix estat tenen conseqüències diferents.

(és el famós "efecte papallona", segons el qual el batejar de les ales d'una papallona a les illes Canàries pot produir un huracà a Florida). La majoria de petites alteracions tenen efectes que s'estenen sobre molts nodes de la xarxa.

Entre $K=2$ i $K=3$ sembla existir un abisme. Els genomes reals, però, tenen moltíssims gens amb més de 2 entrades i tot i així no sembla probable, a priori, que segueixin patrons aleatoris

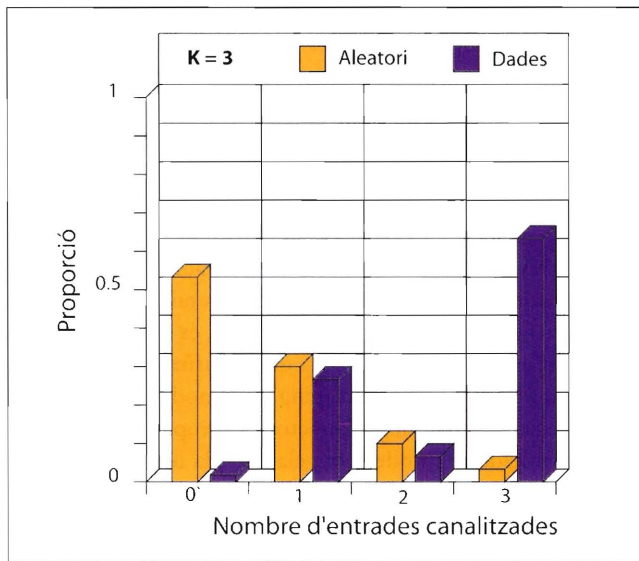


Fig. 3. Distribució del nombre d'entrades canalitzades en 53 gens eucariòtics dotats de $K = 3$ entrades (dades), versus el nombre esperat d'entrades canalitzades si les normes de control genètic fossin escollides a l'atzar d'entre les 256 possibles funcions booleanes de 3 entrades (aleatori). La majoria dels gens reals de 3 entrades tenen les tres canalitzades. (Kauffman 2003).

(l'anàlisi de l'expressió gènica per microarrays mostra que cada tipus cel·lular té un patró característic i constant). Així doncs, com podria una xarxa booleana emular l'elevada connectivitat del genoma sense caure en règims caòtics? Existeixen dos paràmetres que poden dur una xarxa cap al règim ordenat fins i tot quan K és superior a 2, i ambdós representen un biaix en les funcions booleanes de cada node. El primer és la canalització de les funcions booleanes: consisteix en que una o més de les entrades d'un node determini per si sol el resultat de la funció, és a dir l'estat d'activació del node. El segon paràmetre, anomenat P , consisteix en que el nombre major d'ocurrències de les entrades (per exemple, dues actives sobre una d'inactiva en un node de $K=3$ entrades) designi l'estat d'activació del node. Les observacions fetes per Kauffman i Steve Harris sobre dades de gens reals han demostrat que les cèl·lules prefereixen les entrades canalitzades (Fig. 3). És més, Raeymaekers (2002) ha mostrat que a) un biaix en l'elecció de les funcions booleanes que governen cada node, de manera que només quedin les que tenen un significat biològic, incrementa enormement l'habilitat de les xarxes per residir en règim ordenat quan K és major que 2; i b) l'elecció de xarxes on hi ha una majoria (però no una totalitat) de factors inhibidors o activadors també afavoreix el manteniment de les mateixes en règim ordenat. Així doncs, és possible generar una xarxa booleana amb K superior a 2, que resideixi en règim ordenat (Fig. 4 i 5), i que per tant modeli més fidelment un genoma real. Curiosament, i difícilment atribuïble a la casualitat, les cèl·lules eucariotes tenen els gens inactius per defecte i una majoria de factors activadors; les cèl·lules procariotes viceversa. Aquestes dades reals ens indiquen que, no gens sorprenentment, ambdós tipus de cèl·lules tenen xarxes dissenyades per residir en règim ordenat.

En cas que els models de xarxes booleanes, amb les dues modificacions esmentades, representin fidelment els genomes

reals, quin significat té, quins avantatges confereix a la cèl·lula el disposar d'una dinàmica genòmica que resideixi en règim ordenat? Per contestar aquesta pregunta cal considerar les conseqüències biològiques de residir en un règim o en l'altre i especialment en la transició entre ambdós.

Entre el règim ordenat i el caos existeix un lloc, el límit del caos pel costat ordenat, on la capacitat de les xarxes per fer discriminacions sobre un fons sorollós es maximitza. L'explicació proposada per Kauffman és la següent: si la cèl·lula residís en un règim ordenat profund ignoraria gairebé totes les alteracions en l'estat d'activació de gens concrets, per exemple les produïdes per senyals hormonals externs, i resultaria ben poc flexible i prou inútil per a la comunicació intercel·lular. Per altra banda, si residís en un règim caòtic, qualsevol senyal extern podria desbaratar el precís funcionament de la cèl·lula. En canvi, en una cèl·lula que visqui en aquest límit, dos estats d'una mateixa xarxa que difereixin molt poc entre si tendiran a convergir, de manera que seran classificats (integrats per la cèl·lula) com el mateix estat; per exemple el mateix tipus de senyal hormonal extern exposat sobre un fons variable (sorollós). Malgrat el soroll extern, la cèl·lula aconseguiria distingir els senyals. És una sorpresa a mitges que totes les dades biològiques recollides indiquin que les cèl·lules viuen en aquest límit del caos pel costat ordenat. Simplement, si no hi visquessin hi haurien de viure; tant els organismes unicel·lulars com els pluricel·lulars tenen les cèl·lules exposades a una gran diversitat química. Segons Lovelock, la vida és un fenomen social a tots els nivells, i per tant la capacitat de distingir senyals vers soroll —a comunicació efectiva— resulta fonamental per a la supervivència. Kauffman va més enllà i formula la hipòtesi que l'evolució pot tendir a portar els genomes cap al límit del caos, per tal d'optimitzar la seva funcionalitat.

Són ja antics els primers estudis que buscaren una relació entre el nombre de gens i el nombre de tipus cel·lulars d'un organisme. Tot i haver-se demostrat una relació directa aproximada, els mecanismes generadors de més tipus cel·lulars a partir de més gens sempre s'han deixat sense enunciar adduint un vague augment de "complexitat". Els models de xarxes booleanes ens aporten la primera possible explicació d'aquest fenomen: cada xarxa es caracteritza per un nombre d'atractors. S'ha observat que al límit del caos pel costat ordenat el nombre d'atractors diferents és proporcional a l'arrel quadrada del nombre de gens. Kauffman, emprant l'estimació preseqüenciada de 80.000 gens per al genoma humà, calculà que per una xarxa tal el nombre de tipus cel·lulars rondaria els 283; el nombre de tipus cel·lulars estimats és de 256. Actualment el comptatge de gens humans se situa entorn dels 35.000; tanmateix, si considerem no els gens sinó els productes gènics, generats per exemple per empalmament (*splicing*) alternatiu i modificacions postraducionals, la xifra podria molt bé rondar els 80.000 i rescatar ad hoc els càlculs de Kauffman.

Molt interessant també resulta la distribució en forma de llei potencial de la magnitud de les allaus de canvis en dinàmiques que

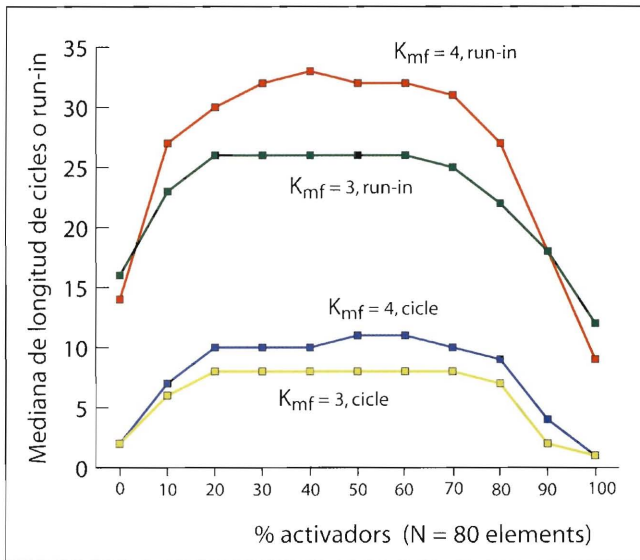


Fig.4. Mediana de longituds de "run-in" i de cicles (en instants de temps) respecte al percentatge de nodes activadors ($N = 80$ nodes), considerant tan sols funcions booleans amb significat biològic. Es mostren dades per $K = 3$ i $K = 4$ entrades. El "run-in" és el nombre d'instants de temps que triga en caure en un atractor (cicle) una xarxa inicialitzada amb l'activitat dels nodes assignada aleatòriament. Un "run-in" curt indica que la xarxa resta lluny del límit del caos pel costat ordenat. Observi's l'homogeneïtat en la longitud dels cicles entre el 20 i el 80% de factors inhibidors. Les xarxes amb una totalitat de factors activadors o inhibidors presenten cicles molt curts. S'utilitza la mediana de la longitud perquè la presència d'atractors i "run-in" molt llargs fa que la mitjana no sigui representativa. (Raeymaekers 2002).

resideixen en règim ordenat. Una allau de canvis és l'alteració en cadena que es produeix en alterar un gen o grup de gens determinat de la xarxa. Segons aquesta distribució, la majoria de les alteracions rebudes per la cèl·lula alterarien poc o gens la transcripció gènica; biològicament es correspondrien a ordres dirigides a les cèl·lules perquè executin una funció pròpia, com per exemple secretar. Un petit nombre d'alteracions, en contrast, podrien desencadenar canvis majors que, per exemple, poguessin desembocar en un canvi d'atractor. De ser certa la relació entre atractors i tipus cel·lulars, estaríem parlant de diferenciació cel·lular. Els diferents passos necessaris per assolir un tipus cel·lular diferenciat concret podrien correspondre a successives allaus de magnitud elevada, de manera que la cèl·lula anés canviant d'atractor fins a estancar-se en el definitiu. Per altra banda, la distribució en llei potencial té un límit de tall superior quant al nombre de nodes de la xarxa que poden ser alterats en l'allau. Aquest límit se situa entre el doble i el

triple de l'arrel quadrada del nombre de nodes de la xarxa. Buscant la comparació en la mosca *Drosophila*, la major allau coneguda és la provocada per l'hormona ecdisona a la muda de la larva. Aquesta afecta 155 "puffs" de cromosomes gegants (segons Kauffman) i 348 gens (segons White et al, 1999), que se situen per sota i entorn de 254 i 381 ($2 \times 3 \times 127$, l'arrel quadrada de 16.000 gens).

Els models de xarxes booleans ens suggereixen doncs que no només els gens en si són importants, sinó que la seva connectivitat resulta fonamental per la dinàmica de la cèl·lula. Ens permeten comprendre perquè, en experiments de microarrays, la gran majoria de gens no mostra canvis apreciables d'expressió entre diferents tipus cel·lulars o diferents estadis de desenvolupament. Encara més, ens permet visualitzar com es poden diferenciar les cèl·lules de manera irreversible, fluint per diferents atractors fins assolir aquells d'on és més difícil sortir -sense una alteració tan gran que condueixi la cèl·lula a un règim caòtic i per tant la destrueixi. L'observació xocant que la major part dels nodes de la xarxa s'estabilitzen de seguida, sense importar l'estat inicial, demostra que fins i tot en les xarxes aleatòries apareixen grans grups de nodes que forcen mútuament el seu estat d'activació, tot formant estructures altament ordenades. El model, agosaradament simple, de xarxes booleans amb estats binaris d'activació (actiu/inactiu) ha estat ratificat per models més complexos que mostraven dinàmiques idèntiques, tot i ser construïts contemplat nivells d'activació més escalonats dels nodes.

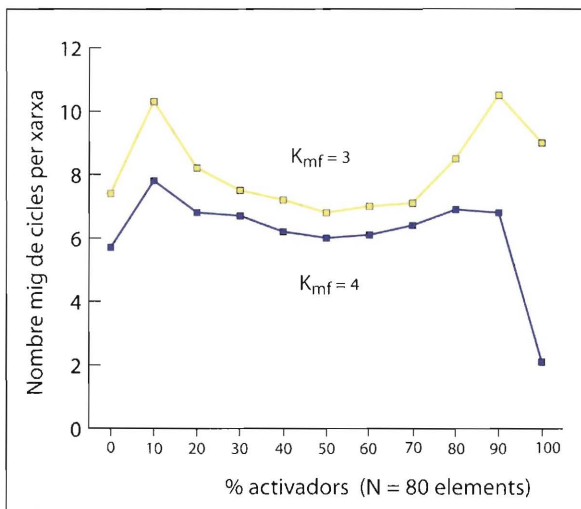


Fig.5. Nombre mig de cicles per xarxa respecte al percentatge de nodes activadors, considerant nodes regits per funcions booleans amb significat biològic. Observi's l'augment acusat del nombre de cicles en xarxes properes a una totalitat de factors o bé activadors o bé inhibidors, i la immediata caiguda en els casos límit. (Raeymaekers 2002).

Bibliografia:

- Kauffman, S. (2003) Investigaciones. En particular el capítol 8 titulat "Leyes para la co-construcción de una biosfera". Col·lecció Metatemas, Tusquets ed.
- Raeymaekers, L. (2002) Dynamics of Boolean Networks Controlled by Biologically Meaningful Functions. *J. Theor. Biol.* 218: 331-341.