

La revolució pendent: com les noves tecnologies bioinformàtiques permetran la integració i comprensió de la biologia i la medicina, i donaran pas a la medicina predictiva, personalitzada, preventiva i participativa

Jordi Naval
Anaxomics

Els termes *bioinformàtica* i *biologia computacional* es refereixen a la investigació de la biologia, tant en els seus vessants teòric i de recerca com en les seves aplicacions pràctiques a la medicina, biotecnologia, etc., mitjançant l'aplicació conjunta de tècniques i programaris d'anàlisi estadística i matemàtica sobre les grans quantitats de dades biològiques obtingudes i emmagatzemades en sistemes informàtics. Cada cop més, la bioinformàtica, juntament amb la nova ciència de la biologia de sistemes, ha esborrat fronteres entre les diferents disciplines de la biologia aplicada a la salut (genètica, biologia molecular, biologia cel·lular, biotecnologia, fisiologia, medicina), fins a convertir-se per si mateixa en una ciència integradora i multidisciplinària, que pretén, en darrer terme, modelar i analitzar sistemes biològics complexos.

Com es comentarà més endavant, la medicina predictiva, personalitzada, preventiva i participativa (medicina P4) és un pas més en aquesta direcció integradora, que canviarà la manera en què entenem la pràctica de la medicina. En aquest nou paradigma,

la bioinformàtica i la biologia de sistemes hi tindran un paper fonamental.

1

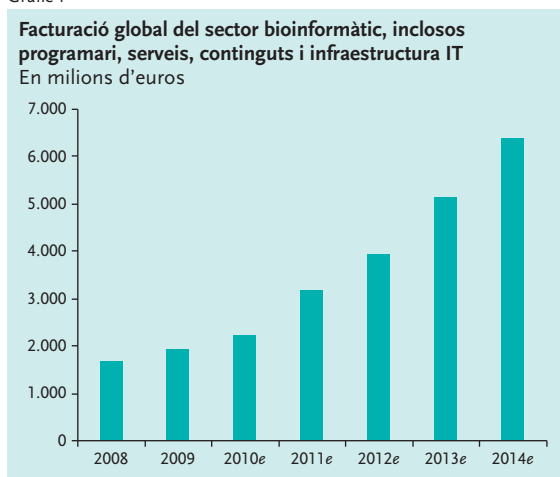
Les eines bioinformàtiques clàssiques i la seva evolució

El mercat de la bioinformàtica d'àmbit mundial ha evolucionat amb un creixement anual acumulat del 15% els darrers anys, i s'espera que la seva taxa de creixement augmenti fins a un 24%, per arribar a un mercat global de 6,4 mil milions d'euros el 2014 (gràfic 1).

Actualment, les tecnologies fonamentals relacionades amb la bioinformàtica es poden agrupar en les categories següents:

- *Genòmica, proteòmica, metabolòmica i farmacogenòmica*: són les ciències i tecnologies que estudien la relació entre un gen (genòmica), la seva expressió com a proteïna (proteòmica), el diferent conjunt de metabòlits que es poden aïllar

Gràfic 1



Font: elaboració pròpia.

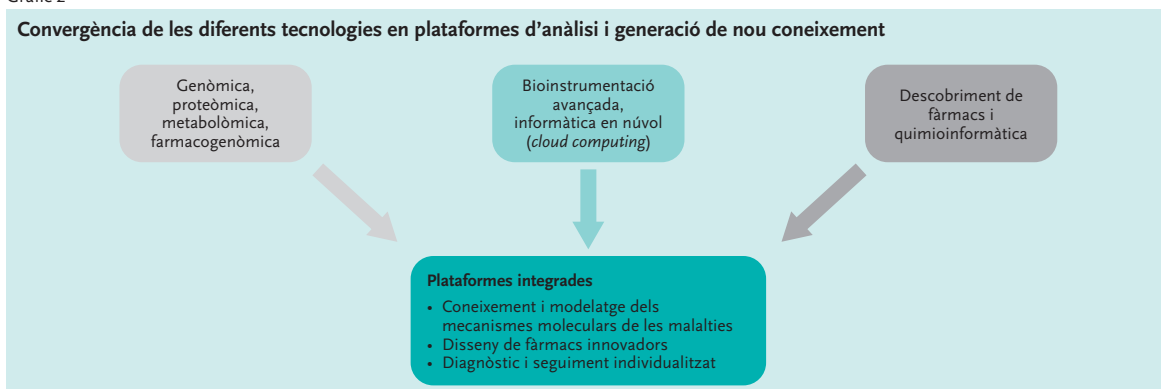
de l'organisme (metabolòmica) i com el cos respon a diferents medicaments en funció del perfil genètic (farmacogenòmica). Aquestes tecnologies estan força desenvolupades actualment i permeten l'anàlisi no només de la seqüència de DNA individual d'una persona (amb un cost que aviat baixarà dels 1.000 euros per genoma), sinó també l'estudi en el temps del patró d'expressió proteica o de metabòlits amb un nivell de detall suficient per poder diferenciar molècula a molècula com i quan varia la nostra bioquímica en funció d'una malaltia o de l'aplicació d'una teràpia. Aquestes eines s'utilitzen, principalment, per al diagnòstic i la identificació de biomarcadors d'evolució de les malalties. Per la seva banda, la farmacogenòmica permet establir la relació entre els fàrmacs que s'administren a un pacient i el seu perfil genètic, així com amb els canvis observats en el seu perfil proteòmic i metabolòmic. Idealment, permetrà la personalització de les teràpies farmacològiques en funció del tipus de pacient i de la resposta que es vagi obtenint amb el temps.

- **Descobriment de fàrmacs i quimioinformàtica:** s'estima que l'aplicació de mètodes bioinformàtics en el descobriment de fàrmacs pugui

reduir els costos de desenvolupament de nous medicaments en un 33% i el temps de desenvolupament, en un 30%. Aquests mètodes estan basats en la predicció de l'efecte dels compostos sobre les xarxes biològiques humanes i en com aquestes modulacions de la xarxa estan relacionades amb els efectes observables desitjats (curació de les malalties) i indesitjats (problemes d'efectes adversos). En una primera etapa, les eines bioinformàtiques permetran el modelat de compostos que inhibeixin específicament les proteïnes o mecanismes que desitgem, sense modificar o alterar el comportament d'altres parts de la xarxa. O, des d'un altre enfocament, permetran conèixer exactament com els compostos actuals (o aquells que es trobin en productes naturals o es desenvolupin per química sintètica) interaccionen amb les seves dianes terapèutiques. Addicionalment, la tendència futura de les agències reguladores (EMA i FDA) és demanar una justificació del mecanisme d'acció dels fàrmacs basada en la predicció d'eficàcia i seguretat mitjançant mètodes de biologia de sistemes. Les agències volen optimitzar els esforços en recerca i discriminar els projectes de desenvolupament de nous medicaments cada cop en etapes més inicials, amb l'objectiu d'optimitzar esforços econòmics i no posar en perill la salut dels pacients amb compostos que tenen un potencial d'efectes adversos elevat. Aquestes línies estratègiques es recullen en els documents "Critical path" de l'FDA (www.fda.gov) i "Innovative medicines initiative" de l'EMA (www.imi.europa.eu).

- **Bioinstrumentació:** és evident que, sense els sofisticats equips actuals d'obtenció de dades sobre les propietats dels sistemes biològics, no hi podria haver bioinformàtica com a tal. Els sistemes de seqüenciació (tant de gens com de proteïnes) evolucionaran espectacularment i produiran cada cop més informació en menys temps i a un cost menor gràcies als seqüenciadors de nova generació basats en *microarrays*. Igualment ho estan fent les plataformes d'espectrometria de

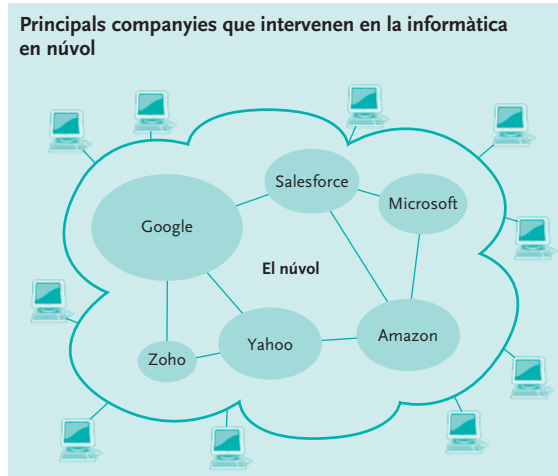
Gràfic 2



masses i sistemes combinats, la cristal·lografia i anàlisi d'estructures moleculars (que farà un salt espectacular gràcies al sincrotró Alba) i l'anàlisi d'imatge *in vivo* d'alt rendiment, entre moltes d'altres. Caldrà desenvolupar nous sistemes de microfluidimetria per mesurar de manera massiva i barata milers de paràmetres en una petita quantitat de sang o altres fluids, així com sistemes de captura i identificació de proteïnes individuals no basats en anticossos. Tots aquests sistemes, juntament amb la informatització dels historials mèdics individuals, està generant una enorme quantitat d'informació individual, de tots els aspectes de la salut, del genotip als signes clínics.

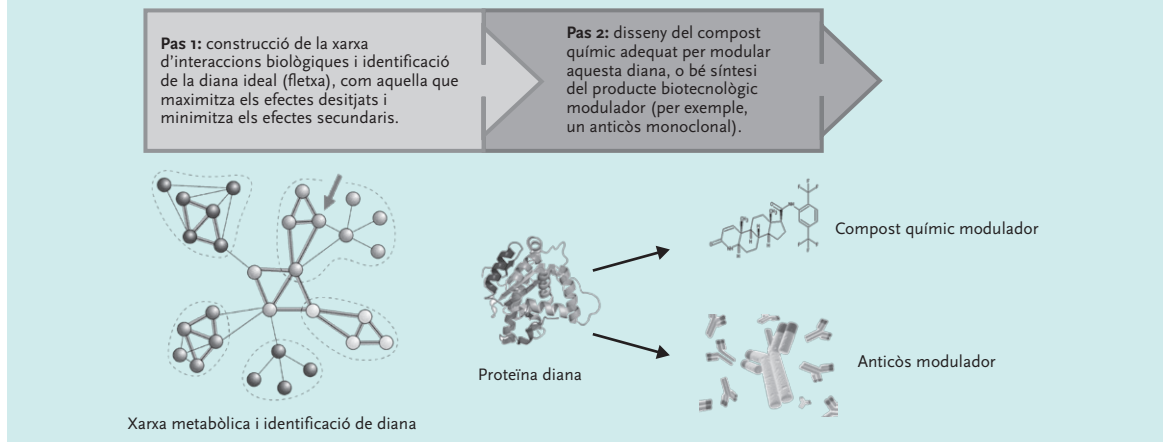
- **Infraestructura tecnològica:** sembla probable que els serveis d'infraestructura tecnològica (principalment, xarxes informàtiques i maquinari i equips) evolucionin cap a l'anomenada *informàtica en núvol* (*cloud computing*) per la típica forma de núvol amb què se sol representar la xarxa Internet, de manera que la potència de càlcul necessària per a qualsevol de les tasques descrites (anàlisi de seqüències genètiques, de dades de proteòmica/metabolòmica, càlcul dinàmic d'estructures de proteïnes, modelat de xarxes biològiques complexes i modelat d'interaccions entre molècules químiques i dianes proteiques, entre d'altres) pugui deixar de ser una limitació

Gràfic 3



i convertir-se en un simple pagament per servei o pagament per temps i potència de càlcul. Les significatives innovacions en virtualització i en graella de càlcul (*grid computing*), així com la millora en l'accés a Internet de gran velocitat, en el marc d'una economia feble, han accelerat el desenvolupament de la informàtica en núvol, que ofereix a demanda la potència de càlcul dels grans servidors de les corporacions més actives a la xarxa, com Amazon i Google. En conseqüència, preveiem que ja no es requerirà la construcció o adquisició de grans infraestructures o computadors, sinó que simplement s'adquiriran "minuts"

Gràfic 4

Procés de descobriment de nous fàrmacs mitjançant la combinació de biologia de sistemes i quimioinformàtica

o “hores” de càlcul de manera flexible als grans proveïdors, de manera que l'usuari només necessitarà un PC i connexió a Internet per efectuar els mateixos càlculs que avui requereixen costoses infraestructures.

2**La sinergia de la biologia de sistemes i la quimioinformàtica en *drug discovery***

La biologia de sistemes és una nova ciència en desenvolupament que té com a objectiu integrar tota la informació “òmica” existent (genòmica, proteòmica i metabolòmica) i posar-la en relació amb fenòmens observables (per exemple, estats de salut, malaltia o efectes dels medicaments), mitjançant la construcció de xarxes d'interconnexió entre elements biològics, seguit de la modelització d'aquestes xarxes per explicar els efectes que podem observar. En termes habituals, es tractaria d'explicar i predir els efectes dels medicaments mitjançant el coneixement i la modelització de totes les rutes bioquímiques implicades en el seu mecanisme d'acció.

El procés teòric per al descobriment de nous fàrmacs basats en la combinació de biologia de sistemes

i quimioinformàtica té dues etapes diferenciades. En primer lloc, com s'ha esmentat, cal construir la xarxa metabòlica complexa i identificar quins punts del mapa són clau per a la consecució d'efectes desitjats, tot minimitzant els efectes indesitjats. Aquests punts sensibles serien les dianes dels medicaments que volem dissenyar. En una segona etapa, un cop identificades les dianes (normalment, proteïnes amb una determinada funció metabòlica o de senyalització clau), mitjançant tècniques quimioinformàtiques es dissenyaria el compost químic que és més específic per a aquesta diana i que presenta les característiques fisicoquímiques adequades per ser utilitzat com a medicament.

Aquesta nova perspectiva en xarxa per al tractament de malalties que cada cop més es demostra que són multifactorials i fortament interrelacionades (obesitat, diabetis, càncer, malalties respiratòries, cardiovasculars, neurològiques, etc.) presenta nombroses implicacions en el procés de descobriment de fàrmacs, ja que el punt que volem modular ja no serà una proteïna individual, sinó que la nostra diana pot ser ara una ruta metabòlica completa (Pujol *et al.*, 2010). En aquest sentit, és probable que vegem cada cop més medicaments formats per combinacions de molècules actives contra un conjunt de

múltiples dianes terapèutiques seleccionades prèviament.

Algunes de les tecnologies i models matemàtics d'aplicació a la biologia de sistemes tot just estan en la seva infantesa. Si aquesta aproximació té èxit en els propers anys, representarà el següent pas cap al descobriment de nous fàrmacs basats en el coneixement i modelat del seu mecanisme d'acció.

3

La biologia de sistemes com a nova plataforma clau en la futura medicina P4

El terme *medicina predictiva, personalitzada, preventiva i participativa* (medicina P4) va ser encunyat pel Dr. Leroy Hood en un article fonamental el 2009 (Hood i Galas, 2009). L'autor, una autoritat en biologia de sistemes, sosté que estem a prop de veure un canvi de paradigma, en el qual la medicina passarà de ser principalment “reactiva”, a ser predictiva, personalitzada, preventiva i participativa. Possiblement, ja no haurem d'esperar que el pacient estigui “malalt” per actuar, sinó que podrem actuar molt abans, de manera personalitzada i adaptable, i amb més efectivitat. Els aspectes bioinformàtics que es descriuen a continuació seran clau en aquest desenvolupament:

- Ja hem comentat com les tècniques diagnòstiques i òmiques permeten milions de mesures sobre cada pacient individual, començant pel seu genoma (que podrà ser seqüenciat completament per menys de 1.000 euros d'aquí a tres anys, i potser per sota de 100 euros d'aquí a cinc anys), fins arribar al nivell d'expressió de proteïnes o metabòlits en diferents punts de temps (per exemple, en estat sa o en estat malalt), i juntament amb altres dades macroscòpiques que van des de l'evolució de la pressió arterial fins a imatges detallades del cervell en funcionament. El desafiament per a la bioinformàtica en el sentit més ampli serà com compilar, normalitzar i analitzar aquesta enorme quanti-

tat de dades per generar hipòtesis d'aplicació mèdica.

- Les noves eines bioinformàtiques acoblades a les potents noves instrumentacions permetran la predicció de l'aparició de les malalties i el seu diagnòstic precoç en funció de cada malalt i de cada moment de la seva vida. Un pas intermedi abans de la individualització total serà l'estratificació dels pacients en categories on la teràpia mèdica aplicada sigui més efectiva. Les noves tecnologies permetran un seguiment de l'efectivitat de les teràpies al llarg del temps, mitjançant el monitoratge de biomarcadors rellevants.
- Addicionalment, el coneixement cada cop més exhaustiu de la relació entre l'estructura química dels medicaments, les seves dianes terapèutiques en el cos humà, els mecanismes d'acció sobre els quals actuen i els efectes que podem observar, tant terapèutics com d'efectes adversos, obren un nou camí per al disseny racional de nous medicaments més efectius i segurs.

Posant en relació els paràgrafs anteriors, es podrà testar ben aviat, mitjançant tècniques bioinformàtiques, l'efecte predit d'una nova estructura química sobre l'expressió de proteïnes o metabòlits en el cos humà al llarg del temps, havent definit prèviament quins efectes volem observar i quins volem evitar. Així, podrem avaluar ràpidament i de manera molt econòmica processos de recerca que actualment duren anys, amb un cost econòmic, social i humà molt important.

La medicina P4 té encara un cert camí per endavant per demostrar la seva efectivitat en termes mesurables en la població. El que és clar és que queden per resoldre temes molt importants relacionats, sobretot, amb aspectes econòmics i d'organització dels sistemes de salut. Per exemple, una qüestió que caldrà solucionar ben properament és qui, com i quan pagarà i gestionarà la informació mèdica que es pugui obtenir del genoma individual d'una persona.

Creiem que Catalunya en particular disposa de la combinació de grups d'investigació i d'empreses tecnològiques adequada per convertir-se en clúster regio-

nal capdavanter en l'àmbit mundial en aquest nou paradigma de la medicina P4 (vegeu el quadre 1). Per exemple, Luxemburg ha iniciat un projecte de partenariat amb l'Institute for Systems Biology (Seattle, EUA) que implica la inversió pública i privada de més de 100 milions de dòlars en els propers cinc anys, amb l'objectiu d'establir al país un model integrat que posi en marxa la medicina P4. Seria necessària, però, una decidida política de suport a la recerca que impliqui les companyies biofarmacèutiques, els centres públics de recerca i les pimes innovadores.

4

La bioinformàtica i biologia de sistemes aplicada a les energies renovables i al medi ambient

Resulta obvi per al lector que l'aplicació de les noves tecnologies bioinformàtiques i de biologia de sistemes a la medicina són extensibles a altres àmbits de la biotecnologia, com ara la veterinària, la producció industrial biotecnològica o l'agricultura. Per exemple, una de les primeres aplicacions de la biologia de sistemes ha estat la producció en llevats de l'artemisinina, un compost efectiu contra certs tipus de malària, mitjançant el disseny computacional *de novo* des d'una perspectiva de metabolisme en xarxa, d'una ruta metabòlica de síntesi del producte. Doncs bé, una altra aplicació de la biologia de sistemes és en la modelització i modificació d'ecosistemes microbians a escala industrial per a: 1) optimització dels processos biològics de degradació de la lignina de la biomassa, per tal de produir cultius amb millor rendiment per a la producció de biocombustibles (en aquest cas, parlariem de la biologia d'ecosistemes, ja que aquesta degradació és duta a terme de manera natural per un conjunt de fongs, bacteris i altres microorganismes, que caldria modelar i redissenar conjuntament); 2) comprensió, modelatge i disseny d'ecosistemes d'algues o altres microorganismes per a la producció directa de biodièsel o d'hidrogen, i 3) bioremediació i segrest de CO₂. Actualment, alguns programes de

noves energies, principalment als Estats Units, ja integren aquesta visió en els seus plans estratègics (genomicscience.energy.gov/).

5

La necessària col·laboració entre les companyies farmacèutiques, les pimes innovadores i els centres de recerca, i el paper dinamitzador de l'Administració

La indústria farmacèutica està canviant radicalment la seva aproximació a la recerca inicial en descobriment de nous fàrmacs. Estem veient el pas d'una estratègia d'internalització i protecció màxima del coneixement, en forma de bases de dades i sistemes d'anàlisi propietat de les grans companyies, a un entorn on la indústria cerca activament col·laboracions amb altres companyies competidores, pimes amb visions innovadores i entitats de recerca públiques i privades. La tendència clara és que l'augment exponencial del volum i la qualitat de la informació biològica accessible de domini públic fa que sigui molt més efectiva una estratègia de col·laboració on es comparteix informació de caràcter bàsic i hom s'aprofita de l'existent. Aquesta visió està encoratjada per les agències reguladores governamentals com la FDA i l'EMA, mitjançant les iniciatives "Critical path" i "Innovative medicines initiative", respectivament. Bases de dades de literatura com PubMed, de compostos químics com PubChem o de dades òmiques com KEGG, OMIM, DrugBank, SwissProt i desenes d'altres, són ja repositoris públics de referència de la màxima qualitat. Addicionalment, nombroses aplicacions i programari d'anàlisi de codi obert fan traslladar el valor afegit de la possessió de les dades cap a visions i implementacions radicalment innovadores i integradores de tot aquest coneixement, orientades a oferir solucions d'aplicació ràpida, per exemple, en descobriment de nous fàrmacs. En aquest sentit, cada cop més la indústria biofarmacèutica forma consorcis i busca la col·laboració interna i amb altres entitats, amb l'objectiu de

Quadre 1

Empreses de bioinformàtica a Catalunya

Anaxomics (www.anaxomics.com)

Anaxomics és una empresa de serveis orientada a oferir serveis de biologia de sistemes, principalment a la indústria farmacèutica i a altres grups de recerca. Aquesta companyia és activa, sobretot, en les àrees de predicció d'efectes adversos clínics en les fases inicials del descobriment de nous medicaments, així com en la identificació de noves indicacions per a fàrmacs coneguts. Anaxomics ha desenvolupat tecnologies pròpies, així com sistemes per a la integració de tecnologies de química computacional d'altres companyies.

Intelligent Pharma (www.intelligentpharma.com/)

Intelligent Pharma desenvolupa tecnologies basades en química computacional per oferir serveis avançats de descobriment de nous fàrmacs. La companyia integra completament els projectes de recerca per optimitzar els processos de recerca de les companyies farmacèutiques. La tecnologia pròpia d'Intelligent Pharma (Helios) permet la determinació de nous compostos actius o la determinació del mecanisme d'acció de la molècula d'interès, mitjançant una solució de *virtual screening* basada en lligands. El perfil farmacològic d'una molècula depèn dels seus camps moleculars superficials. Helios compara els camps moleculars d'un compost nou en desenvolupament amb els de molècules amb mecanismes d'acció coneguts. La identificació d'anàlegs per analogia no-estructural permet determinar els mecanismes d'acció dels compostos en estudi.

Chemotargets (www.chemotargets.com)

Chemotargets ofereix serveis de farmacologia *in silico* que inclouen:

- Perfil d'efectes adversos d'una molècula nova, basada en la projecció del seu perfil farmacològic predit sobre una xarxa d'efectes adversos descrits per a fàrmacs coneguts.
- Perfil de *targets*, basat en tècniques de quimiinformàtica, que prediu la capacitat d'unió d'un compost per a un conjunt de 4.700 *targets* proteics ben caracteritzats.
- Identificació de *hits*, a partir d'un *target* o un perfil de *targets*, i mitjançant un *screening* químic *in silico* contra un catàleg comercial que inclou actualment més de 8 milions de compostos.
- Planificació química, que permet expandir virtualment l'espai químic al voltant d'una estructura principal, per metodologies d'*isosteric fragment transformation* (SHIFT).

Aromics (www.aromics.com)

Aromics ofereix una plataforma de recerca translacional mitjançant l'ús de ciències òmiques en l'àmbit de la recerca de nous medicaments i biomarcadors. En concret, Aromics està especialitzada en proteòmica, genòmica i farmacogenòmica, en combinació amb tecnologies *in vitro* i *in vivo*.

Bioingenium (www.bioingenium.net)

Aquesta companyia aplica el seu saber fer (*know-how*) al desenvolupament, optimització i escalat de processos de producció de proteïnes basats en expressió en bacteris o en llevats. Aquests productes són d'aplicació biomèdica per a la salut humana i animal. Bioingenium aplica eines de programari pròpies per a l'optimització d'aquests bioprocessos.

Gendiag (www.gendiag.com)

Gendiag és una companyia biotecnològica que desenvolupa projectes de medicina personalitzada mitjançant eines diagnòstiques i descobriment de noves dianes terapèutiques. Gendiag desenvolupa plataformes diagnòstiques basades en farmacogenètica i farmacogenòmica en oncologia, cardiovascular, sistema digestiu i metabolisme de fàrmacs, així com en l'anàlisi d'expressió genòmica en aquestes malalties.

Prous Institute for Biomedical Research (www.prousresearch.com)

Prous Institute és una companyia centrada en el descobriment de nous medicaments i en les tecnologies *in silico*. Pel que fa a nous medicaments, aquesta companyia utilitza l'aproximació anomenada *epistemic drug discovery* (EDD), que inclou tecnologies quimiinformàtiques, bioinformàtiques i de recerca bàsica. Addicionalment, Prous Institute ha desenvolupat potents eines de mineria de dades que permeten predir seguretat, eficàcia, farmacocinètica, mecanisme d'acció i altres aspectes d'un fàrmac a partir de la seva estructura química. Aquesta tecnologia permet, addicionalment, construir i entrenar models predictius a partir de dades conegudes.

Microart (www.microart.cat)

Microart és una companyia d'investigació i enginyeria de programari especialitzada en sistemes d'informació innovadors que utilitza la web semàntica i tecnologia multiagent, principalment aplicada a l'extracció de coneixement en bases de dades distribuïdes. Entre d'altres, Microart participa en diversos projectes de recerca internacionals per crear sistemes de suport a les decisions en diferents tipus de tumor utilitzant dades clíniques i dades de metabolòmica i genòmica.

millorar la seva producció de recerca, com ara l'European Bioinformatics Network (www.ebi.ac.uk) o el Predictive Safety Consortium (www.c-path.org/pstc.cfm). Les pimes bioinformàtiques catalanes que apareixen a la llista del quadre 1 serien un molt bon exemple de companyies que volen desenvolupar aquest rol innovador i integrador de la informació pública per posar-la al servei de la innovació de la indústria biofarmacèutica. Creiem que l'Administració ha de tenir un paper clau dinamitzant i facilitant la formació de consorcis públics-privats, que integrin les companyies farmacèutiques, els centres de recerca avançats i les pimes innovadores, fomentant, al mateix temps, la cooperació internacional.

6

I què hi pot haver després del futur?

En aquest article ens hem estat referint a la medicina "individual" i "personal", ja que el primer repte és la comprensió i modelatge complet de la biologia de cadascú de nosaltres. Una de les tendències que es comencen a entreveure és que les malalties com a entitats diferenciades també estan relacionades entre elles en forma de xarxa, de manera que existeixen vincles metabòlics entre, posem per cas, l'obesitat i l'asma, principalment perquè poden compartir processos comuns relacionats amb la inflamació. A aquesta nova ciència òmica se la comença a anomenar *diseasòmica* (A.-L. Barabasi, 2007). Però si anem un pas més enllà, podem veure que el nostre "sistema metabòlic" no està aïllat dins la nostra pell, sinó que interacciona fortament amb els altres sistemes metabòlics de les persones que ens envolten, mitjançant les nostres relacions socials (familiars, amics, feina, etc.). Per exemple, s'ha estudiat que la contribució a

l'obesitat d'una xarxa social d'amics amb el mateix fenotip és aproximadament la mateixa que la contribució de la genètica individual. Per què no podem aplicar en un futur les mateixes metodologies bioinformàtiques que estem aplicant a l'estudi i modelatge d'un sistema aïllat, a l'estudi d'un sistema de sistemes, és a dir, d'un grup humà? De fet, autors catalans (R. Guimerà i M. Sales-Pardo, 2009) han demostrat que les tecnologies matemàtiques d'anàlisi de xarxes que hom utilitza en biologia de sistemes són les mateixes que s'utilitzen per a l'anàlisi de les xarxes socials. Tecnologies de Semantic Web for Health Care i xarxes socials com PatientsLikeMe (www.patientslikeme.com) hi poden tenir un paper clau.

7

Referències bibliogràfiques

- BARABASI, A.-L. (2007). "Network Medicine –from obesity to the 'diseasome'". *N Engl J Med*, 357(4), pàg. 404-407.
- GUIMERÀ, R.; SALES-PARDO, M. (2009). "Missing and spurious interactions and the reconstruction of complex networks". *Proc Natl Acad Sci USA*, 106(52), pàg. 22073-22078.
- HOOD, L.; GALAS, D. (2009). "Systems biology and emerging technologies will catalyze the transition from reactive medicine to predictive, personalized, preventive and participatory (P4) medicine". *IBC*, 1(2), pàg. 6, 1-4.
- PUJOL, A.; MOSCA, R.; FARRÉS, J.; ALOY, P. (2010). "Unveiling the role of network and systems biology in drug discovery". *Trends Pharmacol Sci.*, 31(3), pàg. 115-123.