

Estudi de l'evolució i el tractament del càncer  
des d'una perspectiva de la biologia de sistemes

Jordi Serra  
Tesis doctoral  
2015





## Acta de qualificació de tesi doctoral

Curs acadèmic:

Nom i cognoms

Jordi Serra i Musach

Programa de doctorat

DOCTORAT EN MATEMÀTICA APLICADA

Unitat estructural responsable del programa

## Resolució del Tribunal

Reunit el Tribunal designat a l'efecte, el doctorand / la doctoranda exposa el tema de la seva tesi doctoral titulada

Acabada la lectura i després de donar resposta a les qüestions formulades pels membres titulars del tribunal, aquest atorga la qualificació:

NO APTE

APROVAT

NOTABLE

EXCEL·LENT

(Nom, cognoms i signatura)		(Nom, cognoms i signatura)	
President/a		Secretari/ària	
(Nom, cognoms i signatura)	(Nom, cognoms i signatura)	(Nom, cognoms i signatura)	(Nom, cognoms i signatura)
Vocal	Vocal	Vocal	Vocal

\_\_\_\_\_, \_\_\_\_\_ d'/de \_\_\_\_\_ de \_\_\_\_\_

El resultat de l'escrutini dels vots emesos pels membres titulars del tribunal, efectuat per l'Escola de Doctorat, a instància de la Comissió de Doctorat de la UPC, atorga la MENCIÓ CUM LAUDE:

SÍ

NO

(Nom, cognoms i signatura)	(Nom, cognoms i signatura)
President de la Comissió Permanent de l'Escola de Doctorat	Secretari de la Comissió Permanent de l'Escola de Doctorat

Barcelona, \_\_\_\_\_ d'/de \_\_\_\_\_ de \_\_\_\_\_

A la meva mare  
*sense la teva valentia,*  
*res d'això hagués estat possible.*

I als meus avis Roser i Josep.

M'agradaria començar agraint al meu director de tesi per deixar-me la possibilitat de realitzar el doctorat sota la seva direcció i supervisió. Gràcies Miquel Àngel per l'ajuda i consells rebuts durant aquests anys i per l'oportunitat d'introduir-me a la recerca en la biologia de sistemes i computacional

Agrair a en Francesc de la UPC per acceptar ser el tutor i pels seus consells i idees a l'hora de desenvolupar els projectes.

Gràcies a la Conxi del LRT2 per l'ajuda rebuda al final d'aquesta etapa.

Gràcies a en Manel per deixar-me un raconet i accedir al servidor del PEBC.

També gràcies a la Carme de la UPC per la seva *quasi-infinta* paciència per resoldre dubtes burocràtics.

Gràcies als membres del laboratori. Francesca, per la paciència que has tingut per deixar que un *manasses* com jo participés en els experiments. Gràcies també a en Gorka, Núria, Laia, Helena, Griselda i a la resta de companys que hem coincidit fugaçment pel *lab*.

Gràcies als companys del PEBC, a en Jose, Sergi, Antonio, Holger, Carmen, Olga pels bons moments compartits.

Als companys bioestadístics Antoni i Daniel pels vostres consells de la UBS. Gràcies Antoni pel suport a l'hora de resoldre dubtes i la teva pedagogia per explicar problemes complicats! Gràcies també a la Rebeca pel seu suport i consells.

I per acabar, gràcies a tu Mireia. Aquests últims mesos hagués estat impossible superar-los sense el teu suport i comprensió.



# Índex

<b>I</b>	<b>Introducció</b>	<b>15</b>
1	INTRODUCCIÓ A LA BIOLOGIA DE SISTEMES I A LES XARXES COMPLEXES	17
1.1	Cap a la biologia de sistemes . . . . .	17
1.2	Introducció a la biologia de sistemes . . . . .	18
1.3	Introducció a les xarxes complexes . . . . .	20
1.3.1	Xarxes complexes biològiques . . . . .	20
1.3.2	Xarxes aleatòries . . . . .	21
1.3.3	Xarxes d'escala lliure . . . . .	22
1.3.4	Clústers i mòduls . . . . .	23
1.3.5	Motius . . . . .	24
1.3.6	Estructura jeràrquica . . . . .	24
1.3.7	Robustesa de les xarxes d'interaccions proteiques . . . . .	25
1.3.8	La centralitat subgraf. L'índex d'Estrada . . . . .	26
1.4	L'interactoma proteic humà . . . . .	27
2	INTRODUCCIÓ AL CÀNCER	31
2.1	Mecanismes moleculars involucrats en el càncer . . . . .	31
2.2	Característiques bàsiques del càncer . . . . .	32
2.3	Estadis del càncer de mama . . . . .	34
2.4	Teràpies del càncer . . . . .	35
<b>II</b>	<b>Assumpcions. Objectius</b>	<b>39</b>
3	ASSUMPCIONS	41
3.1	Assumpcions . . . . .	41
4	OBJECTIUS I	43

5	OBJECTIUS II	45
<b>III Mètodes</b>		<b>47</b>
6	MATERIALS I MÈTODES I	49
6.1	La xarxa de l'interactoma . . . . .	49
6.2	Dades d'expressió gènica . . . . .	50
6.2.1	Font de les dades . . . . .	50
6.2.2	Dades d'expressió del càncer de mama . . . . .	50
6.2.3	Dades d'expressió del càncer colorectal . . . . .	51
6.3	Dades d'expressió de les cèl·lules MCF7 . . . . .	52
6.4	Dades d'expressió de les cèl·lules MCF7 transfectades amb SCNA . . . . .	52
6.5	Mètode de la cascada de fallades . . . . .	52
7	MATERIALS I MÈTODES II	55
7.1	La xarxa de l'interactoma . . . . .	55
7.2	Dades d'expressió, de les IC <sub>50</sub> i de mutació de gens . . . . .	55
7.3	Anàlisi dels processos biològics . . . . .	56
7.4	Assaigs cel·lulars . . . . .	56
7.5	Activitat de la xarxa . . . . .	56
7.6	Estimació dels tractaments . . . . .	58
7.6.1	Estimació de la resposta cel·lular a tractaments de fàrmacs . . . . .	58
7.6.2	Estimació de l'efecte de la combinació de fàrmacs . . . . .	58
<b>IV Resultats</b>		<b>59</b>
8	RESULTATS I	61
8.1	Anàlisi dels gens diferencialment expressats . . . . .	61
8.1.1	Gens diferencialment expressats en el càncer de mama . . . . .	61
8.1.2	Gens diferencialment expressats en el càncer de colon . . . . .	62
8.2	La xarxa de l'interactoma . . . . .	63
8.3	Anàlisi topològica de la xarxa de l'interactoma . . . . .	64
8.3.1	Robustesa topològica respecte de les proteïnes del càncer . . . . .	64
	Anàlisi del paràmetre $\alpha$ . . . . .	66
	Anàlisi del paràmetre $r_i$ . . . . .	68
	El model de la cascada de fallades . . . . .	71



Anàlisi respecte de les proteïnes del càncer de mama . . . . .	72
8.3.2 Anàlisi respecte de les vies de senyalització . . . . .	74
8.3.3 Anàlisi respecte de les proteïnes del càncer de colon . . . . .	75
8.4 Robustesa de la progressió del càncer respecte l'estadi inicial . . . . .	76
8.5 Robustesa de la progressió del càncer respecte del tractament . . . . .	78
8.6 Robustesa de la progressió del càncer respecte a la pertorbació gènica	79
8.7 Les proteïnes involucrades en l'autofàgia mostren patrons diferents als del càncer . . . . .	81
8.8 Robustesa topològica respecte la distància dels vèrtexs . . . . .	82
<b>9 RESULTATS II</b>	<b>85</b>
9.1 L'activitat de la xarxa captura informació de la resposta de fàrmacs	85
9.2 Els processos biològics influeixen en les associacions entre l'activitat de la xarxa els fàrmacs . . . . .	88
9.3 Les mutacions conductores del càncer influeixen en les associacions .	90
9.3.1 Diferències respecte de l'estatus de mutació dels gens conduc- tors del càncer . . . . .	90
9.3.2 Concordança entre l'activitat i els perfils d'expressió gènica .	92
9.4 Coherència amb les vies de senyalització i processos biològics . . . . .	92
9.4.1 Concordança entre els fàrmacs respecte de l'estatus de muta- ció dels gens . . . . .	93
9.4.2 Expressió de les dianes dels fàrmacs . . . . .	93
9.5 Predicció de sinergies mitjançant l'activitat . . . . .	94
<b>V Discussió</b>	<b>99</b>
<b>10 DISCUSSIÓ I</b>	<b>101</b>
10.1 Les proteïnes del càncer en l'interactoma . . . . .	101
10.2 Anàlisi topològica de la xarxa de l'interactoma . . . . .	102
10.3 La xarxa de l'interactoma . . . . .	104
<b>11 DISCUSSIÓ II</b>	<b>107</b>
11.1 Activitat de la xarxa . . . . .	107
11.2 Combinació de fàrmacs . . . . .	110

ÍNDEX	10
<b>VI Conclusions</b>	<b>115</b>
12 CONCLUSIONS I	117
13 CONCLUSIONS II	119
<b>VII Bibliografia</b>	<b>121</b>
BIBLIOGRAFIA	123
<b>VIII Annexes</b>	<b>143</b>
A DESCRIPCIÓ DE LA XARXA DE L'INTERACTOMA INTACT	145
B ANÀLISI TOPOLÒGICA DE LA XARXA DE L'INTERACTOMA INTACT	147
C CASCADES DE FALLADES RELATIVES A LES VIES KEGG	149
C.1 Vies KEGG que revelen patrons diferents als del càncer . . . . .	149
C.2 Vies KEGG amb patrons similars als del càncer . . . . .	151
D CLASSIFICACIÓ DE FÀRMACS	153
E DIFERÈNCIES ENTRE CARACTERÍSTIQUES DELS FÀRMACS	155
F DIFERÈNCIES EXPLICADES PER LES MUTACIONS DELS GENS	157
G FÀRMACS ASSOCIATS A L'ESTATUS DE MUTACIÓ DELS GENS	159
<b>IX Currículum</b>	<b>161</b>
CURRÍCULUM	163

# Índex de figures

8.2.1	Distribució del grau de la xarxa de l'interactoma . . . . .	64
8.3.1	Estratègia de l'anàlisi . . . . .	65
8.3.2	Càrrega vs grau . . . . .	66
8.3.3	anàlisi del paràmetre $\alpha$ . . . . .	67
8.3.4	Anàlisi del paràmetre $r_i$ . . . . .	69
8.3.5	Anàlisi de la robustesa de l'interactoma respecte de la progressió del càncer de mama . . . . .	73
8.3.6	Anàlisi respecte de la resposta immune . . . . .	75
8.3.7	Anàlisi de la robustesa respecte de la progressió del càncer colon . .	76
8.4.1	Robustesa topològica dels estadis del càncer respecte de l'estadi inicial	77
8.5.1	Robustesa topològica respecte del tractament . . . . .	79
8.6.1	Robustesa respecte de la pertorbació gènica . . . . .	80
8.7.1	No robustesa respecte de les proteïnes associades a l'autofàgia . . .	82
8.8.1	Robustesa de la xarxa respecte de les distàncies entre vèrtexs . . . .	84
9.1.1	Estratègia de l'anàlisi . . . . .	86
9.1.2	Diferències entre classes de fàrmacs . . . . .	87
9.2.1	Distribució de la centralitat dels vèrtexs . . . . .	89
9.2.2	Distribució dels coeficients de correlació segons la centralitat dels vèrtexs . . . . .	90
9.3.1	Expressió gènica dels fàrmacs identificats respecte de mutacions en els gens conductors del càncer . . . . .	91
9.4.1	Expressió de les concentracions dels fàrmacs associats als gens conductors de tumors . . . . .	93
9.4.2	Expressió de les dianes associades a les mutacions de <i>PTEN</i> i <i>R1</i> .	94
9.5.1	Combinacions de fàrmacs . . . . .	97
A.0.1	Xarxa de l'interactoma IntAct . . . . .	145
B.0.1	Anàlisi topològica en la xarxa IntAct . . . . .	147

---

C.1.1	Fallades respecte de les vies KEGG amb patrons diferents als del càncer	149
C.2.1	Fallades respecte de les vies KEGG amb patrons similars als del càncer	151
F.0.1	Diferències entre els fàrmacs respecte de l'estatus dels gens . . . . .	157

# Índex de taules

6.1.1	Exemple del fitxer HPRD . . . . .	49
8.1.1	Gens diferencialment expressats en el càncer de mama . . . . .	62
8.1.2	Gens diferencialment expressats en el càncer de colon . . . . .	62
8.2.1	Components connexes de l'interactoma . . . . .	63
8.3.1	Valors de $\alpha$ determinats per a cadascun dels estadis . . . . .	68
9.2.1	Processos biològics sobrerrepresentats en els vèrtex més centrals . . .	89
D.0.1	Classificació dels fàrmacs . . . . .	153
E.0.1	Diferències entre les característiques dels fàrmacs . . . . .	155
G.0.1	Fàrmacs associats a l'estatus de mutació dels gens . . . . .	160



## Part I

# Introducció





# 1

## Introducció a la biologia de sistemes i a les xarxes complexes

### 1.1 Cap a la biologia de sistemes

Durant el segle XX la recerca en biologia va experimentar una evolució important i transcendent que va permetre elucidar els mecanismes hereditaris, els mecanismes moleculars i plantejar la complexitat cel·lular com a sistema. A la segona meitat del segle XX, es van desenvolupar les tècniques per seqüenciar cadenes d'ADN (Sanger i Tuppy, 1951) i mesurar l'expressió gènica, que a finals de segle es van sistematitzar, permetent les primeres seqüenciacions de genomes sencers d'organismes simples (elegans Sequencing Consortium, 1998; Blattner et al., 1997; Adams et al., 2000) i la descripció massiva de patrons quantitius d'expressió gènica, és a dir, a escala genòmica (Lockhart et al., 1996; Velculescu et al., 1995). Amb el desenvolupament d'aquestes tècniques es van començar a generar una quantitat enorme de dades i va sorgir la necessitat de definir i desenvolupar eines bioinformàtiques per tractar-les, extreure'n resultats i interpretar-les. Amb les primeres eines bioinformàtiques que permetien analitzar les dades sistemàticament es van poder realitzar estudis integrant diferents tipus de dades i estudiar la cèl·lula des de la perspectiva de sistema complex. Així, a cavall entre els dos segles, va néixer la biologia de sistemes (Kitano, 2002).

La seqüenciació completa de genomes, i en particular la del genoma humà a l'any 2001 (the International Human Genome Sequencing Consortium, 2001), ha estat una fita important a partir de la qual ha estat possible identificar tots els gens que l'integren i construir-ne el seu mapa genòmic (Dunham et al., 2012). Gràcies als bioxips, que permeten interrogar milers de gens, si no tots, d'un genoma, s'ha pogut estudiar l'activitat dels gens de forma massiva i obtenir una fotografia dels

patrons d'expressió gènica en unes condicions donades o desitjades i, d'aquesta manera, detectar aquells gens actius o inhibits i les possibles relacions entre ells (DeRisi et al., 1997; Cho et al., 1998). L'ús dels bioxips com a eina ha contribuït a realitzar estudis més complexos, permetent elucidar els fenotips derivats de mutacions gèniques (Ross-Macdonald et al., 1999), identificar aquells gens que podrien contribuir en el desenvolupament d'una malaltia (Heller et al., 1997) o, inclús, predir dianes terapèutiques per un tractament específic (Marton et al., 1998).

L'ús dels bioxips però, va començar a produir una quantitat molt gran de dades i va sorgir la necessitat de definir i desenvolupar eines bioinformàtiques pel seu tractament i interpretació. Eines com la clusterització han estat àmpliament aplicades a les dades d'expressió gènica per definir signatures de gens (Eisen et al., 1998; Alizadeh et al., 2000) i caracteritzar, per exemple, el càncer de mama (Perou et al., 1999). Amb l'ús d'eines bioinformàtiques i la capacitat de processar grans quantitats de dades, s'ha pogut aprofundir en la complexitat del sistema cel·lular. Per exemple, Lamb et al. van integrar dades de diversos càncers per descriure els patrons de l'expressió de *CCND1*, l'activitat del qual està associada al càncer, i van descriure en detall els mecanismes que regulen la seva expressió. Les eines bioinformàtiques van permetre, també, descriure el fenotip cel·lular mitjançant l'activitat gènica i de seguida es va materialitzar la idea de que els patrons d'expressió transcripcional en la cèl·lula reflectien el fenotip cel·lular. Mitjançant aquesta idea ha estat possible refinar i millorar la classificació clínica de tumors, identificar-ne els tipus cel·lulars involucrats (Perou et al., 2000) i fer pronòstics acurats del càncer (van de Vijver et al., 2002).

## 1.2 Introducció a la biologia de sistemes

La biologia de sistemes és la branca científica que s'aproxima a la biologia natural o als fenòmens de la vida a través dels sistemes dinàmics i complexos. Ja a mitjans del segle XX es van començar a realitzar estudis sobre la complexitat cel·lular. Un exemple rellevant és l'estudi de Delbrück, a finals del 1940, per respondre a la qüestió de com cèl·lules amb el mateix genotip generen cèl·lules amb diferents fenotips. Per explicar-ho, va modelitzar l'evolució del fenotip cel·lular mitjançant un model biestable format per dues partícules moleculars que es retroalimentaven positivament pel qual el fenotip es va determinar segons l'activació d'aquestes partícules en cada instant de temps (Vidal, 2009). Malgrat estudis similars a aquests que interrogaven

la complexitat del sistema cel·lular, no és fins a finals de segle que la biologia de sistemes es va forjar com una disciplina (Kitano, 2002).

Per comprendre completament un sistema biològic cal descriure (i) l'arquitectura, (ii) la dinàmica i (iii) construir mètodes de control i disseny del sistema. Per a descriure l'arquitectura del sistema cal identificar tots els elements que l'integren, els nivells del sistema que el formen, com els elements s'organitzen a través d'aquests nivells i quines funcions tenen o en quines participen. Tanmateix, la identificació dels elements no és un pas trivial, i encara menys conèixer-ne les seves funcions. Tot i que les tecnologies de seqüenciació i dels bioxips han permès obtenir un mapa genètic acurat i l'activitat dels gens en diferents condicions o processos cel·lulars (Dunham et al., 2012), encara no s'han pogut descriure de manera exhaustiva les funcions gèniques, ni obtenir un mapa complert ni del transcriptoma ni del proteoma (Organization, 2010; Legrain et al., 2011).

Quan se'n coneix l'arquitectura, cal entendre el seu funcionament i dinàmica. Cal descriure'n els patrons dinàmics, els mòduls funcionals i com s'organitzen. Estudiar la resposta del sistema modificant la seva estructura i veure quin paper tenen els elements en la funcionalitat del sistema. Per exemple, en l'interactoma proteic, estudiar la resposta a la supressió o alteració dels seus elements, que pot ser a causa de mutacions gèniques, i que puguin està involucrats en el desenvolupament del càncer (Barabasi i Oltvai, 2004). Entendre la robustesa és un punt clau per comprendre el sistema i els seus punts febles estructurals o funcionals. Amb la comprensió de la dinàmica del sistema i la seva robustesa, es podran construir models per estudiar-ne el seu comportament.

Un dels objectius finals de la biologia de sistemes és modelitzar el sistema en qüestió per descriure'n la dinàmica i predir o inferir solucions a diferents problemàtiques derivades d'alteracions estructurals o funcionals. Després de la comprensió de la dinàmica del sistema, el següent pas natural és dissenyar i construir mètodes de control per modelitzar-ne l'activitat mitjançant diferents condicions i trobar solucions a diferents problemàtiques del sistema. Per exemple, el coneixement exhaustiu del sistema cel·lular podria permetre modelitzar la seva mecànica i dinàmica, estudiar per quines alteracions concretes del sistema les cèl·lules esdevenen canceroses i interrogar-ne la seva robustesa. D'aquesta manera, es podria prevenir el desenvolupament d'un càncer o determinar-ne teràpies per eradicar-lo.

Molts dels sistemes complexos reals, i en particular els biològics, es poden modelit-

zar mitjançant xarxes complexes i representar de manera senzilla la interacció dels seus elements mitjançant les arestes. Així, un coneixement profund de les xarxes complexes és un requeriment pel desenvolupament de la biologia de sistemes.

## 1.3 Introducció a les xarxes complexes

### 1.3.1 Xarxes complexes biològiques

Molts sistemes reals solen tenir una gran quantitat d'elements i la interacció dels seus elements sol tenir un efecte multiplicatiu, enlloc de ser una combinació lineal dels seus elements. Això fa que siguin altament complexos i molt difícil d'explicar tan la seva estructura com el seu funcionament. Molts d'aquests sistemes es poden modelitzar mitjançant xarxes que per la seva topologia, estructura i propietats solen ser complexes. Aquestes xarxes reals són difícils de descriure a causa del gran nombre de vèrtexs i arestes que tenen i de com aquestes arestes es distribueixen entre els vèrtexs, a més de que els seus elements s'organitzen a diferents nivells del sistema, dotant a la xarxa d'una topologia complexa (Barabási i Albert, 1999; Barabasi i Oltvai, 2004). Aquestes xarxes reals complexes solen ser riques en propietats que les fan comunicatives —en el sentit que la informació pot passar d'un vèrtex a un altre ràpidament—, robustes —solen tenir una resiliència alta envers a canvis ambientals o estructurals interns—, i els seus elements solen organitzar-se topològicament segons la seva naturalesa i funcionalitat —s'organitzen en diferents nivells i mòduls funcionals. Així, un coneixement profund en la teoria de xarxes complexes és crucial per entendre molts dels sistemes reals que es poden trobar a la natura, en particular el sistema cel·lular i cadascun dels seus nivells.

La dinàmica del sistema complex cel·lular respon a l'orquestració de l'activitat de tots els seus elements i components que interaccionen entre ells. En l'exemple anterior s'ha introduït el cas en que es modelitza la generació del fenotip cel·lular en funció de dues partícules moleculars. Des d'un punt de vista formal, aquest sistema es pot representar mitjançant una xarxa amb dos vèrtexs, les dues partícules moleculars, i un conjunt de dues arestes, la interacció de cadascuna d'elles amb l'altra. Generalitzant aquest cas, un sistema complex es pot modelitzar per una xarxa tal que els elements del sistema estan representats pels vèrtexs i les interaccions entre els elements per les arestes. A més a més, aquesta xarxa complexa pot ser la unió de diverses xarxes de classe diferent, cadascuna d'elles representant un nivell diferent

del sistema amb la seva pròpia dinàmica.

Hi ha diferents tipus de xarxes biològiques —xarxes de regulació gènica, xarxes metabòliques, xarxes d'interacció física entre proteïnes, etc.— i cadascuna d'elles queda determinada pel tipus de sistema que modelitza, els seus elements i les interaccions entre ells. Per exemple, les xarxes de regulació gènica representen gens i les seves interaccions. En aquest marc, les interaccions gèniques no necessàriament representen interaccions físiques entre gens, sinó que representen la regulació directa o indirecta entre els gens via proteïnes o altres molècules. Les xarxes es poden construir de diferents maneres, modelitzant el sistema des de punts de vista diferents (Bansal et al., 2007). Un altre exemple és la xarxa de l'interactoma humà que modelitza les interaccions entre proteïnes. En aquesta xarxa no es coneix quina és el sentit del flux d'informació ni de l'ordre cronològic de les interaccions. És a dir, que només indiquen si la interacció es produeix o no (Rolland et al., 2014).

Les xarxes biològiques solen ser diferents de les xarxes aleatòries, ja que alguns dels seus elements solen tenir un caràcter rellevant en la xarxa i que es reflecteix en la topologia de la xarxa. Per altra banda, els elements de les xarxes aleatòries, no solen tenir comportament diferenciats entre ells (Barabasi i Oltvai, 2004).

### 1.3.2 Xarxes aleatòries

La teoria de grafs (xarxes des de la perspectiva matemàtica) aleatoris va ser introduïda per Paul Erdős i Alfréd Rényi a mitjans del segle XX i gràcies als seus estudis sistemàtics va esdevenir una branca de la teoria de grafs. Erdős i Rényi van estudiar els grafs  $G(n, m)$  de  $n$  vèrtexs i  $m$  arestes que es generen distribuïnt aleatòriament les arestes entre les  $\binom{n}{2}$  possibles parelles de vèrtexs i l'evolució de les xarxes quan  $n$  i  $m$  creix i el comportament de les seves propietats usant el mètode probabilístic (Erdős i Rényi, 1960, 1959). El seu mètode ha resultat útil per mesurar l'esdeveniment de propietats en les xarxes. En el seu mètode, es construeix l'espai mostral  $\mathcal{G}(n, m)$  de tots els grafs  $G(n, m)$  de  $n$  vèrtexs i  $m$  arestes, que conté exactament  $|\mathcal{G}(n, m)| = \binom{\binom{n}{2}}{m}$  elements. Llavors, es pot calcular la probabilitat de que una xarxa compleixi una certa propietat  $\mathcal{P}$ . Alguns exemples són l'existència de subgrafs concrets —arbres, cicles, etc.—, la mida de subgrafs, el nombre de components connexes, la mida de la component connexa, o el diàmetre d'un graf (Erdős i Rényi, 1960, 1959; Chung, 2001).

Una altra manera de modelitzar l'evolució d'un graf aleatori  $G(n, p)$  amb  $n$  vèrtexs

és mitjançant el model binomial amb probabilitat  $p$ , que és la probabilitat de que qualsevol parell de vèrtexs estiguin connectats. Essencialment, els grafs aleatoris binomials i els generats pel mètode d'Erdoős i Rényi tenen les mateixes propietats (Newman, 2003).

Segurament, el resultat més conegut derivat dels estudis d'Erdős i Rényi és la distribució del grau d'un graf aleatori que es pot aproximar per una distribució de Poisson de paràmetre  $\lambda \sim \langle k \rangle = \frac{2m}{n}$ , el grau mitjà del graf (Erdős i Rényi, 1960; Bollobás, 2001). D'aquesta manera, en un graf aleatori s'espera que tots els vèrtexs tinguin un grau similar i que s'aglutinin al voltant de  $\langle k \rangle$ , essent molt poc probable els vèrtexs amb un grau molt allunyat de  $\langle k \rangle$ . A més, a causa de la distribució equitativa de les arestes, els vèrtexs són ràpidament assequibles, anant d'un vèrtex a un altre amb pocs passos per a qualsevol parella de vèrtexs. Els camins més curts de  $G$  entre qualsevol parella de vèrtexs, els que defineixen la distància entre els elements de  $G$ , són relativament petits i, en particular, també ho serà el diàmetre de la xarxa  $G$ ,  $\text{diam}(G)$ . Aquesta propietat es coneix amb el nom de món petit (Albert i Barabási, 2002; Watts i Strogatz, 1998; Milgram, 1967) i es caracteritza per  $\text{diam}(G) \sim \frac{\log n}{\log \langle k \rangle}$  (Chung, 2001; Barabasi i Oltvai, 2004; Albert i Barabási, 2002).

Una altra propietat de les xarxes aleatòries és el seu coeficient de clusterització petit. Com que les arestes estan equitativament distribuïdes per a tota la xarxa, és a dir que no hi ha zones més denses que d'altres, el coeficient de clusterització és petit i té un caràcter homogeni, que es pot aproximar per  $p$  (Newman, 2003).

### 1.3.3 Xarxes d'escala lliure

Moltes de les xarxes que modelitzen sistemes reals solen allunyar-se de tenir un comportament aleatori (Barabási i Albert, 1999) i ser riques amb propietats que no tenen les xarxes aleatòries. Una propietat rellevant de les xarxes reals és que solen ser lliure d'escala i la distribució del grau dels seus elements es pot fitar per una funció potencial  $P(X_k) = c(\gamma)k^{-\gamma}$ , on  $c$  és una constant que depèn de  $\gamma$ , típicament amb  $\gamma \in (2, 3)$  (Barabási i Albert, 1999). A diferència de les xarxes aleatòries, que tenen un vèrtex tipus al qual la resta de vèrtexs s'hi aproximen, les xarxes d'escala lliure no en tenen cap ni cap patró en la distribució del grau que distingeixi uns elements dels altres. Sigui quin sigui el grau  $k$ , es té que  $\frac{P(X_{\alpha k})}{P(X_k)} = \frac{c(\gamma)(\alpha k)^{-\gamma}}{c(\gamma)k^{-\gamma}} \sim \alpha^{-\gamma}$  ( $\alpha$  real), és a dir que la quantitat d'elements amb grau  $\alpha k$  respecte del nombre d'elements de grau  $k$  es comporta de manera semblant independentment del grau  $k$  escollit, i és

per això que es diuen d'escala lliure. Les funcions potencials tenen la propietat que en escala logarítmica són lineals amb pendent l'exponent  $-\gamma$ . D'aquesta manera, en una xarxa d'escala lliure s'espera que hi hagi molts elements de grau petit respecte de pocs elements amb un grau gran. Així doncs, les arestes no es distribueixen equitativament i s'espera que hi hagi zones més denses que d'altres o que cada vèrtex clusteritzi en un entorn —encara que aquestes propietats depenen de com es distribueixen els vèrtexs, en particular aquells que són de grau elevat (Newman, 2002). És a dir, que les xarxes d'escala lliure tenen un coeficient de clusterització més gran que les xarxes aleatòries (Barabási i Albert, 1999).

#### 1.3.4 Clústers i mòduls

Els elements de les xarxes reals s'organitzen en mòduls funcionals o a través dels diferents nivells dels sistemes. Per exemple, en les cèl·lules, les proteïnes poden organitzar-se per regular-se elles mateixes en diferents estadis d'una funció cel·lular (alberts, 1998), formar complexos (alberts, 1998) o formar diferents mòduls especialitzats en algun procés concret dins d'una mateixa via de senyalització (Rives i Galitski, 2003). Tanmateix, en una xarxa complexa aquests clústers no defineixen components connexes diferents sinó que interaccionen entre ells i poden formar un superclúster per regular la dinàmica del sistema (Dobrin et al., 2004). La manera que interaccionen aquests clústers defineix la organització de la xarxa (Ravasz i Barabási, 2003).

En una xarxa  $G$ , un mòdul o clúster és un conjunt de vèrtexs que estan connectats entre ells, sovint per algun propòsit funcional en el sistema (Yook et al., 2004) —per exemple, una via de senyalització pot està formada per diferents mòduls que corresponen a diferents clústers de proteïnes, cadascun d'ells especialitzats en algun procés (Rives i Galitski, 2003). Des del punt de vista topològic, els clústers es caracteritzen per tenir més arestes entre els seus elements que amb els elements de fora. Si  $v \in G$ , el coeficient de clusterització  $cl(v)$  de  $v$  és una mesura local que reflecteix la densitat d'arestes en un entorn del vèrtex  $v$  i ho fa comptant la quantitat de triangles que tenen  $v$  com a vèrtex respecte dels possibles triangles.

Com que un conjunt de vèrtexs en un clúster estaran més connectat entre ells, en particular ho estaran prenent els vèrtexs tres a tres, és a dir hi haurà una sobrerrepresentació de triangles de vèrtexs. Tanmateix, els vèrtexs poden definir-se mitjançant altres estructures entre els vèrtexs, com per exemple quadrats. Un subgraf sobre-

representat en una xarxa és un tret característic de la xarxa i s'anomena motiu (Barabasi i Oltvai, 2004).

### 1.3.5 Motius

En una xarxa, un motiu és un subgraf que té una representació més gran del que s'espera per atzar. Les xarxes reals, solen tenir motius concrets que les caracteritzen i que descriuen de la manera que interaccionen i es regulen els seus elements. Per exemple en les xarxes de regulació, els grafs en forma de triangle, llaços *feed-forward*, i de quadrat, llaços *bi-fan*, són motius característics pels quals els seus elements s'autoregulen entre ells (Milo et al., 2002; Dobrin et al., 2004). El nombre de subgrafs dins d'una xarxa d'ordre  $n$  creix exponencialment amb  $n$  i fa que torbar-los sigui un problema irresoluble (Itzkovitz et al., 2003).

En una xarxa, els motius no solen estar isolats uns dels altres sinó que interaccionen entre ells formant un clúster de motius homòlegs que, al mateix temps, clústers amb motius diferents s'uneixen formant un superclúster.

### 1.3.6 Estructura jeràrquica

Les xarxes biològiques tendeixen a tenir les propietats d'escala lliure i clusterització. En una xarxa clusteritzada, cada clúster està relativament aïllat de la resta d'elements de la xarxa. En canvi, en les xarxes d'escala lliure, els *hubs* estan connectats amb molts elements de la xarxa, cosa que fa improbable l'existència de mòduls aïllats. Per què aquestes dues condicions aparentment paradoxals coexisteixin en una xarxa, aquesta s'organitza de manera jeràrquica (Barabasi i Oltvai, 2004). És a dir, els clústers es combinen entre ells forma jeràrquica. Una característica de les xarxes jeràrquiques és que el coeficient de clusterització en funció del grau és  $cl(k) \sim k^{-1}$ . És a dir, els vèrtexs amb pocs veïns clusteritzen mentre que els *hubs* fan d'intermediaris entre els clústers (Ravasz i Barabási, 2003).

El problema de trobar clústers és un problema obert. Els mètodes per trobar clústers es poden basar només en propietats topològiques de les xarxes (Clauset et al., 2004) o integrar la topologia amb les funcions de les proteïnes (Yook et al., 2004). Diferents mètodes poden donar resultats diferents.



### 1.3.7 Robustesa de les xarxes d'interaccions proteiques

Una característica important dels sistemes complexos és la seva robustesa, és a dir la capacitat del sistema per respondre a alteracions, tan internes com externes, per tal de mantenir la seva estructura i el seu funcionament normal. La robustesa d'una xarxa complexa està determinada per la seva topologia. La disfunció (absència) de vèrtexs augmenta, en general, la distància entre els parells de vèrtexs restants, ja que poden alterar-se els camins més curts entre parelles de vèrtexs. El mal funcionament de vèrtexs, té doncs un impacte en la comunicació de la xarxa i, en general, augmenta el diàmetre d'aquesta o, inclús, pot afectar a la seva connectivitat disseminant-se en components connexes més petites.

En una xarxa Erdős-Rényi, en la qual els seus elements tenen un comportament homogeni, tots els elements contribueixen d'una manera semblant a la robustesa de la xarxa, és a dir, el mal funcionament de cada vèrtex provoca la mateixa quantitat de danys en la xarxa. Si el nombre de vèrtexs que es suprimeixen és superior a un llindar crític, que s'ha provat ser el 30% de vèrtexs, la xarxa sofreix danys importants i la component principal de la xarxa s'esmicola en components més petites (Albert et al., 2000).

Per altra banda, la resposta de les xarxes Barabási-Albert a danys aleatoris, per la seva topologia, no depèn de cap llindar crític i poden suportar una quantitat gran de vèrtexs suprimits conservant la seva integritat. Com que en aquestes xarxes la majoria de vèrtexs tenen grau petit, quan un vèrtex pateix un mal funcionament, per atzar serà un vèrtex poc central. En canvi, a diferència de les xarxes Erdős-Rényi, aquestes xarxes mostren una vulnerabilitat gran quan els vèrtexs són seleccionats segons la seva centralitat. En aquest cas, la resposta de la xarxa sí que depèn d'un llindar crític, que s'ha provat ser el 18% de vèrtexs, que una vegada superat, la xarxa perd la seva integritat disseminant la component principal en components connexes més petites a mesura que la quantitat de vèrtexs suprimits augmenta (Albert i Barabási, 2002).

La robustesa topològica de la xarxa que modelitza un sistema pot reflectir-ne la robustesa funcional. Jeong et al. van provar que els gens essencials en el llevat tendeixen a codificar proteïnes amb un nombre elevat d'interaccions, és a dir, a ocupar un lloc central de la xarxa, mentre que el nombre de proteïnes essencials amb poques interaccions és baix. Goh et al. van estudiar la relació entre els gens associats a malalties humanes, gens essencials i la centralitat de les proteïnes que codifiquen

aquests gens dins de la xarxa d'interaccions proteiques. Van veure també que els gens essencials codifiquen proteïnes centrals. En canvi, els productes d'aquells gens associats a malalties que no són essencials no tenen aquest comportament d'ocupar posicions centrals. Van integrar també els nivells d'expressió gènica provant que aquells gens l'activitat dels quals es sincronitza amb la resta de gens tendeixen a ser gens essencials, mentre que els gens no essencials associats a malalties mostren el patró complementari, amb una correlació dels nivells negativa (Goh et al., 2007).

### 1.3.8 La centralitat subgraf. L'índex d'Estrada

La centralitat *betweenness* d'un vèrtex d'una xarxa és una mesura local que reflecteix la influència del vèrtex entre la comunicació de dos vèrtexs qualssevol de la xarxa (Freeman, 1977). Aquesta mesura està basada en la idea que la informació passa a través dels camins més curts entre dos vèrtexs. És a dir, que reflecteix el nombre de camins més curts entre dos vèrtexs qualssevol que passen pel vèrtex considerat. Tanmateix, la informació en una xarxa no ha de viatjar, necessàriament, pels camins més curts, sinó que pot anar d'un vèrtex a un altre mitjançant camins aleatoris. En aquest sentit, la centralitat *betweenness* pot ometre informació rellevant entre la comunicació de vèrtex en una xarxa.

La centralitat subgraf és una mesura local de centralitat dels vèrtexs d'una xarxa que considera els camins de qualsevol longitud. Tanmateix, per emfatitzar la importància que poden tenir els camins més curts en una xarxa, penalitza els camins en funció de la seva longitud. Sia  $G$  una xarxa. La centralitat subgraf d'un vèrtex  $i$  de la xarxa  $G$  es defineix com

$$C_G(i) = \left(A^k\right)_{ii} \quad (1.1)$$

on  $A$  és la matriu d'adjacència de la xarxa  $G$ . És a dir, l'entrada  $(i, i)$  de l'exponencial de la matriu d'adjacència que, per definició de la matriu exponencial, es pot escriure com

$$C_G(i) = \sum_{k \geq 0} \frac{(A^k)_{ii}}{k!}. \quad (1.2)$$

Com que l'entrada  $(i, i)$  de la potència  $k$ -èsima de la matriu d'adjacència  $A^k$  és el nombre de camins tancats de longitud  $k$  amb origen el vèrtex  $i$  (Biggs, 1993), es pot pensar que la mesura  $C_G(i)$  està basada en els camins de totes les longituds possibles amb origen  $i$ , és a dir, que considera tots els subgrafs amb vèrtex  $i$ . Tanmateix, amb

la idea que els camins més curts són més centrals entre la comunicació de vèrtexs en una xarxa, la centralitat subgraf penalitza els camins més llargs en funció de la seva longitud mitjançant  $k!$ .

L'índex d'Estrada és una mesura global de la xarxa que es defineix a partir de la centralitat subgraf dels seus vèrtexs  $C_G(i)$  i es defineix com

$$C(G) = \sum_{i \in V} C_G(i) = \sum_{k \geq 0} \frac{(A^k)_{ii}}{k!} \quad (1.3)$$

on  $V$  és el conjunt de vèrtexs de la xarxa  $G$ .

## 1.4 L'interactoma proteic humà

A diferència de l'actual mapa genòmic humà, en el qual hi ha descrits gairebé la totalitat dels gens que codifiquen proteïnes (Dunham et al., 2012), en el cas proteic hi ha poques proteïnes i interaccions proteiques descrites (Kim et al., 2014; Organization, 2010). En la última dècada han sorgit diverses iniciatives i projectes per descriure el conjunt de proteïnes humanes (o proteoma) (Organization, 2010; Paik et al., 2012) i l'interactoma (Rolland et al., 2014; Rual et al., 2005a; Stelzl et al., 2005). Malgrat que en l'actualitat els mapes interactòmics són incomplets, han estat útils per predir noves funcions proteiques, noves interaccions o per estudiar el fenotip i causes moleculars d'una malaltia, o noves dianes terapèutiques.

L'interactoma proteic humà és (i) el conjunt de les proteïnes humanes que s'expressen en qualsevol dels tipus cel·lulars diferents que formen el cos humà (proteoma) i (ii) les seves interaccions físiques. Les majors dificultats per descriure l'interactoma recauen en la complexitat de descriure el proteoma i determinar el gran nombre d'interaccions. El producte d'un gen pot variar tan en el tipus de producte (transcrits, proteïnes, etc.) com en el nombre d'elements diferents de cada tipus de producte. Un exemple d'això és el nombre de proteïnes diferents que es poden sintetitzar a partir d'un mateix gen. Per causes diferents, entre elles l'*splicing alternatiu* o mutacions posteriors a la translació, un gen pot expressar diferents isoformes d'una proteïna. A més a més, encara no es coneixen quants gens expressen una sola proteïna ni se n'han descrit, de manera completa, els mecanismes pels quals ho fan. Així doncs, a més a més de la dificultat de conèixer tots els productes canònics de cada gen, s'hi afegeix la diversitat del producte a causa dels mecanismes d'expressió o per

alteracions o mutacions de les cadenes d'ADN. Si s'obvia aquest fet i es considera una variant proteica representativa per a cada gen, s'obtindrien aproximadament 20 000 proteïnes (Kim et al., 2014) i de l'ordre de  $10^8$  interaccions possibles.

Hi ha diversos paràmetres que caracteritzen la fidelitat en que una xarxa representa l'interactoma proteic (Rual et al., 2005b). La mida de la mostra de l'espai interactòmic, la sensibilitat de l'assaig i del mostreig, i la precisió —falsos positius i negatius. Per poder construir una xarxa d'interaccions proteiques optimitzant aquests paràmetres, s'ha de fer de forma sistemàtica i prenent una mostra d'elements distribuïda uniformament en l'espai interactòmic i que no estigui esbiaixada per la localització, processos biològics, etc. (Rual et al., 2005b; Rolland et al., 2014). Tanmateix, amb les tecnologies actuals no és possible prendre una mostra gaire gran de l'espai interactòmic (Rolland et al., 2014). Un mètode alternatiu per construir una xarxa que millora el recobriment de l'espai interactòmic és per recopilació de resultats en la literatura científica. Aquest mètode però, introdueix diferents tipus de biaixos com la quantitat d'articles publicats per gen-proteïna —distribució no uniforme de les interaccions, ni dels processos biològics associats, ni de la localització— o diferents tipus de tècniques usades en els experiments —diferents comportaments dels falsos positius i negatius— que poden afectar a la topologia de la xarxa respecte la de l'interactoma.

Amb tot, encara s'està molt lluny d'aconseguir un mapa prou fidel de l'interactoma proteic humà i inferir-ne el caràcter general a partir de les xarxes que es poden construir actualment (Stumpf et al., 2005; Han et al., 2005). Alguns autors han xifrat la mida de l'espai interactòmic amb 130 000 elements —prenent una sola variant proteica per gen codificant proteïna (Rual et al., 2005b). Si fos així, actualment se'n recobreix entre el 5% i 25%, segons el mètode i els criteris escollits (Rolland et al., 2014; Keshava Prasad et al., 2009; Xenarios et al., 2000; Orchard et al., 2014; Stark et al., 2006).

Tradicionalment s'han descrit les xarxes que modelitzen l'interactoma proteic com xarxes d'escala lliure. Tot i que les xarxes que modelitzen interactomes proteics es poden aproximar a xarxes d'escala lliure, aquest comportament pot ser a causa de la quantitat de proteïnes i interaccions considerades i no es pot extrapolar a la topologia de l'interactoma sencer. De xarxes que no són d'escala lliure, en particular aleatòries i exponencials, es poden generar subxarxes prenent vèrtexs i arestes per atzar amb una topologia d'escala lliure (Han et al., 2005).

Malgrat que els estudis amb bioxips han revolucionat els estudis de malalties i han introduït la dimensió molecular perfeccionant els mètodes clínics clàssics a l'hora de fer pronòstics de risc de malalties com el càncer o classificar els tumors (Perou et al., 2000; van de Vijver et al., 2002), aquesta eina només introdueix informació d'un nivell del sistema biològic concret i no cobreix tota la diversitat biològica del sistema en el qual s'emmarca una malaltia com qualsevol dels càncers. Diferents estudis d'un mateix càncer infereixen signatures de gens que sovint comparteixen pocs gens, si no cap, encara que les prediccions finals són similars (Chang i Nevins, 2006; Chuang et al., 2007; Shen et al., 2008). Encara més, això passa inclús estudiant amb metodologies diferents un mateix conjunt de dades de càncer de mama (Fan et al., 2006). Així, hi ha una certa diversitat biològica en la redundància de l'activitat cel·lular. És per això, que signatures diferents per a mètodes diferents poden pronosticar comportaments similars per a una mateixa malaltia, ja que encara que els gens no siguin els mateixos, sí que reflecteixen els mateixos patrons, o molt similars, de l'activitat cel·lular i activació de les vies de senyalització (Shen et al., 2008). En aquest marc doncs, no té sentit identificar biomarcadors individuals com a responsables del desenvolupament o progressió d'un càncer, sinó que és més acurat identificar conjunts de gens o elements del sistema que, d'alguna manera, la seva activitat està relacionada en les condicions donades. Una manera de complementar la informació d'expressió gènica és integrant informació d'interaccions entre proteïnes o la xarxa de l'interactoma proteic que poden ajudar a revelar conjunts de gens implicats en el desenvolupament o progressió d'un càncer (Chuang et al., 2007).

La progressió del càncer es produeix mitjançant l'adquisició de mutacions gèniques que alteren els mecanismes de la cèl·lula. L'efecte d'aquestes mutacions són canvis de perfils de l'expressió gènica cel·lular. Aquests canvis d'expressió gènica poden reflectir-se en augments o disminucions en el producte proteic i, per tant, en la inhibició o promoció de senyals a través de l'interactoma (Taylor et al., 2009). Tanmateix, per la mateixa complexitat de l'interactoma, els patrons topològics associats a la dinàmica del càncer encara no són coneguts.



# 2

## Introducció al càncer

### 2.1 Mecanismes moleculars involucrats en el càncer

En les darreres dècades, la incidència del càncer ha augmentat fins a ser la família de malalties amb més incidència, inclús per davant de les malalties cardiovasculars (Ferlay et al., 2015). Durant el 2012 es van comptabilitzar 14.1 milions de casos nous i 8.2 milions de morts a causa del càncer.

Les causes del desenvolupament i progressió del càncer són diverses i complexes. L'edat en sol ser un factor important, ja que sol ser un procés que requereix temps i que es desenvolupa al llarg de la vida d'un individu. Les causes poden ser des d'ambientals —exposició a agents mutàgens, hàbits culturals com l'alimentació, etc.— fins a causes genètiques i moleculars. A més, aquestes causes poden estar relacionades entre elles augmentant la complexitat de la malaltia.

Cadascun dels càncers depèn del teixit a on s'ha desenvolupat però també del genotip i fenotip cel·lular de l'individu, que pot diferenciar càncers del mateix tipus o teixit mitjançant signatures moleculars diferents, en les quals també hi pot haver influït els hàbits culturals de l'individu. Des d'aquesta perspectiva, el càncer és una família de malalties complexes que totes elles es caracteritzen per (i) una proliferació cel·lular descontrolada i (ii) la capacitat de les cèl·lules d'envair teixits veïns o emigrar i colonitzar teixits llunyans a on replicaran nous tumors, és a dir de metastatitzar (Stratton et al., 2009; Hanahan i Weinberg, 2000)

Cadascun dels càncers s'origina en una sola cèl·lula que ha adquirit una sèrie d'avantatges evolutius que li confereixen una major supervivència i proliferació respecte de les altres cèl·lules normals o no canceroses. Aquests avantatges que adquireix la cèl·lula són a causa de les mutacions somàtiques que han patit i adquirit les seves cèl·lules ancestrals. Una cèl·lula cancerosa és el resultat d'un llinatge cel·lular que ha

acumulat les mutacions (somàtiques) suficients per les quals la cèl·lula descendent esdevé maligna.

Les cèl·lules del cos humà, tan aquelles que són normals com les que mostren algun tipus de fenotip de neoplàsia alterada, pateixen constantment alteracions en el seu ADN, tan per causes internes com per agents mutàgens externs. El mateix sistema cel·lular corregeix moltes d'aquestes alteracions però algunes d'elles, poques, són capaces de convertir-se en mutacions fixes en el genoma cel·lular. Una vegada que una cèl·lula ha adquirit una mutació, aquesta s'estendrà pel llinatge cel·lular alhora que se n'incorporaran de noves i la cèl·lula evolucionarà cap a una cèl·lula cancerosa. A mesura que una cèl·lula adquireix mutacions, augmenta la inestabilitat genòmica que predisposa a la cèl·lula a adquirir-ne més (Stratton et al., 2009; Hanahan i Weinberg, 2011).

Les causes que provoquen una mutació en l'ADN cel·lular són diverses i complexes. Els mateixos mecanismes moleculars poden provocar mutacions en l'ADN cel·lular —per exemple, errors en la replicació de l'ADN. L'exposició a agents mutàgens externs també augmenta el risc de patir mutacions —l'exposició al fum del tabac o a la radiació ultraviolada. Altres factors que tenen un paper important en el risc de càncer són les malalties poc freqüents hereditàries, que augmenten la susceptibilitat de patir mutacions, o els virus, que solen introduir fragments sencers d'ADN en la cèl·lula. Aquests casos solen està relacionats amb tipus de càncers concrets —per exemple el càncer de mama a causa de l'anèmia de Fanconi o el càncer de cèrvix a causa del virus del papil·loma humà (D'Andrea i Grompe, 2003; D'Andrea, 2010).

Les cèl·lules canceroses també adquireixen resistència terapèutica a causa de les mutacions gèniques, ja sigui adquirint-les abans o com a resposta de la teràpia. S'ha observat que una cèl·lula cancerosa pot restar en una fase de repòs fins que en unes condicions d'entorn adequades, el seu comportament maligna pot tornar a recidivar i/o desenvolupar metàstasis (Stratton et al., 2009).

## 2.2 Característiques bàsiques del càncer

En el teixit normal, format per cèl·lules normals, les cèl·lules estan en equilibri o sincronitzades amb les cèl·lules del seu entorn per tal de dur a terme els processos o funcions necessàries per mantenir la pròpia integritat i, també, la del teixit. L'activitat cel·lular està orquestrada per xarxes complexes de fluxos de senyals que tan



poden provenir de l'exterior i interioritzar-lo com ser fluxos interns de la cèl·lula. Per exemple, si la cèl·lula detecta factors de creixement, es generen cascades de senyals que participen en la regulació del creixement cel·lular. En la carcinogènesi, les cèl·lules perden progressivament aquest equilibri i guanyen una certa independència respecte del seu entorn, autoregulant-se amb criteri propi. D'aquesta manera poden evitar els senyals i mecanismes cel·lulars que reparen els danys en la cèl·lula o inclús el programa de mort cel·lular. Per exemple, quan l'expressió dels oncogenes augmenta, i amb aquests les senyals de les seves proteïnes, també ho fa la proliferació cel·lular i el creixement tumoral. Tanmateix, si l'augment de senyal en una cèl·lula és massa gran, aquesta pot ser induïda cap a la senescència o la mort programada, com és el cas de les oncoproteïnes RAS o MYC (Hanahan i Weinberg, 2011).

La cèl·lula que s'ha transformat en una cèl·lula cancerosa ha estat capaç de superar o evadir aquestes limitacions per tal d'autoregular-se ella mateixa. Les mutacions poden afectar de diferents maneres en els mecanismes de la cèl·lula. Normalment, quan una cèl·lula adquireix una mutació, afecta a nivell molecular activant-se o desactivant-se vies de senyalització associades a mecanismes cel·lulars. Un dels mecanismes que té la cèl·lula per regular els nivells de proteïna són els llaços de retroalimentació negativa. Hi ha oncogenes que la seva activitat està regulada per aquest tipus de llaços de manera que atenua els senyals de les seves oncoproteïnes si aquests són massa elevats, per tornar-los a nivells normals. Una mutació pot alterar aquests mecanismes de regulació i, com a conseqüència, els gens supressors de tumors poden alterar la seva activitat o que augmentin els nivells de les oncoproteïnes. D'aquesta manera, a mesura que les cèl·lules adquireixen mutacions amb pèrdua o canvis de funcions, també adquireixen les característiques suficients per tal de tenir una proliferació ràpida i il·limitada i poder envair teixits veïns o metastatitzar a teixits llunyans.

A l'any 2000 Hanahan i Weinberg van descriure els sis trets característics de les cèl·lules canceroses (Hanahan i Weinberg, 2000):

- senyalització de proliferació sostinguda,
- evasió dels supressors de creixement,
- resistència a la mort cel·lular,
- activació de la replicació il·limitada,
- inducció a l'angiogènesi i

- activació de la invasió i metàstasi.

Deu anys més tard, Hanahan i Weinberg van proposar

- la inestabilitat genòmica i
- la inflamació promotora de tumors

com dos factors que faciliten l'adquisició de les característiques necessàries de les cèl·lules canceroses (Hanahan i Weinberg, 2011). La inestabilitat genòmica es sol adquirir mitjançant el sistema de manteniment de l'ADN mitjançant el desgast dels telòmers que protegeixen els extrems dels cromosomes i queden exposats a patir mutacions.

A més, Hanahan i Weinberg van proposar, també, dos trets característics emergents de les cèl·lules canceroses (Hanahan i Weinberg, 2011):

- evasió de la destrucció immunològica i
- reprogramació del metabolisme d'energia cel·lular.

El sistema immunològic és responsable de reconèixer i eliminar la majoria de cèl·lules canceroses incipients. Des d'aquesta lògica, els tumors sòlids que apareixen han aconseguit evadir, d'alguna manera, el control del sistema immunològic. S'ha observat, que en individus immunocompromesos augmenta la incidència de càncers (Hanahan i Weinberg, 2011). Així, el sistema immunològic podria actuar de barrera per al desenvolupament de tumors, eliminant les cèl·lules canceroses.

## 2.3 Estadis del càncer de mama

El càncer de mama és el segon càncer més freqüent a nivell mundial i el més freqüent en les dones amb una incidència de 1.67 milions el 2012 i amb una mortalitat total de 522 000 defuncions ocupa el segon lloc del rànquing pels països desenvolupats, mentre que és el primer càncer amb més mortalitat en els països subdesenvolupats (Ferlay et al., 2015).

Com la majoria dels càncers, el càncer de mama es desenvolupa i progressa mitjançant l'adquisició de mutacions en el genoma cel·lular que confereix a la cèl·lula avantatges evolutius mitjançant canvis en els seus mecanismes. S'estima que entre el 90% i el 95% dels càncers són a causa de mutacions somàtiques mentre que la resta són a causa de les mutacions germinals, aquelles que l'individu ha heretat a través

dels gàmetes dels progenitors (Laloo i Evans, 2012). El càncer de mama progressa a través de diversos estadis clínics i patològics iniciant-se, generalment, en els conductes mamaris i d'origen epitelial. L'estadi hiperplàsia ductal atípica (HDA) és una alteració en la qual les cèl·lules poden mostrar una proliferació descontrolada però sense cap caràcter maligna. L'estadi carcinoma ductal *in situ* (CDIS) és l'estadi en el qual les cèl·lules es desregulen i proliferen de manera descontrolada però no surten de la seva localització. Es considera un estadi precursor del carcinoma ductal invasiu (CDI), en el qual les cèl·lules ja han adquirit la capacitat de sortir del teixit a on s'ha originat i poden invadir teixits veïns. EL CDI és el tipus de càncer de mama més usual, amb una freqüència del 80% dels casos, aproximadament (Weigelt et al., 2005b). La metastasi (M) es considera el pas següent del CDI. En aquest estadi, les cèl·lules han arribat als nodes limfàtics i han aconseguit penetrar en els sistema limfàtic o penetrar en els vasos sanguinis. Les cèl·lules canceroses es disseminen pel cos de l'individu colonitzant nous teixits o òrgans a on generaran nous tumors (Weigelt et al., 2005b).

## 2.4 Teràpies del càncer

El desenvolupament i progressió del càncer és un procés lent i complex en el qual la cèl·lula va adquirint mutacions que alteren l'estructura de les seves xarxes de senyalització i que es manifesten amb l'adquisició d'avantatges evolutius mitjançant guanys o pèrdues de funcions cel·lulars. Així doncs, més enllà de que uns quants gens hagin rebut una o més mutacions, la carcinogènesi és la causa d'una desregulació de les xarxes de senyalització cel·lular. Les vies de senyalització formen una estructura complexa de xarxes en el qual no són independents unes de les altres, sinó que una molècula o proteïna pot formar part de diverses vies. Així, les vies de senyalització no operen isolament sinó que estan orquestrades formant patrons complexos en els quals un canvi en una via pot tenir un efecte transversal i modificar els patrons d'activitat de diverses vies (Kreeger i Lauffenburger, 2010), ja sigui directament —via interaccions proteïna-proteïna— o indirectament —via canvis transcripcionals.

A causa de la complexitat del desenvolupament i progressió del càncer a nivell cel·lular, les teràpies contra als càncers tenen, generalment, una efectivitat menor a la desitjada (Janku et al., 2012). Un dels tractaments estàndards és la quimioteràpia per citotoxicitat que danya la cèl·lula induint-la a la mort cel·lular. Tanmateix, aquests tipus de teràpies no només afecten a les cèl·lules canceroses sinó que també

ataquen a les cèl·lules normals, produint efectes secundaris no desitjats. A més, les cèl·lules canceroses solen mostrar resistència a aquestes teràpies, ja sigui per la mateixa complexitat cel·lular i heterogeneïtat molecular o adquirida mitjançant la resposta a la teràpia (Holohan et al., 2013).

Els fàrmacs dirigits a dianes específiques poden millorar els tractaments contra als càncers, tan per l'efectivitat com pels efectes secundaris, mitjançant la identificació de marcadors i dianes concretes. En aquest sentit poden ser útils les anàlisis integrant dades de diferents nivells del sistema cel·lular. Per exemple, amb l'anàlisi de les dades d'expressió gènica associada al càncer o a respostes terapèutiques es poden identificar patrons d'expressió gènica associats al desenvolupament o progressió del càncer (Sadanandam et al., 2013). A partir d'aquests patrons es poden definir dianes concretes derivades d'oncogens o que formen part de vies de regulació el mecanisme de les quals està associada al càncer. Per exemple, pacients amb càncer de mama amb el protooncogen *PIK3CA* mutat —activador la via PI3K-AKT-mTOR— que no responien a teràpies estàndards contra al càncer sí que han respost a teràpies dirigides a inhibir la via PI3K-AKT-mTOR (Janku et al., 2012; Juvekar et al., 2012). Tanmateix, els fàrmacs dirigits a dianes concretes poden resultar insuficients ja que van dirigits a mecanismes molt concrets de la cèl·lula. Una manera de cobrir un rang més ampli de mecanismes cel·lulars és mitjançant fàrmacs amb múltiples dianes o combinació de fàrmacs dirigits a diferents vies de senyalització (Juvekar et al., 2012).

Les cèl·lules canceroses solen respondre als tractaments mostrant o generant resistència als fàrmacs evadint les teràpies de diverses maners. Moltes de les quimioteràpies provoquen danys en l'ADN induint a la cèl·lula a la mort cel·lular. Tanmateix, les cèl·lules sovint han adquirit una major eficàcia en els mecanismes de reparació atenuant els efectes de la quimioteràpia. En les teràpies dirigides a dianes concretes, les dianes poden patir mutacions provocant una menor eficàcia del fàrmac o, inclús, augmentar els nivells de l'expressió gènica fent perdre l'eficàcia dels inhibidors a causa de l'augment del nombre de molècules. Les cèl·lules canceroses també poden respondre a la teràpia generant resistència mitjançant mecanismes més complexos. Per exemple, les cèl·lules canceroses poden respondre a teràpies dirigides amb diana EGFR generant resistència mitjançant l'activació de la via PI3K-AKT. En aquest cas però, la diana no s'altera i el fàrmac la segeix inhibint. La cèl·lula cancerosa, supleix els mecanismes inhibits mitjançant la via PI3K-AKT.

Una manera d'augmentar l'eficàcia de les teràpies o tractaments, és mitjançant la

combinació de fàrmacs. Per exemple, moltes de les quimioteràpies citotòxiques són més efectives en combinació amb altres fàrmacs (Chabner i Roberts, 2005). L'objectiu de la combinació de fàrmacs és trobar una teràpia alternativa amb un efecte additiu o sinèrgic atacant a mecanismes cel·lulars diferents provocant la menor toxicitat possible. Tanmateix, la combinació de fàrmacs és un problema complex a causa de l'infinitat de combinacions possibles. Per exemple, Juvekar et al. han observat, mitjançant models *in vivo* de càncer de mama amb la via PI3K activada, que la teràpia amb l'inhibidor NVP-BKM120 de PI3K augmentava l'activitat de MAPK i els nivells de PARP i que malgrat el tractament amb olaparib per inhibir PARP no tenia efecte en l'activitat de MAPK, la combinació dels dos inhibidors tenia un efecte sinèrgic en la reducció del tumor, suggerint que d'alguna manera la inhibició de PARP evita les rutes alternatives de PI3K cap a MAPK (Juvekar et al., 2012). Tanmateix, el problema de la combinació és un problema complex a causa de la quantitat possible de combinacions. Per exemple, si es consideren, aproximadament, els 1 200 fàrmacs contra el càncer que estan en desenvolupament, la combinació de dues drogues dóna 719 400 combinacions possibles, mentre que combinacions de tres fàrmacs dóna més de 287 milions de combinacions possibles. Una manera d'acotar aquest nombre de combinacions és mitjançant la integració de dades biològiques dels diferents nivells del sistema cel·lular.

El càncer progressa mitjançant l'adquisició de mutacions que alteren els seus mecanismes. Aquestes mutacions es reflecteixen a diferents nivells del sistema. Per exemple, a mesura que l'entropia genòmica augmenta també ho fa l'entropia del transcriptoma (van Wieringen i van der Vaart, 2011). La resposta de la cèl·lula al tractament, també es reflecteix en els nivells del sistema cel·lular (Breitkreutz et al., 2012) i la integració de dades d'expressió amb dades d'interaccions han sigut útils per identificar fàrmacs amb dianes concretes (Komurov et al., 2012).



## Part II

# Assumpcions. Objectius





# 3

## Assumpcions

### 3.1 Assumpcions

- Malgrat que els gens poden codificar per diferents productes i que les proteïnes poden ser ràpidament degradades o durar molt de temps en la cèl·lula i que podria afectar en la representació de l'interactoma, s'assumeix que hi ha una correlació positiva entre l'expressió gènica i els nivells de proteïna en la cèl·lula que es manté.
- Tal i com s'ha introduït, els gens poden codificar per diferents isoformes proteiques. Quan es construeix l'interactoma proteic humà, s'identifiquen les isoformes en una sola proteïna. És a dir, tenim una relació biunívoca entre gens i proteïnes.
- Es descarten els homodímers, és a dir els llaços en la xarxa.



# 4

## Objectius I

El principal objectiu d'aquesta part del treball és caracteritzar la relació entre el desenvolupament i progressió del càncer i la topologia de l'interactoma humà.

Més concretament, mitjançant l'ús del model de cascada de fallades,

- identificar els patrons topològics relacionats als estadis progressius del càncer,
- identificar els patrons topològics en els estadis del càncer respecte de l'estadi inicial,
- identificar els patrons topològics relacionats a la pertorbació gènica i la resposta al tractament,
- identificar els patrons topològics diferents o oposats als patrons relacionats al càncer i descriure'n els mecanismes involucrats.



# 5

## Objectius II

El principal objectiu d'aquesta part del treball és caracteritzar el càncer mitjançant l'activitat de la xarxa de l'interactoma humà associada al càncer.

Més concretament,

- definir una mesura que capturi l'activitat de la xarxa de l'interactoma associada al càncer,
- capturar els mecanismes involucrats en el càncer a través de la xarxa de l'interactoma mitjançant la mesura definida,
- capturar l'impacte i les diferències de tractaments contra el càncer en la xarxa de l'interactoma mitjançant la mesura definida,
- capturar l'impacte i les diferències de tractaments contra el càncer entre l'estatus de mutació dels gens conductors del càncer en la xarxa de l'interactoma mitjançant la mesura definida,
- identificar fàrmacs amb mecanismes d'acció complementaris que la seva combinació generi sinergia, mitjançant la mesura definida.



## **Part III**

# **Mètodes**





# 6

## Materials i Mètodes I

### 6.1 La xarxa de l'interactoma

Per representar l'interactome proteic humà s'han construït dues xarxes usant dades de dues bases de dades diferents, la versió núm. 7 de *Human Protein Reference Database* (HPRD) (Keshava Prasad et al., 2009) i la versió del 29/09/2011 de *IntAct Molecular Interaction Database* (IntAct) (Orchard et al., 2014). Ambdues bases de dades contenen informació d'interaccions proteiques recopilada de la literatura científica assegurant, així, una evidència experimental per a cada interacció.

**Taula 6.1.1:** Exemple del fitxer de les interaccions HPRD. Per construir la xarxa, s'han considerat aquelles proteïnes anotades per Entrez (Maglott et al., 2005) (columnes 2 i 4)

Protein A HPRD identifier	Protein A Entrez Gene	Protein B HPRD identifier	Protein B Entrez Gene	Experiment Type	PubMed
02995	4089	04484	4093	yeast 2-hybrid	16189514,15231748
01898	7429	00017	71	in vitro	3510866,6893424
03014	1173	03690	1856	yeast 2-hybrid	16189514
02319	10134	00017	71	in vivo	10958671
02458	9087	00017	71	in vitro	8416954
02549	4130	00017	71	in vitro	7820861
03014	1173	04318	10320	yeast 2-hybrid	16189514
02991	4306	00017	71	in vitro	8612804

Per construir les xarxes d'interaccions proteiques humanes, s'han filtrat les dades (exemple de les dades HPRD Taula 6.1.1) prenent només aquelles proteïnes anotades per Entrez (Maglott et al., 2005). S'han agafat les identificacions Entrez per etiquetar els vèrtexs de la xarxa. Per construir la xarxa s'ha usat el paquet `networkx` (Hagberg et al., 2008) del llenguatge Python. Una vegada construïda la xarxa de les dades filtrades, s'ha agafat la component principal i se li han eliminat els llaços. S'ha definit la xarxa de l'interactoma a partir de la subxarxa obtinguda.

## 6.2 Dades d'expressió gènica

### 6.2.1 Font de les dades

Les dades d'expressió gènica del càncer de mama i del càncer de colon s'han obtingut del repositori de dades *Gene Expression Omnibus* (GEO) (Barrett et al., 2013). Les dades d'expressió gènica contenen els nivells de producte gènica en la cèl·lula en diferents mostres de tipus de teixits diferents.

### 6.2.2 Dades d'expressió del càncer de mama

En el cas del càncer de mama, els teixits de les mostres són

- normal (N),
- hiperplàsia ductal atípica (HDA),
- carcinoma ductal *in situ* (CDIS),
- carcinoma ductal invasiu (CDI),
- metàstasi (M).

Les dades d'expressió gènica corresponen a diversos experiments realitzats amb el bioxip *Affymetrix Human Genome U133 Plus 2.0 Array*. Les referències de les dades, per comparar els nivells de producte gènica entre aquestes condicions, són

- GSE16873 (Emery et al., 2009), N→HDA,
- GSE14548 (Ma et al., 2009), N→DCIS,
- GSE3744 (Richardson et al., 2006), N→CDI,
- GSE3893 (Schuetz et al., 2006), CDIS→CDI,
- GSE2741 (Weigelt et al., 2005a), CDI→M.

Els bioxips contenen fragments curts d'ADN que corresponen a fragments de gens. Cada gen està representat per més d'una sonda. Per quantificar els nivells de producte gènica, s'introdueix el material cel·lular de la mostra en el bioxip i s'hibriden amb les sondes aquells fragments complementaris als fragment de les sondes. Les hibridacions estan marcades amb un tint que es quantifica amb un escàner. La quantificació dels nivells del producte gènica en la cèl·lula es fa a partir dels nivells d'intensitat dels tints.

Per identificar aquelles sondes que amb canvis significatius del nivell del producte gènic entre les condicions, en aquest cas mostres de teixit normal o cancerós, s'ha usat el mètode SAM mitjançant el paquet samr de R (Tusher et al., 2001). El mètode SAM compara els nivells de producte gènic mitjançant l'estadístic *t-student* modificat

$$\frac{\bar{x}_i^1 - \bar{x}_i^2}{s_i + s_0}$$

on  $x_i$  correspon als nivells d'expressió de la sonda  $i$  i els superíndexs 1 i 2 corresponen a les dues condicions que es comparen. En el denominador,  $s_i$  correspon a la desviació estàndard dels nivells de la sonda  $i$ . L'estadístic es corregeix amb la constant  $s_0$  —determinada pel mateix mètode— per corregir les desviacions petites i evitar que les sondes amb variabilitat petita no es detectin com a sondes significatives per error. Per mesurar la significança de l'estadístic, és a dir la diferència dels nivells entre les condicions, realitza un test de permutacions en les etiquetes de les condicions i genera una distribució empírica d'estadístics que s'usa per calcular la significança de l'estadístic observat.

Quan s'han obtingut les sondes significatives (amb un nivell de significança del 5%), s'identifiquen els gens que representen. Com que en el bioxip hi ha diverses sondes per un mateix gen, en el cas d'obtenir diverses sondes significatives per un mateix gen, s'ha considerat la sonda amb el canvi més extrem. D'aquesta manera, s'ha obtingut una llista de gens i el signe del canvi entre les condicions, és a dir si augmenta (es sobreexpressa) o disminueix (infraexpressa) el nivell del seu producte gènic.

### 6.2.3 Dades d'expressió del càncer colorectal

En el cas del càncer colorectal, els teixits de les mostres corresponen a

- normal (N),
- adenoma (A),
- carcinoma (C).

Per comparar els nivells de producte gènic entre les condicions del càncer colorectal, s'ha usat l'experiment amb referència GSE4183 (Galamb et al., 2010).

Per identificar els gens diferencialment expressats, s'ha procedit de manera anàloga al cas del càncer de mama, afegint la condició de que el canvi de nivells entre les

condicions sigui de factor 2, és a dir més del doble o menys de la meitat.

### 6.3 Dades d'expressió de les cèl·lules MCF7

Pel cas de l'estudi MCF7 xeografts s'han usat les dades de l'experiment amb referència GSE7327 (Massarweh et al., 2008).

Per calcular els gens diferencialment expressats s'ha procedit anàlogament al cas de les dades d'expressió del càncer de mama.

### 6.4 Dades d'expressió de les cèl·lules MCF7 transfectades amb SCNA

L'anàlisi d'expressió gènica entre les cèl·lules MCF7 i MCF7 transfectades s'ha realitzat usant les dades GSE31180 (Serra-Musach et al., 2012), i s'ha procedit de manera anàloga al cas del càncer de mama.

### 6.5 Mètode de la cascada de fallades

Un vèrtex de la xarxa pateix una fallada quan aquest vèrtex no és funcional. Des del punt de vista topològic significa que és suprimit de la xarxa, juntament amb les arestes incidents. Cada vèrtex de la xarxa té una càrrega inicial i una capacitat de la càrrega. Quan un vèrtex falla, la seva càrrega es transmet als seus veïns, de manera uniforme respecte la càrrega dels veïns, no del nombre d'interaccions. Els veïns del node que falla, incrementen la seva càrrega amb la càrrega transmesa. Si la nova càrrega del vèrtex és superior a la capacitat definida, el vèrtex falla i transmet. Aquest procés es repeteix fins que s'assoleix un estat estacionari en el qual ja no hi ha més vèrtexs que fallin.

Per iniciar les fallades, es pren un conjunt de vèrtexs inicials que són forçats a fallar. En les anàlisis realitzades, els conjunts de vèrtexs inicials corresponen als gens/proteïnes diferencialment expressats.

Per definir les capacitats dels vèrtexs, s'han integrat les dades d'expressió amb les dades d'interaccions. D'aquesta manera, s'ha pogut definir la capacitat de cada vèrtex reflectint l'activitat del gen associat.

Sia  $G = (V, E)$  la xarxa de l'interactoma amb  $V$  el conjunt dels seus vèrtexs (gens/proteïnes) i  $E$  el de les arestes (interaccions entre proteïnes). Per cada vèrtex  $i \in V$  de la xarxa es defineix la càrrega  $L_i$  del vèrtex  $i$  com

$$L_i = \left( k_i \sum_{j \in \Gamma_i} k_j \right)^\alpha \quad (6.1)$$

on  $\Gamma_i$  és el conjunt dels veïns del vèrtex  $i$ ,  $k_j$  el grau del vèrtex  $j$  (nombre d'arestes incidents) i  $\alpha$  és un paràmetre real que modela la definició de  $L_i$ .

La capacitat de càrrega màxima del vèrtex  $i$ ,  $C_i$ , es defineix a partir de la càrrega però també dels nivells d'expressió de la proteïna que representa. Sia  $S$  el conjunt de les sondes associades als vèrtexs  $V$  en un experiment i  $S_i \subset S$  el conjunt de les sondes que apunten al gen associat al vèrtex  $i$ . Per a cada sonda  $s$ , es tenen  $m$  nivells d'expressió corresponent a les  $m$  mostres  $s_0, \dots, s_{m-1}$ . Considerem aquests nivells reordenats de menor a major  $s_{(0)}, \dots, s_{(m-1)}$ , a on  $(j)$  indica que en la posició  $j$  hi ha el  $j$ -èsim valor més petit. Es defineix el rang normalitzat dels nivells de la proteïna  $i$  com

$$r_i = \frac{\max_{s \in S_i} \{s_{(m-1)} - s_{(0)}\}}{\max_{s \in S} \{s_{(m-1)} - s_{(0)}\}}. \quad (6.2)$$

La capacitat del vèrtex  $i$  es defineix com la càrrega del vèrtex més una un increment que depèn de del rang normalitzat

$$C_i = L_i + \tau r_i L_i = (1 + \tau r_i) L_i \quad (6.3)$$

d'aquesta manera, la capacitat reflecteix la variabilitat de l'expressió dels gens, en el sentit que aquells gens que tenen un rang d'expressió major tenen una capacitat major en la xarxa.

L'increment que reben els veïns d'un vèrtex  $i$  quan falla és

$$\Delta L_{ji} = L_i \frac{L_j}{\sum_{\ell \in \Gamma_i} L_\ell} \quad (6.4)$$

que es pot entendre que la càrrega  $L_i$  es distribueix respecte de la càrrega dels seus veïns, reben una proporció més gran aquells veïns amb una càrrega més gran. Com que

$$\sum_{j \in \Gamma_i} \Delta L_{ji} = \sum_{j \in \Gamma_i} L_i \frac{L_j}{\sum_{\ell \in \Gamma_i} L_\ell} = L_i \frac{\sum_{j \in \Gamma_i} L_j}{\sum_{\ell \in \Gamma_i} L_\ell} = L_i,$$

la càrrega distribuïda coincideix amb la càrrega del vèrtex  $i$ . Llavors, la nova càrrega del vèrtex  $j$  és

$$L_j + \Delta L_{ji} \tag{6.5}$$

que per simplificar la notació se l'anomena  $L_j$ . En aquest punt, si  $L_j > C_j$  el vèrtex  $j$  falla i transmet la seva càrrega als seus veïns actius (que no han fallat), i així successivament produint una cascada de fallades fins que no es produeixen més fallades.

Donat un vèrtex inicial  $i_0 \in V$  i fixats els paràmetre  $\alpha$  i  $\tau$ , les fallades es computen segons els passos

1. es defineix  $M = \{i_0\}$
2. per a cada vèrtex  $i$  de  $M$ , es redefineixen les càrregues dels seus veïns segons l'expressió (6.5).
3. s'eliminen els vèrtexs del conjunt  $M$  i les arestes incidents.
4. es redefineix el conjunt  $M$  com el conjunt dels vèrtexs que tenen la càrrega major que la capacitat.
5. es repeteixen els passos 2, 3 i 4 fins que no es produeixen més fallades.
6. es defineix  $cf(i_0; \alpha, \tau)$  com el nombre de vèrtexs que han fallat segons el vèrtexs inicial  $i_0$ .

Per implementar aquest algorisme, s'ha escrit un programa amb el llenguatge Python.

# 7

## Materials i mètodes II

### 7.1 La xarxa de l'interactoma

Per representar la xarxa de l'interactoma s'ha usat la versió núm 9 de la base de dades *Human Protein Reference Database* que conté 9 670 proteïnes i 39 172 interaccions (Keshava Prasad et al., 2009) i que està construïda mitjançant la recopilació en la literatura científica. Aquestes dades s'han filtrat prenent només les proteïnes anotades per Entrez (Maglott et al., 2005), prenent la component principal de la xarxa d'aquestes dades i eliminant els llaços. La xarxa final conté 9 267 proteïnes i 36 893 interaccions.

### 7.2 Dades d'expressió, de les $IC_{50}$ i de mutació de gens

Les dades d'expressió normalitzades, de les  $IC_{50}$  i de mutació de gens s'han obtingut del seu estudi original (Garnett et al., 2012).

Les dades d'expressió corresponen a l'expressió basal gènica de 595 línies cel·lulars de diversos càncers. Per assignar un valor d'expressió a cada gen representat en el bioxip, s'ha considerat la sonda amb més variabilitat, és a dir amb la desviació estàndard més gran. Pels gens en l'interactoma que no estan representats en el bioxip (1 428), se'ls hi ha assignat el valor d'expressió esperat del bioxip a través de les sondes.

Les dades de concentració que inhibeixen l'activitat cel·lular al 50% ( $IC_{50}$ ) corresponen a 130 fàrmacs diferents. Cada fàrmac s'ha classificat com a *únic node* o *múltiple node* segons si afecta a una única diana o a diverses dianes, Taula D.0.1.

### 7.3 Anàlisi dels processos biològics

Les dades dels processos biològics s'han obtingut del repositori *Gene Ontology Consortium* (Ashburner et al., 2000). S'ha obtingut i muntat la base de dades del 05/2012 en MySQL. Mitjançant la base de dades es poden assignar els processos biològics als gens. Per fer una anàlisi d'enriquiment en un conjunt de gens donat, només s'han considerat aquells termes amb una freqüència major al 5% en el conjunt a avaluar. Per determinar l'enriquiment d'un terme en un conjunt de gens, s'ha efectuat un test de Fisher amb un nivell de significança del 5%.

### 7.4 Assaigs cel·lulars

Totes les línies cel·lulars usades en aquest estudi, han estat cultivades seguint protocols estàndards. La viabilitat cel·lular s'ha avaluat mitjançant assajos de methylthiazol tetrazolium (MTT, Sigam-Aldrich), que mesura la quantitat de cèl·lules vives. La línia cel·lular epitelial de mama humana MCF10A (ATCC) s'ha usat com a control no tumoral. Els fàrmacs s'han adquirit de Sigam-Aldrich (metformina) i de LC laboratories. Els resultats corresponen, com a mínim, a quatre assajos independents per a cada combinació de fàrmacs, amb quatre rèpliques per a cada dosi de fàrmacs o combinació de fàrmacs.

### 7.5 Activitat de la xarxa

L'activitat de la xarxa s'ha definit a partir de la comunicabilitat amb pesos (Crofts i Higham, 2009). Primer, s'han integrat les dades d'expressió a la xarxa per calcular els pesos de les arestes. Donades les dades d'expressió d'una línia cel·lular, per a cada aresta  $(i, j)$  amb  $i, j$  vèrtexs de la xarxa, el pes de l'aresta es defineix com

$$w_{ij} = \frac{x_j}{\sum_{k \in \Gamma_i} x_k} x_i + \frac{x_i}{\sum_{k \in \Gamma_j} x_k} x_j$$

on  $\Gamma_i$  és el conjunt dels veïnes del vèrtex  $i$  i  $x_i$  el valor d'expressió del gen corresponent. Tot seguit, s'ha definit la matriu de la xarxa  $W$  a partir dels pesos de les arestes, essent  $w_{ij}$  l'element en la posició  $(i, j)$  de la matriu. S'observa que  $w_{ji} = w_{ij}$ , és a dir, la matriu  $W$  és simètrica. Per reduir la variabilitat dels elements de la matriu  $W$  i siguin comparables entre ells, la matriu  $W$  s'ha normalitzat (Higham et al.,



2007) mitjançant

$$\frac{w_{ij}}{\sqrt{d_i d_j}}$$

on  $d_i = \sum_j w_{ij}$ .

A partir d'aquesta matriu, s'ha definit la mesura de centralitat d'un vèrtex  $i$  com

$$C(i) = (e^W)_{ii} \quad (7.1)$$

és a dir, l'entrada  $i$ -èsima de l'exponencial de la matriu de pesos. Per la definició d'exponencial, es pot escriure

$$e^W = \sum_{k \geq 0} \frac{W^k}{k!} \quad (7.2)$$

on  $W^k$  correspon a la potència  $k$ -èsima de la matriu  $W$ . Com que la matriu  $W$  és simètrica, per calcular l'exponencial de la matriu, s'ha diagonalitzat la matriu de pesos

$$W = PDP^{-1}$$

on  $D$  correspon a la matriu diagonal dels valors propis de  $W$  i  $P$  la matriu dels vectors propis. Llavors, es té que

$$e^W = e^{PDP^{-1}} = Pe^D P^{-1}$$

i l'exponencial d'una matriu diagonal correspon a la matriu de l'exponencial dels seus elements, és a dir

$$e^D = \begin{pmatrix} e^{d_1} & & \\ & \ddots & \\ & & e^{d_n} \end{pmatrix}$$

Finalment, s'ha definit la mesura d'activitat de la xarxa a partir de l'Estrada Índex (Estrada, 2000) respecte de la pesos  $W$ ,

$$NA(G) = \sum_{i \in V} C(i) \quad (7.3)$$

que dóna una mesura global de la xarxa.

## 7.6 Estimació dels tractaments

### 7.6.1 Estimació de la resposta cel·lular a tractaments de fàrmacs

Per estimar la resposta cel·lular respecte de les concentracions de fàrmacs, s'ha usat l'equació

$$\frac{f_a}{f_u} = \left( \frac{D}{D_m} \right)^m \quad (7.4)$$

a on  $f_a$  és la fracció de cèl·lules afectades i  $f_u$  les no afectades,  $D$  és la dosi del fàrmac i  $D_m$  la dosi necessària per reduir l'activitat cel·lular a la meitat i que es coneix amb el nom de *median-effect equation* (Chou, 2006). Prenent logaritmes,

$$\log \left( \frac{f_a}{f_u} \right) = m \log(D) - m \log(D_m) \quad (7.5)$$

que s'ha ajustat mitjançant una regressió lineal programada en Python.

Per comprovar els resultats usant la regressió lineal de l'Equació 7.5, s'ha ajustat l'Equació 7.4 mitjançant la implementació en Python del mètode Newton-Raphson (Stoer et al., 2002), deixant com a paràmetres lliures  $D_m$  i  $m$ .

### 7.6.2 Estimació de l'efecte de la combinació de fàrmacs

Per estimar l'efecte de la combinació de dos fàrmacs, respecte als efectes produïts pels fàrmacs individualment, s'ha aplicat l'índex de combinació

$$CI = \frac{D_1}{D_m} + \frac{D_2}{(D_m)_2} \quad (7.6)$$

a on  $D_i$  correspon a la dosi del fàrmac  $i$  i  $(D_m)_i$  la concentració necessària del fàrmac  $i$  per reduir l'activitat cel·lular a la meitat. En general, si  $CI < 1$  indica sinergia, si  $CI > 1$  antagonisme i si  $CI = 1$  additivitat dels fàrmacs. Tanmateix, s'ha considerat (Chou, 2006) (i) sinergia si  $CI < 0.9$ , (ii) additivitat si  $CI \in (0.9, 1.1]$ , (iii) antagonisme si  $CI > 1.1$ .

## **Part IV**

# **Resultats**



# 8

## Resultats I

### 8.1 Anàlisi dels gens diferencialment expressats

#### 8.1.1 Gens diferencialment expressats en el càncer de mama

Per identificar els gens diferencialment expressats entre estadis del càncer, s'ha usat el mètode SAM (Tusher et al., 2001), que compara els nivells del producte gènic entre condicions diferents —en aquest cas, nivells en les cèl·lules del teixit corresponent a estadis diferents del càncer— i realitza un test de permutacions per determinar la significança dels canvis d'aquests nivells.

S'han comparat els nivells de producte gènic en les cèl·lules entre mostres de teixit normal (N) i carcinoma ductal *in situ* (CDIS), teixit normal i carcinoma ductal invasiu (CDI), teixit de carcinoma ductal *in situ* i de carcinoma ductal invasiu, i, finalment, entre teixit de carcinoma ductal invasiu i de metàstasi (M).

Per a cada comparació dels nivells de producte gènic en condicions del càncer diferents, s'ha obtingut el conjunt de gens diferencialment expressats (amb un FDR 5%). Aquest conjunt s'ha dividit en dos conjunts, un definit pels gens sobreexpressats i un altre definit pels gens infraexpressats. Tot seguit s'han identificat les proteïnes corresponents dins de la xarxa de l'interactoma, Taula 8.1.1. Finalment, s'ha usat cadascun d'aquests conjunts de proteïnes obtinguts per definir el conjunt inicials de vèrtexs en la xarxa per l'anàlisi de la cascada de fallades.

**Taula 8.1.1:** Taula dels gens diferencialment expressats en el càncer de mama i els corresponent nombres de proteïnes dins de la xarxa de l'interactoma (HPRD). Teixit de mama normal (N), hiperplàsia ductal atípica (HDA), carcinoma ductal *in situ* (CDIS), carcinoma ductal invasiu (CDI) i metàstasi (M).

Infraexpressats					Sobreexpressats				
	HDA	CDIS	CDI	M		HDA	CDIS	CDI	M
N	728 (496)	1 748 (939)	2 718 (1 526)		N	850 (566)	2 917 (1 691)	1 085 (644)	
CDIS			24 (14)		CDIS			268 (170)	
CDI				22 (11)	CDI				128 (79)
Diferencialment expressats									
	HDA	CDIS	CDI	M		HDA	CDIS	CDI	M
N	1 578 (1 062)	4 665 (2 630)	3 803 (2 170)						
CDIS			292 (184)						
CDI				150 (90)					

### 8.1.2 Gens diferencialment expressats en el càncer de colon

Per detectar els gens que capturen diferències entre estadis del càncer de colon, s'ha usat el mateix mètode que en el càncer de mama, l'algoritme SAM (Tusher et al., 2001). En aquest cas, s'han comparat els nivells de producte gènic en les cèl·lules de teixit normal (N) i teixit d'adenomes (A) —les cèl·lules mostren una proliferació atípica però es tracta d'una alteració benigne, sovint considerada un estadi previ per desenvolupar carcinomes—, i entre teixit d'adenomes i teixit de carcinomes (C) —les cèl·lules han adquirit les característiques malignes. S'han identificat aquells gens amb un canvi de nivell de producte gènic de més del doble, en un sentit o en l'altre, Taula 8.1.2. Els conjunts dins de la xarxa de l'interactoma definits per aquests gens s'han usat com a conjunts inicials en el mètode de la cascada de fallades.

**Taula 8.1.2:** Taula dels gens diferencialment expressats en el càncer de colon i els corresponent nombres de proteïnes dins de la xarxa de l'interactoma (HPRD). Teixit normal (N), teixit d'adenomes (A) i teixit de carcinomes (C).

	Total	infraexpressats	sobreexpressats
N→A	822 (460)	463 (236)	359 (224)
A→C	303 (195)	44 (19)	259 (176)

## 8.2 La xarxa de l'interactoma

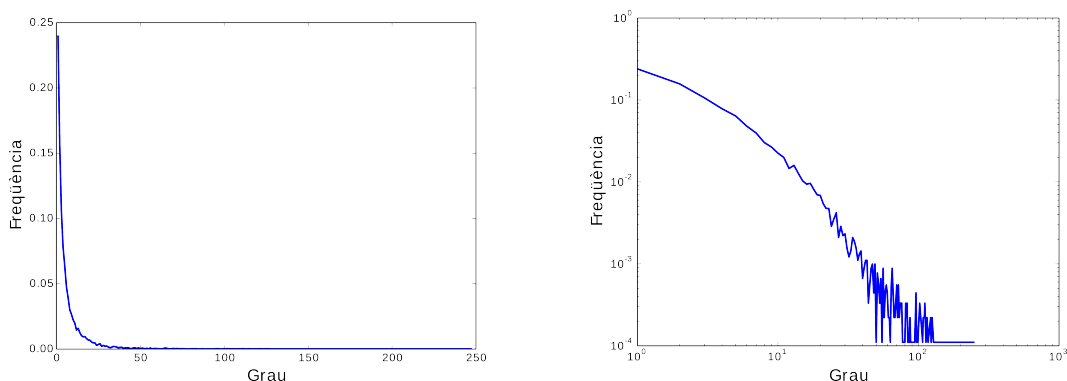
Per representar la xarxa de l'interactoma s'han obtingut les dades d'interacció entre proteïnes humanes de dues fonts diferents, *Human Protein Reference Database* (HPRD) (Keshava Prasad et al., 2009) i *IntAct Molecular Interaction Database* (IntAct, Figura A.0.1) (Orchard et al., 2014) per construir dues representacions de la xarxa de l'interactoma proteic humà. Aquestes bases de dades estan formades per dades d'interacció proteiques recollides de la literatura científica, assegurant un experiment científic com a evidència de les interaccions.

El nombre de proteïnes que conté HPRD és 9 461 i un total de 37 081 interaccions entre elles. Aquestes dades s'han filtrat, excloent aquelles proteïnes no anotades per Entrez (Maglott et al., 2005). Amb les dades filtrades s'ha construït un xarxa formada per 273 components connexes. Malgrat aquest nombre elevat de components connexes, la gran quantitat d'informació de l'interactoma està continguda en la component principal, Taula 8.2.1.

**Taula 8.2.1:** Les tres components connexes més grans. La component connexa principal  $G_0$  és la que conté la majoria d'informació de l'interactoma.

component	núm. proteïnes	núm. interaccions
<b>1ra.</b>	9 045	36 753
<b>2na.</b>	4	4
<b>3ra.</b>	4	3

La xarxa de l'interactoma es defineix prenent la component principal i eliminant els homodímers, és a dir el llaços de la xarxa. Al final s'ha obtingut una xarxa amb 9 045 vèrtexs i 34 853 arestes i el seu diàmetre és 14. Aquesta xarxa té un caràcter proper a les xarxes d'escala lliure, ja que la cua de la distribució del grau té un caiguda lineal en escala logarítmica, Figura 8.2.1.



**Figura 8.2.1:** A la dreta, distribució del nombre d'interaccions de les proteïnes de l'interactoma humà. A l'esquerra la distribució en escala logarítmica. Tal i com es pot apreciar, llevat dels primers graus, la distribució en escala logarítmica té una caiguda lineal (pendent  $-1.89$ ), suggerint una topologia d'escala lliure.

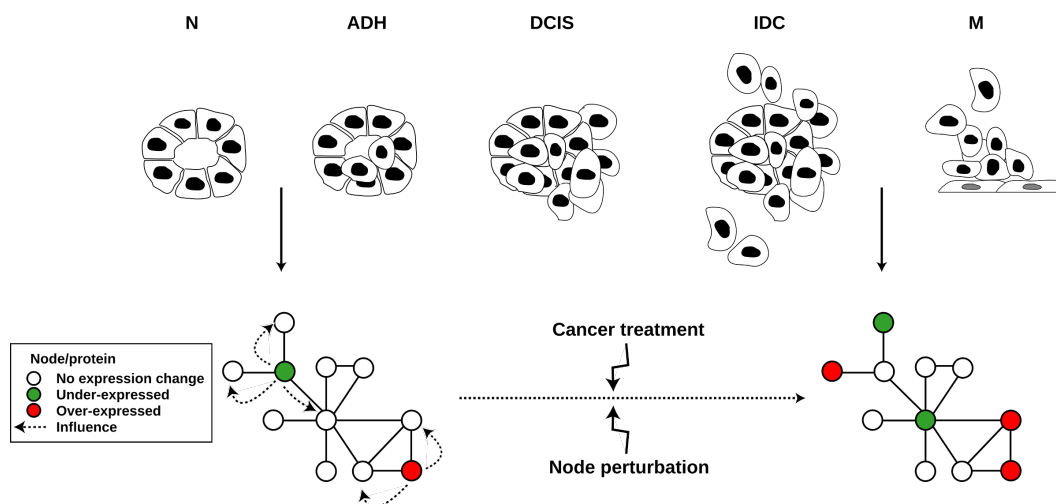
## 8.3 Anàlisi topològica de la xarxa de l'interactoma

### 8.3.1 Robustesa topològica respecte de les proteïnes del càncer

Els càncers es desenvolupen i progressen a través d'una evolució de diversos passos en els quals les cèl·lules adquireixen mutacions genètiques. Aquestes mutacions genètiques es tradueixen en alteracions en els productes gènics que comporten alteracions de funcions i processos que li confereixen avantatges evolutius a la cèl·lula que esdevé cancerosa. Des d'aquest punt de vista, sembla lògic pensar que el desenvolupament o progressió del càncer pot reflectir-se en la topologia de la xarxa de l'interactoma.

Per analitzar la relació entre la dinàmica del càncer i la topologia de la xarxa de l'interactoma, s'ha implementat un mètode basat en cascades de fallades (Wang i Rong, 2009) i s'ha examinat l'impacte que tenen, en la xarxa de l'interactoma, els canvis de nivells proteics que es produeixen en la cèl·lula en la formació d'un càncer (Figura 8.3.1). Wang i Rong han implementat aquest mètode per estudiar la robustesa de la xarxa elèctrica dels Estats Units mitjançant la transmissió de càrrega entre les centraletes, que fallen quan no poden suportar l'increment de càrrega. Un mètode semblant s'ha implementat en la xarxa metabòlica cel·lular (Smart et al., 2008). Aquí, s'ha implementat el mètode usat per Wang i Rong per fer una anàlisi de la xarxa de l'interactoma respecte dels nivells de les proteïnes associades al desenvolupament i progressió del càncer.





**Figura 8.3.1:** Estratègia usada per l'estudi de la topologia de la xarxa de l'interactoma respecte de la dinàmica de la formació del càncer. Per a cada estadi del càncer, s'han identificat aquelles proteïnes que pateixen canvis significatius en el seus nivells cel·lulars. Els conjunts definits per aquestes proteïnes s'han usat com a conjunts inicials per generar les cascades de fallades dins de la xarxa de l'interactoma i se n'han estudiat els patrons topològics generats.

Per analitzar la topologia de la xarxa de l'interactoma respecte dels nivells de les proteïnes, primer s'han integrat les dades d'interacció proteica amb les dades de l'expressió gènica en els diferents estadis del càncer de mama. Una vegada construïda la xarxa de l'interactoma, s'ha definit la càrrega  $L_i$  de cada vèrtex (Equació 6.1) i s'ha calculat el paràmetre  $r_i$  corresponent (el rang d'expressió de la proteïna associada, Equació 6.2), per poder definir la capacitat  $C_i$  de cada vèrtex (Equació 6.3).

L'interactoma proteic humà s'ha representat mitjançant dues xarxes de dues bases de dades diferents, HPRD (Keshava Prasad et al., 2009) i IntAct (Orchard et al., 2014) (Secció 6.1). Les anàlisis s'han realitzat en les dues xarxes, principalment en la HPRD i a mode de comprovació en la IntAct, Figura B.0.1.

Tal i com s'ha introduït anteriorment, la càrrega dels vèrtexs està definida pel grau del vèrtex i el dels seus veïns però depèn, també, del paràmetre  $\alpha$  que modela l'expressió de  $L_i$ . Per definir  $L_i$ , primer s'ha analitzat el paràmetre  $\alpha$  per determinar-ne els seus valors.

La càrrega  $C_i$  depèn del paràmetre  $\tau$ . En aquest cas però, es deixa lliure i es realitza l'anàlisi per diversos valors de  $\tau$  en un malla equidistant. Els resultats es representen en funció d'aquest paràmetre, per a més detalls veure la Figura 8.3.5.

### Anàlisi del paràmetre $\alpha$

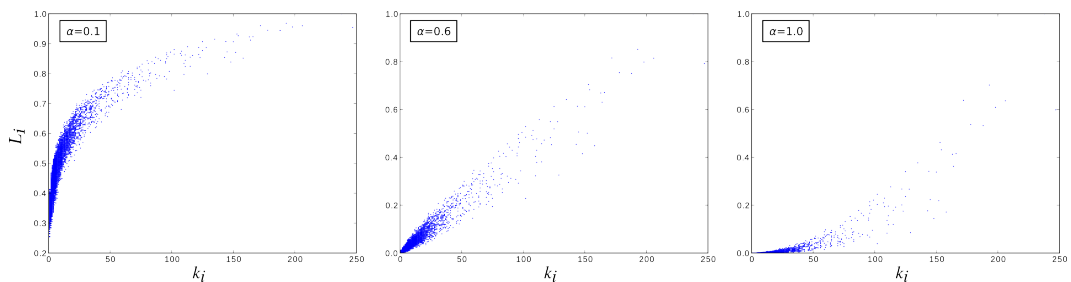
La càrrega d'un vèrtex  $i$  de la xarxa s'ha definit segons l'Equació 6.1,

$$L_i = \left( k_i \sum_{j \in \Gamma_i} k_j \right)^\alpha,$$

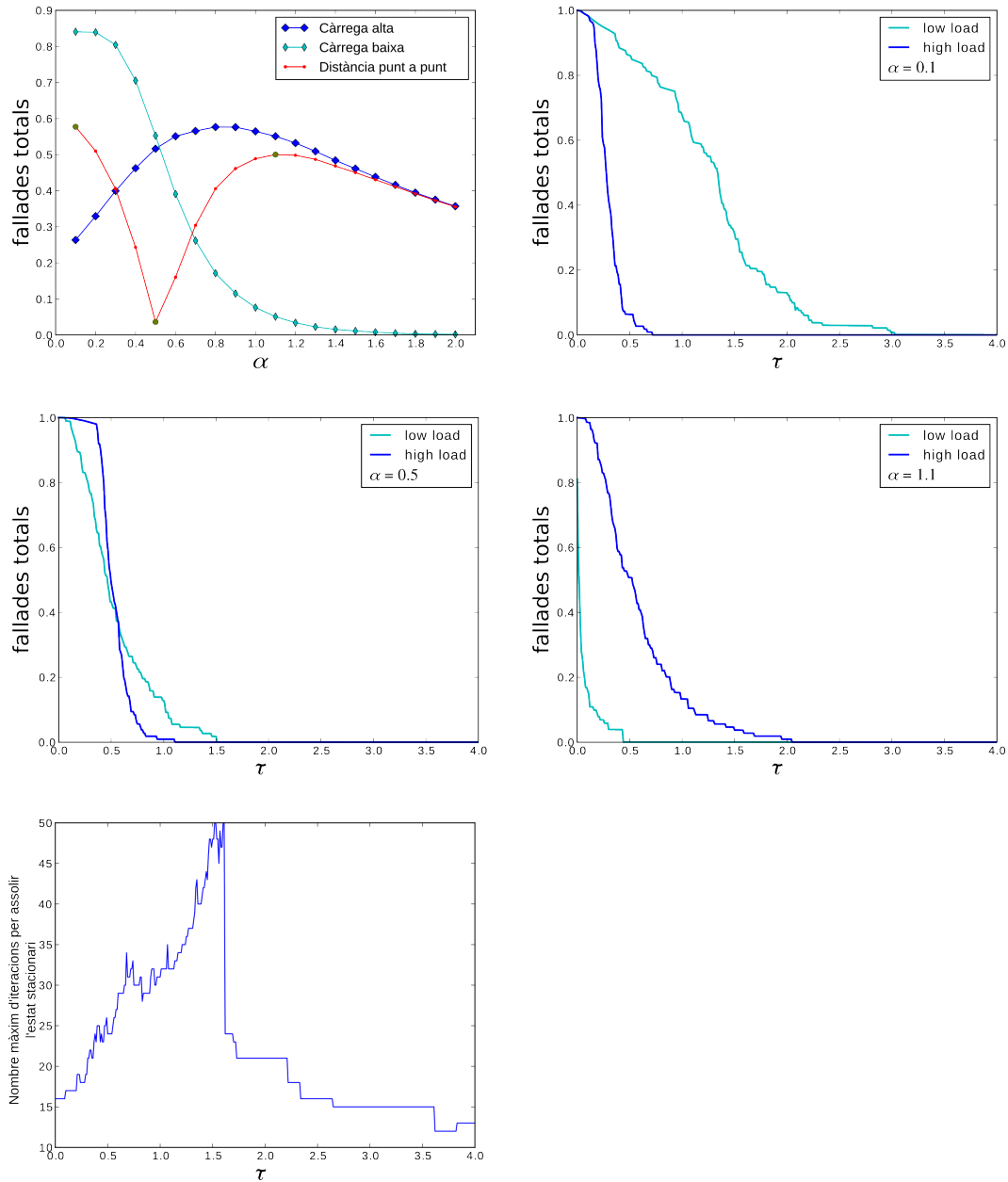
és a dir, a partir dels graus del vèrtexs i dels graus dels seus veïns. D'aquesta manera, aquelles proteïnes que tenen més interaccions i, que alhora, els seus interactors tenen més interaccions, tenen una càrrega major. Al mateix temps, l'expressió de  $L_i$  depèn del paràmetre  $\alpha$  que modela el seu comportament que, bàsicament, es pot resumir en (Figura 8.3.2):

- i.  $\alpha$  petita: una atenuació del caràcter de la càrrega per graus dels vèrtexs grans,
- ii.  $\alpha$  mitjana: un creixement lineal respecte dels graus dels vèrtexs
- iii.  $\alpha$  gran: una atenuació del caràcter de la càrrega per graus dels vèrtexs petits.

Per determinar quan  $\alpha$  és petita, mitjana o gran, s'ha realitzat una anàlisi de les fallades per diferents valors de  $\alpha$  respecte de dos tipus de conjunts inicials: un definit pels vèrtexs amb les 100 càrregues més baixes i un altre per les 100 més altes, amb la intenció de capturar aquells comportaments antagònics, Figura 8.3.3.



**Figura 8.3.2:** Representació de les càrregues  $L_i$  respecte dels graus  $k_i$  dels vèrtexs per tres valors del paràmetre  $\alpha$  (dades d'expressió dels estadis N→CDI). D'esquerra a dreta, per  $\alpha = 0.1$  s'atenuen les càrregues altes; per  $\alpha = 0.6$  un comportament lineal i per  $\alpha = 1.0$  s'atenuen les càrregues baixes.



**Figura 8.3.3:** D'esquerra a dreta i de dalt a baix. La primera figura mostra les fallades totals (projectades a través de les  $\tau$ ) pels dos conjunts definits pels 100 vèrtexs amb càrrega més alta (blau fosc) i pels 100 amb càrrega més baixa (blau clar) per a cada  $\alpha$  (en el cas  $N \rightarrow \text{CDI}$ ). En vermell es mostren les diferències dels nombres de fallades entre els dos conjunts. Els punts verds sobre de la corba vermella corresponen als punts que o bé maximitzen aquesta diferència, pel cas  $\alpha$  petita i gran, o la minimitza,  $\alpha$  mitjana. El segon, tercer i quart gràfic mostren el nombre de fallades totals en funció dels valors de  $\tau$  pels tres valors de  $\alpha$  escollits (per a més detalls del paràmetre  $\tau$  veure al Figura 8.3.5). A mesura que  $\alpha$  creix, s'observa que les diferències del nombre de fallades entre els dos conjunts canvia, començant per una diferència gran i que domina el conjunt de càrregues baixes, passant per un comportament semblant pels dos conjunts i, finalment, la diferència torna a ser gran però dominant el conjunt de les càrregues altes. En la última figura, es representen el nombre d'iteracions per assolir l'estat estacionari en el qual o bé tots els vèrtexs de la xarxa han fallat o bé ja no es provoquen més fallades.

Per a cada valor de  $\alpha$ , s'han calculat els nombres de fallades produïts pels dos conjunts i s'han comparat entre ells, Figura 8.3.3. A partir d'aquesta comparació, s'han considerat aquells valors de  $\alpha$  tals que (Taula 8.3.1)

- i. una  $\alpha$  petita que maximitzi la diferència de fallides, emfatitzant el rol dels vèrtexs amb càrrega baixa,
- ii. una  $\alpha$  mitjana que minimitzi la diferència de fallides i
- iii. una  $\alpha$  gran que maximitzi la diferència de fallides, inhibint el rol dels vèrtexs de càrrega baixa

Cadascuna de les anàlisis, s'han realitzat per cadascun d'aquests tres valors de  $\alpha$ . Tanmateix, els tres casos mostren resultats semblant i es mostraran els corresponents a la  $\alpha$  mitjana.

**Taula 8.3.1:** Valors de  $\alpha$  determinats per a cadascun dels estadis

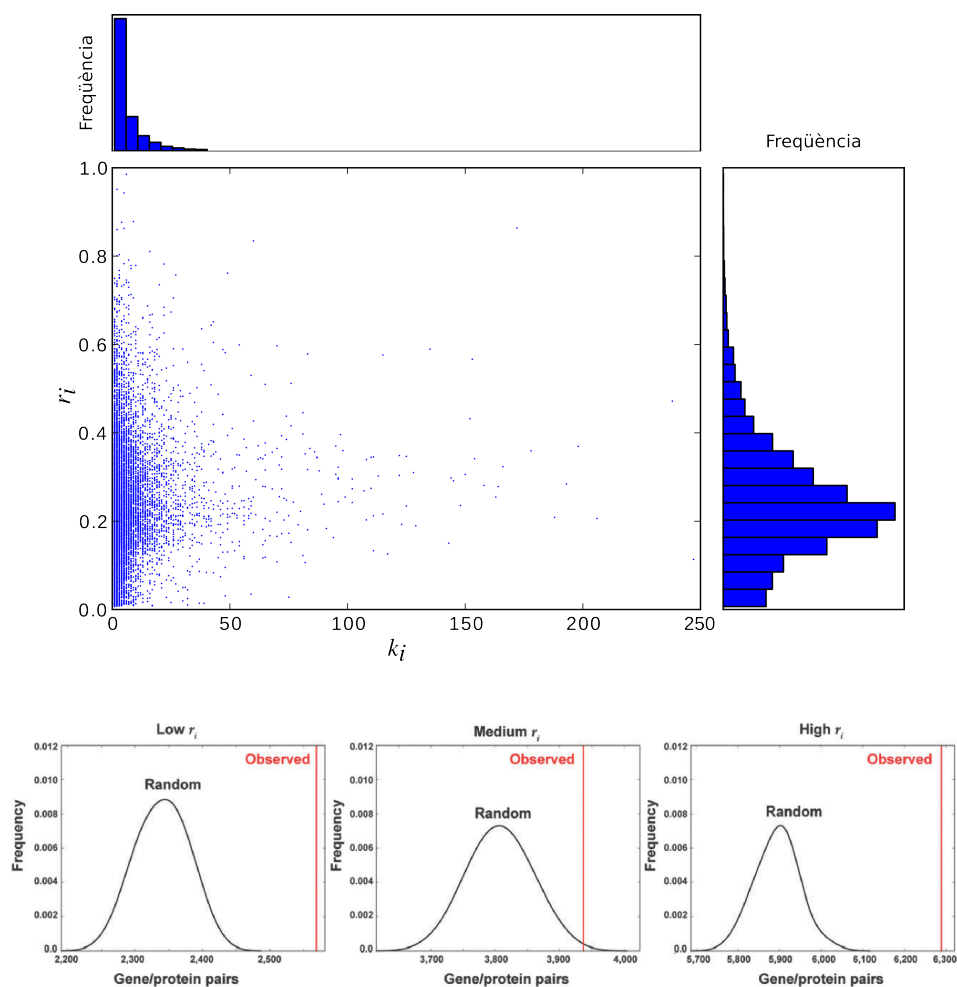
	<b>N→HDA</b>	<b>N→CDIS</b>	<b>N→CDI</b>	<b>CDIS→CDI</b>	<b>CDI→M</b>
$\alpha_{lo}$	0.1	0.1	0.1	0.1	0.1
$\alpha_{med}$	0.4	0.5	0.4	0.4	0.5
$\alpha_{hi}$	1.0	1.1	1.1	1.0	1.1

### Anàlisi del paràmetre $r_i$

La capacitat  $C_i$  de cada vèrtex  $i$  de la xarxa defineix la càrrega màxima que pot suportar cada vèrtex. Quan aquesta és superada, el vèrtex falla i transmet la seva càrrega als vèrtexs veïns. La capacitat es defineix segons l'Equació 6.3,

$$C_i = L_i + \tau r_i L_i,$$

que es pot entendre com la càrrega inicial del vèrtexs més una proporció d'aquesta càrrega, determinada per  $\tau r_i$ . El paràmetre  $r_i$  reflecteix l'expressió dels gens a través



**Figura 8.3.4:** A dalt, gràfica dels rangs  $r_i$  en funció dels graus  $k_i$  dels vèrtexs  $i \in V$ , pel cas  $N \rightarrow \text{CDI}$  —pels altres casos els resultats són similars. Per quantificar la dependència entre  $r_i$  i  $k_i$  s'ha realitzat un test no paramètric  $\tau$  de Kendall mostrant una lleugera correlació positiva,  $\tau \in (0.03, 0.09)$  per a tots els conjunts de dades dels estadis del càncer.

L'histograma a la part superior de la imatge correspon a la distribució de grau, introduït anteriorment.

L'histograma de la dreta correspon a la distribució dels rangs  $r_i$ .

A baix, distribucions del nombre de proteïnes que interaccionen entre elles tals que els seus gens codificants tenen un rang  $r_i$  similar i, en vermell, el nombre observat, pel cas  $N \rightarrow \text{CDI}$  —resultats similars per la resta de casos. Per estudiar aquesta relació entre el rang d'expressió i proteïnes que interaccionen entre elles, s'ha dividit el domini del rang  $r_i$  en terçils, de menor a major en les figura. Per a cada tercil s'ha generat una distribució de nombres d'interaccions de proteïnes tals que els gens corresponents tenen el rang en el mateix tercil. La generació d'aquesta distribució s'ha realitzat permutant aleatòriament els noms dels vèrtexs i comptant-ne les interaccions proteïnes amb rangs en els mateix tercil, resultant-ne les distribucions dibuixades. Tot seguit s'ha calculat la significança del nombre observat, tots tres casos significatius.

de les diferents condicions de l'experiment —els diferents estadis del càncer— i es defineix com el rang màxim dels nivells d'expressió de la proteïna corresponent al vèrtex  $i$ , Equació 6.2. És a dir, com la diferència dels valors màxim i mínim d'expressió del gen corresponent. D'aquesta manera, s'introdueix l'activitat gènica en el model i quant major sigui aquest rang  $r_i$  també ho serà la capacitat  $C_i$ . Des d'un punt de vista biològic, s'assumeix que la capacitat d'una proteïna d'adaptar-se i absorbir la informació transmesa pels seus interactors és major quan major és el rang de l'expressió del gen que la codifica —assumint, també, que l'expressió gènica i els nivells de proteïna en la cèl·lula estan correlacionats positivament (Maier et al., 2009). Aquesta assumpció és introduïda, precisament, pel paràmetre  $r_i$  que quan major és el rang de l'expressió d'un gen major és la capacitat del vèrtex corresponent a la proteïna codificada pel gen i més càrrega podrà absorbir dels vèrtexs veïns.

Per no restringir l'anàlisi a un valor concret dels paràmetres  $r_i$ , s'ha escalat mitjançant el paràmetre  $\tau$ . D'aquesta manera s'han pogut realitzar les anàlisis de la robustesa de la xarxa per diversos valors de la càrrega  $C_i$ , respectant, però, la proporció entre els  $r_i$ , ja que el paràmetre  $\tau$  és el mateix per a tots els vèrtexs  $i$  una vegada ha estat fixat.

Per veure si el rang d'expressió depèn del nombre d'interactors de la proteïna, s'ha calculat la correlació entre els rangs  $r_i$  i els graus  $k_i$  dels vèrtexs de la xarxa (coeficient  $\tau$  de Kendall). Els resultats mostren una lleugera correlació positiva, suggerint una lleugera dependència entre el rang d'expressió d'un gen i el nombre d'interactors que té la proteïna que codifica, Figura 8.3.4.

Per concloure aquesta petita anàlisi, i per veure si els rangs dels nivells de proteïnes que interaccionen entre elles tenen el rang d'expressió similar, s'ha generat una distribució nul·la d'interaccions entre proteïnes amb  $r_i$  similar i s'ha calculat la probabilitat del valor observat en aquesta distribució. Per definir rang similar, s'ha dividit el domini del rang en tercils. Per a cada tercil, s'han computat les parelles de proteïnes que interaccionen entre elles i que tenen el rang  $r_i$  en el mateix tercil. Per construir la distribució nul·la, s'han permutat aleatòriament els noms dels vèrtexs dins de la xarxa —d'aquesta manera es preserva la topologia de la xarxa—, comptant en cada permutació el nombre de vèrtexs adjacents amb  $r_i$  en el mateix tercil, Figura 8.3.4. Si bé les els rangs  $r_i$  i els nombres d'interactors  $k_i$  no estan fortament correlacionats, sí que s'observa que aquelles proteïnes que interaccionen entre elles tenen, significativament, el rang d'expressió similar ( $p < 0.001$ ), tal i com s'aprecia en la Figura 8.3.4.

### El model de la cascada de fallades

Una vegada definits els paràmetres, es poden definir la càrrega (Equació 6.1), capacitat (Equació 6.3) i increment (Equació 6.4) dels vèrtexs,

$$\begin{aligned} L_i &= \left( k_i \sum_{j \in \Gamma_i} k_j \right)^\alpha \\ C_i &= (1 + \tau r_i) L_i \\ \Delta L_{ji} &= L_i \frac{L_j}{\sum_{\ell \in \Gamma_i} L_\ell} \end{aligned}$$

on  $\alpha \in \{\alpha_{low}, \alpha_{med}, \alpha_{high}\}$  pren tres valors corresponents als comportaments de la càrrega  $L_i$  respecte dels graus  $k_i$  (Secció 8.3.1, Figura 8.3.2) i que per a cadascun d'aquests valors s'han calculat les fallades en la xarxa. En cada cas, el paràmetre  $\tau$  pren valors en una malla equiespaiada prou fina per captar els diferents comportaments de les cascades de fallades en la xarxa, Figura 8.3.5.

Sigui  $V$  el conjunt de vèrtex de la xarxa,  $B \subset V$  un subconjunt de vèrtexs i  $v \in V \setminus B$  un vèrtex qualsevol. El nombre de fallades generades pel vèrtex  $v$  en el conjunt  $B$  s'ha anotat com

$$\widetilde{cf}_B(v; \alpha, \tau). \quad (8.1)$$

Com que el nombre màxim de fallades que pot causar el vèrtex  $v$  en el conjunt  $B$  és el nombre d'elements de  $B$ , es té que  $\widetilde{cf}_B(v; \alpha, \tau) \in [0, |B|]$ , on  $|B|$  és la mida del conjunt  $B$ . S'ha definit el nombre de fallades normalitzat com

$$cf_B(v; \alpha, \tau) = \frac{\widetilde{cf}_B(v; \alpha, \tau)}{|B|} \quad (8.2)$$

que està en l'interval  $[0, 1]$ .

Sigui  $A \subset V$  un altre conjunt de vèrtexs. S'ha definit el nombre de fallades relatives al conjunt  $A$  en el conjunt  $B$  com la suma de les fallades generades pels elements de  $A$  en  $B$ ,

$$cf_B(A; \alpha, \tau) = \frac{\sum_{v \in A} cf_B(v; \alpha, \tau)}{|A|} \quad (8.3)$$

Com que per a cada vèrtex  $v$ ,  $cf_B(v) \leq 1$ , es té que el nombre total de fallades en  $B$  respecte de  $A$  serà com a màxim  $|A|$ . Així, l'expressió anterior, Equació 8.3,  $cf_B(A) \in [0, 1]$ . En el cas que  $B = V$ , el nombre de fallides relatives al conjunt  $A$

és el nombre total de fallides en la xarxa

$$cf(A; \alpha, \tau) = \frac{\sum_{v \in A} \widetilde{cf}(v; \alpha, \tau)}{|A| (|V| - 1)} \quad (8.4)$$

que està dins de l'interval  $[0, 1]$ .

### **Anàlisis respecte de les proteïnes del càncer de mama**

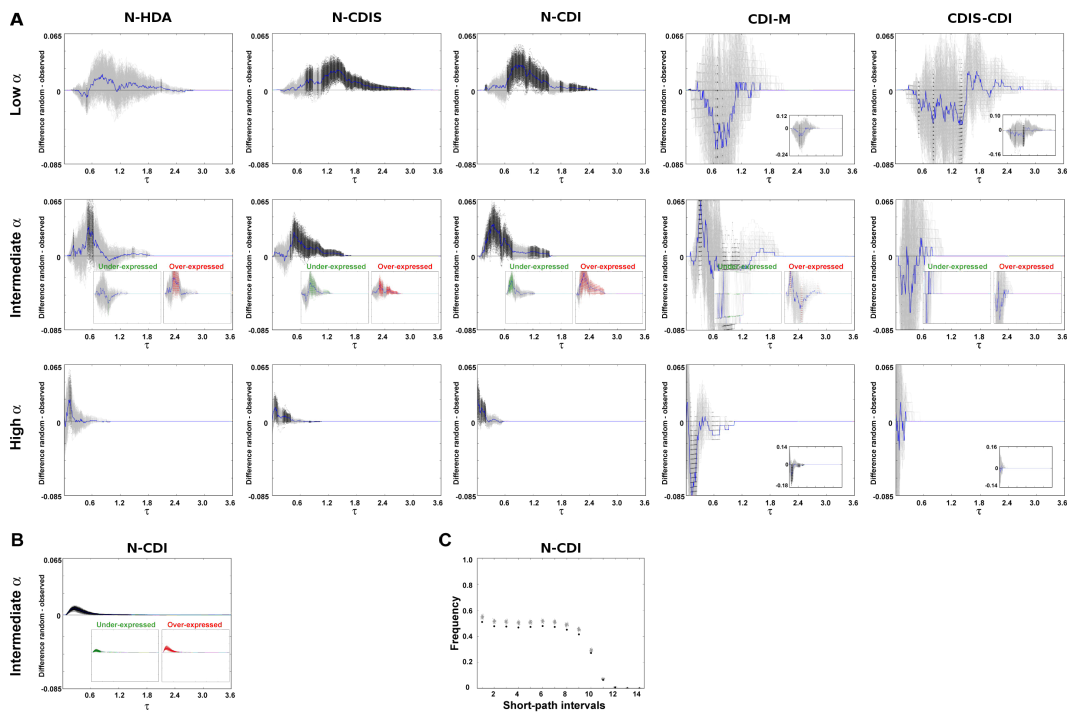
La xarxa de l'interactoma és una xarxa complexa amb molta informació codificada en la seva topologia. A causa de la seva complexitat i del seu coneixement incomplet, i de la mateixa complexitat del càncer, no es coneix com la dinàmica d'un càncer, mitjançant la desregulació gènica, pot afectar a l'interactoma.

Per analitzar el possible impacte del desenvolupament i progressió d'un càncer en l'interactoma, s'han analitzat els possibles patrons topològics que podrien definir en la xarxa de l'interactoma les proteïnes associades als diferents estadis del càncer usant el mètode de la cascada de fallades.

Per generar les fallades dins de la xarxa s'han de definir els vèrtexs inicials que propagaran la seva càrrega cap als seus veïns i generaran la cascada de fallades fins que s'assoleixi un estat estacionari, en el qual ja no es produiran més fallades. Com a vèrtexs inicials es prenen aquells vèrtexs corresponents a les proteïnes del càncer, un conjunt per a cada estadi i pel signe de la desregulació, és a dir si s'infraexpressen o es sobreexpressen.

Per estimar la significança del nombre de fallades generat per aquests conjunts en la xarxa de l'interactoma, s'han generat 500 conjunts de vèrtexs presos a l'atzar i s'han usat com a conjunt inicials per generar la cascada de fallades. S'han construït les distribucions nul·les empíriques per calcular la probabilitat del nombre de fallades que es produeixen en la xarxa. Tanmateix, per no esbiaixar els resultats segons la quantitat d'interactors de les proteïnes i la quantitat de mecanismes en els que puguin participar, s'ha calculat la probabilitat del nombre de fallades observat en una altra distribució nul·la empírica construïda a partir de conjunts aleatoris conservant la mida, la distribució del grau i amb un coeficient de clusterització similar (Wachi et al., 2005; Hernández et al., 2007; Syed et al., 2010). Inclús, es pot pensar que aquestes dues condicions considerades alhora, és una manera d'ajustar per la jerarquia que pot tenir el vèrtex dins de la xarxa, és a dir vèrtexs amb el mateix nombre de veïns i amb una estructura similar de l'entorn d'aquest vèrtex.





**Figura 8.3.5:** Anàlisi dels patrons topològics de la xarxa respecte dels conjunts associats al desenvolupament i progressió del càncer. En els gràfics, s'han representat les diferències entre els nombre de fallades pels conjunts aleatoris i els conjunts inicials definits pels canvis d'expressió. Aquestes diferències estan pintades amb color gris si la diferència no és significativa i amb color negre si ho és. La corba blava és la mitjana de les diferències. Cada punt de l'eix de les abscisses correspon a un valor de  $\tau$ , i l'eix de les ordenades la diferència del nombre de fallades (esperades per atzar menys observades) corresponents. L'escala dels eixos es la mateixa per a tots els gràfics. En les subfigures es mostren o bé el gràfic sencer de les diferències del nombre de fallades, quan sobrepassen els límits del gràfic principal, o bé les diferències del nombre de fallades pels subconjunts infraexpressats (verd) i sobreexpressat (vermell). S'observa que a mesura que el valor de  $\alpha$  creix, els valors de  $\tau$  pels quals s'observen diferències de fallades s'escurça, segurament a causa de que la capacitat dels vèrtexs és major i poden absorbir millor la càrrega transmesa.

- Diferència entre el nombre de les fallades entre els conjunts aleatoris i els inicials donats. Cadascuna de les files correspon un valor de  $\alpha$  diferent. S'observa que el nombre de fallades tendeix a ser positiu, és a dir que son majors les produïdes pels conjunts aleatoris que pels inicials. Aquests comportaments suggereixen una robustesa de la xarxa de l'interactoma respecte dels conjunts associats al càncer.
- Anàlogament a (a), representació del nombre de fallades per una alfa mitjana i pel cas  $N \rightarrow \text{CDI}$  i respecte dels conjunts aleatoris conservant només la mida del conjunt inicial. Tot i que la diferència és molt menor, és més significativa.
- Gràfic corresponent a una  $\alpha$  mitjana i al cas  $N \rightarrow \text{CDI}$  que mostra el nombre de fallades respecte els vèrtexs a distància  $d$  (eix abscisse) dels vèrtexs en els conjunts inicials. Llevat del primer veïnatge, en el qual es formen més fallades, en els conjunts formats per vèrtexs a distància  $2 \leq d \leq 9$ , tenen, aproximadament el mateix nombre relatiu de fallades, que cau ràpidament per  $d \geq 10$ . A més a més, en tots els casos, el nombre de fallades provocades pels conjunts aleatoris (conservant mida, distribució de grau i coeficient de clusterització) és major que el nombre de fallades produït pel conjunt de proteïnes del càncer de  $N \rightarrow \text{CDI}$ .

S'han analitzat els patrons generats per la cascada de fallades pels tres valors de  $\alpha$  per cadascun dels estadis del càncer, Figura 8.3.5. Per totes les alfes, s'observa que pels estadis  $N \rightarrow HDA$ ,  $N \rightarrow CDIS$  i  $N \rightarrow CDI$  el nombre de fallades observat és menor que els esperats per atzar. En canvi en els estadis  $CDI \rightarrow M$  i  $CDIS \rightarrow CDI$  hi ha una tendència d'invertir el patró observat pels altres casos, és a dir que el nombre de fallades observades és major que les esperades per atzar. Els resultats mostren patrons similars pels conjunts inicials definits per proteïnes infraexpressades i sobreexpressades, Figura 8.3.5a.

Si es comparen els resultats obtinguts construint els conjunts aleatoris preservant la mida, la distribució del grau i el coeficient de clusterització (Figura 8.3.5a) amb els resultats obtinguts només preservant la mida (Figura 8.3.5b), la diferència entre el nombre de fallades observat i l'esperat és major en el primer cas que en el segon. Tanmateix, en aquest últim, la diferència és més significativa.

També s'han calculat el nombre de fallades per una  $\alpha$  mitjana respecte del conjunt de proteïnes del càncer en el cas  $N \rightarrow IDC$  que provoquen en els conjunts  $V_d$  definits per vèrtexs a distància  $d$  del vèrtex inicial, és a dir

$$\sum_{v \in A_{N \rightarrow CDI}} \frac{1}{|V_d|} cf_{V_d}(v)$$

on  $A_{N \rightarrow CDI}$  correspon al conjunt inicial format per les proteïnes del càncer i la definició de  $V_d$  depèn de cada vèrtex inicial  $v$ . Primer, per a cada distància  $d$  s'observa que el nombre de fallades observat sempre és menor que el nombre de fallades produïts pels conjunts inicials aleatoris en cadascun dels conjunts  $V_d$ , Figura 8.3.5c. Segon, s'observa que la distribució té un caràcter uniforme per  $d < 10$  que després cau ràpidament.

Aquests resultats suggereixen que hi ha una relació de robustesa entre els estadis del càncer i la topologia de la xarxa de l'interactoma.

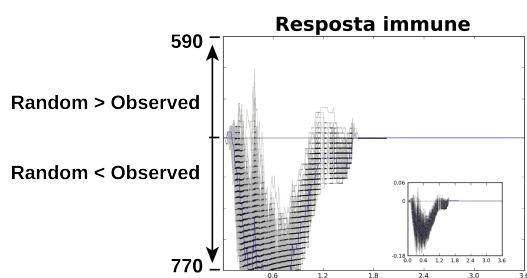
### 8.3.2 Anàlisi respecte de les vies de senyalització

Les vies de senyalització són conjunts d'elements que treballen amb la finalitat de dur a terme una funció cel·lular. Per veure quins patrons topològics estan associats a les proteïnes de les vies de senyalització, s'ha analitzat la robustesa de la xarxa respecte d'aquestes proteïnes per diverses vies de senyalització. Les vies usades s'han obtingut a través del repositori *Kyoto encyclopedia of Genes and Genomes* (KEGG)

(Kanehisa et al., 2015). De les 171 vies analitzades, 136 mostren patrons diferents als dels conjunts de proteïnes associades al càncer, Figura C.1.1.

Per altra banda, s'han trobat 35 vies que mostren patrons similars als de les proteïnes del càncer, Figura C.2.1. Entre aquestes 35 vies, hi ha vies directament relacionades en la formació i progressió de càncers com la via de senyalització MAPK o la del cicle cel·lular. Cal destacar també que, entre les 136 vies revelant patrons diferents als patrons revelats per les proteïnes del càncer, hi ha vies relacionades amb la resposta immune. Encara que el caràcter supressor/promotor del sistema immunològic és un punt controvertit en la literatura científica, es pot pensar que el rol del sistema immunològic és de supressor de tumors o cèl·lules canceroses (Bindea et al., 2010).

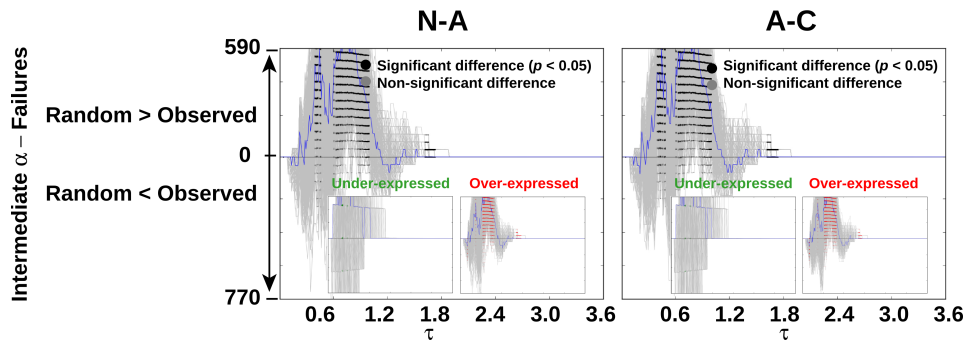
Per comprovar aquests resultats, s'ha definit el conjunt de totes les proteïnes amb el terme *Gene Ontology* (Ashburner et al., 2000) *immune response*. S'ha realitzat l'anàlisi respecte d'aquest conjunt inicial i s'ha observat que el nombre de fallades que provoca és més alt que l'esperat per atzar Figura 8.3.6.



**Figura 8.3.6:** Anàlisi de la topologia de la xarxa de l'interactoma respecte del terme *Gene Ontology* resposta immune (cas  $N \rightarrow \text{CDI}$  i  $\alpha$  intermitja).

### 8.3.3 Anàlisi respecte de les proteïnes del càncer de colon

Per confirmar els patrons observats en el cas del càncer de mama, s'han repetit les anàlisis de la topologia de la xarxa respecte del desenvolupament i progressió del càncer de colon (Galamb et al., 2010). Anàlogament al cas del càncer de mama, s'han definit els conjunts inicials per generar les cascades de fallades per proteïnes associades als estadis del càncer de colon. Els resultats observats en aquest cas són anàlegs als del càncer de mama, és a dir, la xarxa té un caràcter més robust respecte de les proteïnes del càncer de colon, Figura 8.3.7 i Figura B.0.1.



**Figura 8.3.7:** Representació del nombre de fallades pels casos N→A (normal, adenoma) i A→C (adenoma, carcinoma) del càncer colorectal, per  $\alpha$  intermitja. S'observa que els conjunts inicials tendeixen a causar un nombre menor de fallades que els esperats per atzar. Els resultats pels altres valors de  $\alpha$  revelen el mateix patró en la xarxa de l'interactoma.

## 8.4 Robustesa de la progressió del càncer respecte l'estadi inicial

Fins ara s'han exposat les anàlisis de la topologia de la xarxa respecte del desenvolupament i progressió del càncer. S'ha vist, a trets generals, que la dinàmica del càncer revela patrons de robustesa de la xarxa. Per analitzar si hi ha alguna relació topològica entre dos estadis de la progressió del càncer, s'han analitzat els patrons topològics respecte a l'estadi inicial N→HDA en els altres estadis posterior.

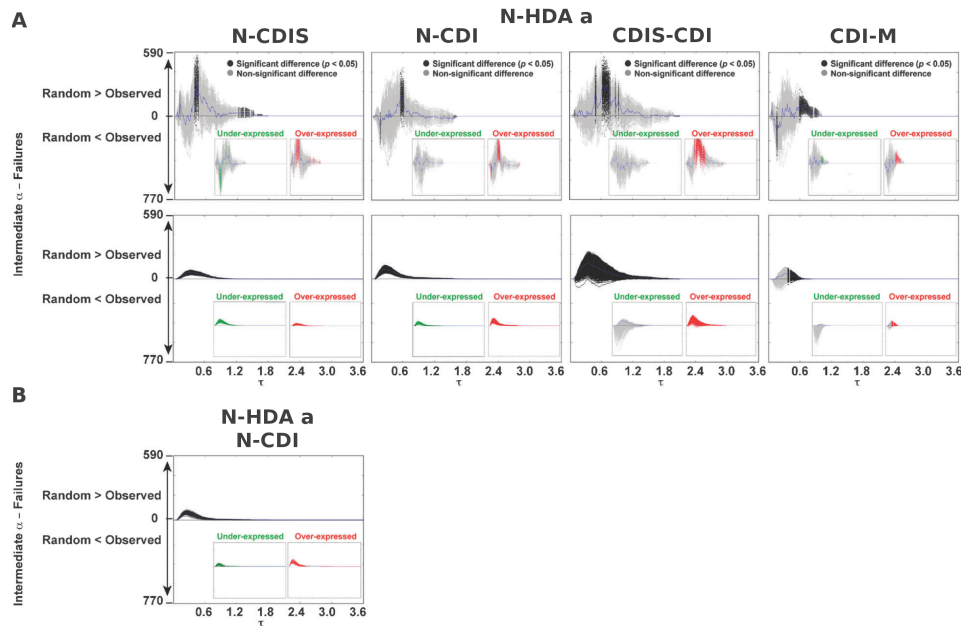
Per realitzar aquesta anàlisi, s'han comptat el nombre de fallades que provoquen les proteïnes de l'estadi N→HDA en els estadis subsegüents. Per calcular la significança del nombre de fallades, provocades pel conjunt inicial s'ha calculat la probabilitat respecte de distribucions nul·les empíriques. Per construir les distribucions, s'han comptat el nombre de fallades que provoquen 500 conjunts construïts prenent vèrtexs a l'atzar. S'han construït tres distribucions respecte de tres tipus de conjunts aleatoris, una conservant només la mida dels conjunts. Una segona conservant, a més, la distribució del grau. I una tercera conservant, a més, el coeficient de clusterització.

Per calcular la significança del nombre de fallades que han patit els conjunts de proteïnes dels estadis subsegüents de l'estadi inicial, s'han calculat la probabilitat del nombre de fallades rebudes en distribucions nul·les empíriques. Aquestes distribucions empíriques s'han construït de la mateixa manera que en el cas anterior.

Els resultats obtinguts, revelen patrons topològics molt similars als patrons observats en els casos anteriors. S'observa que el nombre de fallades que provoca l'estadi inicial

en els estadis subsegüents és menor que l'esperat per atzar, Figura 8.4.1a.

Anàlogament, el nombre de fallades que reben els estadis subsegüents respecte de l'estadi inicial també es menor que l'esperat per atzar. Tanmateix, tot i que la diferència és menor que en el cas anterior, és més significativa, Figura 8.4.1b.



**Figura 8.4.1:** Gràfics del nombre de fallades aleatòries respecte de les observades que generen les proteïnes de l'estadi inicial  $N \rightarrow HDA$  en els estadis subsegüents  $N \rightarrow CDIS$ ,  $N \rightarrow CDI$ ,  $CDIS \rightarrow CDI$  i  $CDI \rightarrow M$ . En el cas (a) es mostren els resultats respecte de conjunts aleatoris conservant mida i distribució del grau. El cas que, a més, conserva el coeficient de clusterització, no es mostra perquè els resultats són similars.

(a) A dalt, nombre de fallades aleatòries respecte de les observades. Aquests gràfics mostren el comportament del nombre de fallades que generen les proteïnes de l'estadi inicial del càncer en les proteïnes dels estadis subsegüents. En aquest cas, el nombre de fallades aleatòries s'han comptat mitjançant conjunts aleatoris del conjunt inicial que genera la cascada de fallades.

A baix, comportament de les fallades que pateixen els estadis subsegüents a l'estadi inicial respecte d'aquest mateix. En aquest cas, el nombre de fallades aleatòries s'han comptat mitjançant conjunts aleatoris anàlegs als conjunts de proteïnes dels estadis subsegüents a l'inicial.

(b) Comportament de les fallades que pateixen les proteïnes dels estadis subsegüents a l'inicial. La diferència del nombre de fallades entre les aleatòries i les observades s'han calculat mitjançant les fallades que provoquen les proteïnes de l'estadi inicial en els conjunts aleatoris anàlegs als subsegüents. Aquests conjunts, però, s'han construït només conservant la mida dels conjunts, no la distribució del grau.

## 8.5 Robustesa de la progressió del càncer respecte del tractament

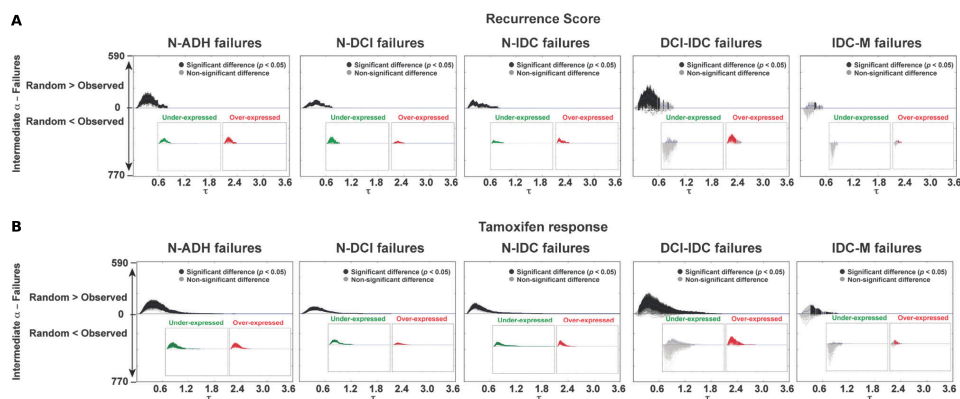
Una altra pregunta interessant és quina relació poden tenir les proteïnes associades a teràpies del càncer amb les proteïnes associades al càncer. Per analitzar aquesta relació, s'ha repetit l'anàlisi de la cascada de fallades per un conjunt de proteïnes associades al tractament del càncer.

Molts dels càncers de mama s'estimulen a través d'hormones com els estrògens. Les cèl·lules tenen receptors d'aquestes hormones que transmeten els senyals cap a l'interior de la cèl·lula i estimulen el seu creixement. Aquests tipus de càncers es classifiquen com a positius en receptors d'estrògens ( $ER\alpha$ ).

Una de les teràpies més comunes pels càncers de mama  $ER\alpha$  positius és el tamoxifè, que inhibeix la unió entre els receptors i els estrògens, impedit d'aquesta manera la cascada de senyals intracel·lular que promou el creixement cel·lular (Dowsett et al., 2005; Massarweh et al., 2008; Osborne, 1998). Per analitzar la possible relació del pronòstic i la predicció amb la formació i progressió del càncer en la xarxa de l'interactoma, s'han definit dos conjunts de proteïnes. Un definit per la signatura de gens associats a la recurrència del càncer de mama després d'haver estat tractats amb tamoxifè (Paik et al., 2004). Un segon corresponent a la resposta de cèl·lules tractades amb estrògens, per estimular el creixement cel·lular.

Per identificar les proteïnes que responen entre el tractament amb estrògens i amb tamoxifè, s'han analitzat els nivells d'expressió gènica entre els dos tractaments per cèl·lules MCF7. Les cèl·lules MCF7 són cèl·lules derivades d'una pacient de càncer de mama. Aquestes cèl·lules són  $ER\alpha$  positius. Per comparar comportaments antagònics, s'han tractat amb estrògens per incentivar el creixement cel·lular i amb tamoxifè per bloquejar la unió entre els estrògens i els seus receptors.

Després de definir els conjunts inicials, per a cada cas s'han analitzat les fallades generades en cadascun dels estadis del càncer, Figura 8.5.1. Ambdós casos, les proteïnes del càncer pateixen menys fallades que les esperades per atzar.



**Figura 8.5.1:**

- Gràfic del nombre de fallades respecte de les proteïnes associades a la recurrència del càncer de mama. Els patrons revelen que el nombre de fallades que pateixen les proteïnes associades al càncer és menor que l'esperada per atzar. Per calcular la significança s'ha seguit la mateixa estratègia usada en casos anteriors, construint una distribució empírica respecte dels conjunts de proteïnes associades al càncer.
- Anàlogament, el nombre de fallades de les proteïnes del càncer, respecte del conjunt de proteïnes que responen al tractament entre estrògens i tamoxifè, és menor que l'esperat per atzar.

## 8.6 Robustesa de la progressió del càncer respecte a la pertorbació gènica

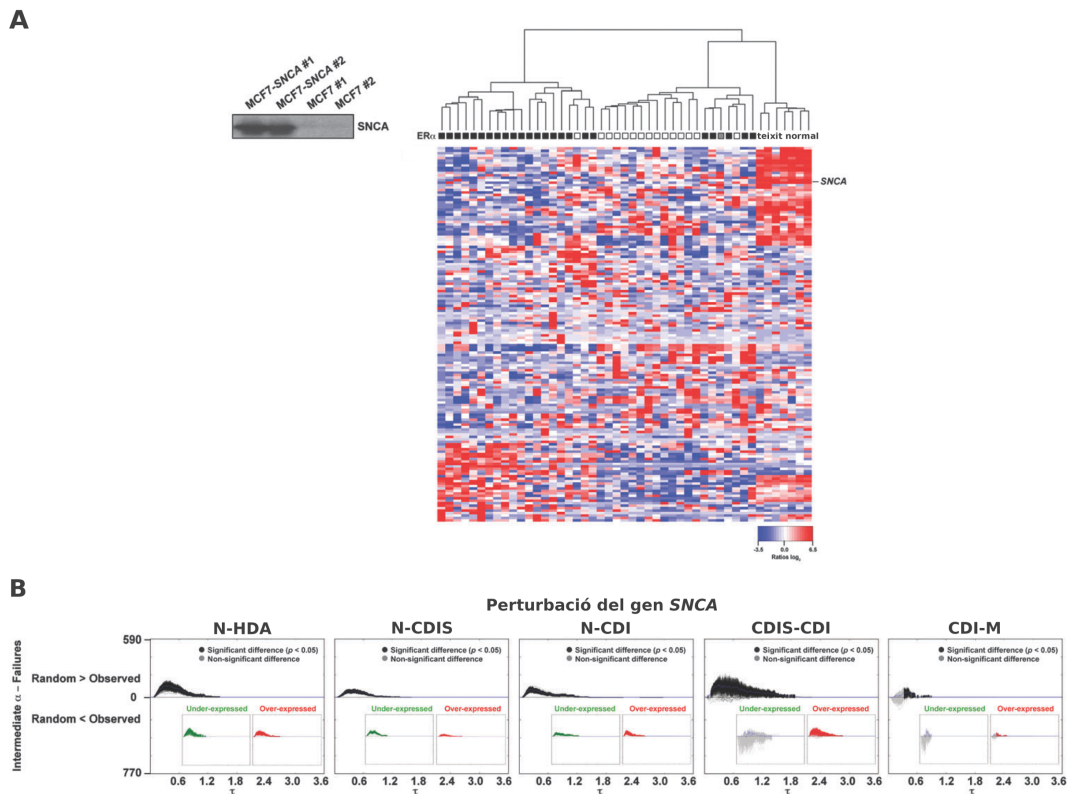
En molts càncers de mama i en les cèl·lules MCF7 el gen *SNCA* està infraexpressat (Chin et al., 2006; Neve et al., 2006). A més, aquest gen té un centralitat relativament gran en l'interactoma ( $k = 39$ ), indicant que pot està involucrat en varis mecanismes o vies. Per veure quina

relació pot haver-hi entre els canvis d'activitat d'aquest gen en les cèl·lules del càncer de mama, s'han transfectat les cèl·lules MCF7 amb

*SNCA* per induir-ne la seva expressió i s'han comparat els nivells d'expressió de les proteïnes en la cèl·lula entre les cèl·lules no transfectades i transfectades.

Tot seguit, s'ha calculat una clusterització dels tumors respecte d'aquests gens diferencialment expressats. S'observa que els gens que diferencien els dos tipus de cèl·lules, transfectades i no transfectades, també separen bé l'estat positiu/negatiu del receptor d'estrogen  $ER\alpha$ , Figura 8.6.1a.

Per analitzar la relació entre aquests gens i els associats al desenvolupament i progressió del càncer, s'ha definit com a conjunt inicial les proteïnes associades als canvis d'expressió entre MCF7 i MCF7 transfectades, i s'ha realitzat un anàlisi de cascada de fallades. Els resultats obtinguts són similars que els exposats fins ara, és a dir, el nombre de fallades que pateixen les proteïnes associades al desenvolupament i progressió del càncer es menor que l'esperat per atzar, Figura 8.6.1b



**Figura 8.6.1:**

- A la dreta, western blot mostrant els nivells de la proteïna SNCA en les cèl·lules MCF7 transfectades i les no transfectades. A la dreta, clusterització no supervisada de tumors primaris i teixit normal usant gens diferencialment expressats entre les MCF7 transfectades i no transfectades. Tal i com s'observa, els gens diferencialment expressats usats per la clusterització distingeixen amb claredat l'estat positiu/negatiu (negre/blanc respectivament) del receptor d'estrogen.
- Gràfic del nombre de fallades en els conjunts associats al desenvolupament i progressió del càncer respecte de la signatura definida per les cèl·lules transfectades amb *SNCA*. El nombre de fallades que reben les proteïnes del càncer és menor que les esperades per atzar.



## 8.7 Les proteïnes involucrades en l'autofàgia mostren patrons diferents als del càncer

Per complementar les anàlisis exposades fins ara, s'ha realitzat una anàlisi per detectar quines característiques biològiques podrien estar associades a patrons diferents als patrons associats al càncer, dins de la xarxa de l'interactoma.

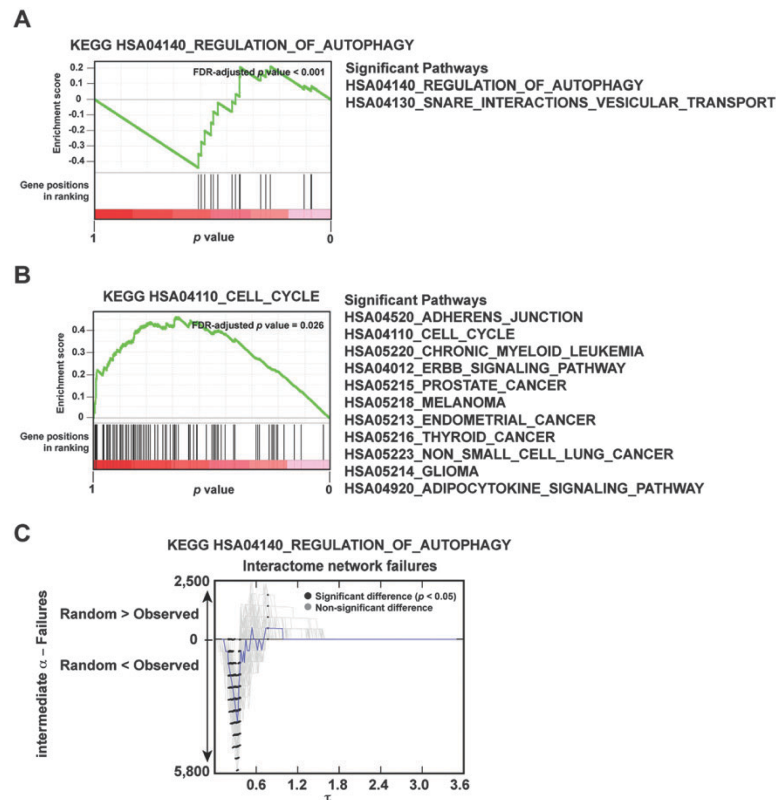
Per realitzar aquesta anàlisi, s'han calculat el nombre de fallades que generen cadascuna de les proteïnes de la xarxa i s'ha calculat la probabilitat de causar el nombre de fallades observat o més. Per calcular aquesta probabilitat per a cada vèrtex  $v$ , s'ha construït una distribució empírica del nombre de fallades calculant el nombre de fallades que causen els vèrtexs de la xarxa amb el mateix grau que  $v$ .

Tot seguit, s'han ordenat les proteïnes segons aquesta probabilitat. Per capturar aquelles vies que puguin tenir una concentració de proteïnes cap a un dels extrems d'aquesta llista —això és causant més o menys nombre de fallades que les esperades—, s'ha aplicat l'eina GSEA (que correspon a un test de Kolmogorov-Smirnov) (Subramanian et al., 2005), que captura com es distribueixen un conjunt d'elements en un altre conjunt que el conté.

Les vies analitzades corresponen a totes les vies KEGG (Kanehisa et al., 2015) i, mitjançant el GSEA, s'ha mesurat com els seus elements es distribueixen dins de la llista de proteïnes de la xarxa de l'interactoma ordenada segons la probabilitat del nombre de fallades que generen. S'han pres aquelles vies que els seus elements tendien a distribuir-se cap als extrems de la llista total.

Seguint la tendència observada en les anteriors anàlisis, les vies associades a un nombre de fallades menor, són aquelles involucrades en el càncer, com per exemple la via cicle cel·lular, Figura 8.7.1b.

Per altra banda, les vies associades a proteïnes que generen un nombre més gran de fallades respecte de l'esperat, són vies que no estan directament relacionades amb el càncer, encara que puguin tenir un rol rellevant en el seu desenvolupament i progressió. Per exemple, la via regulació de l'autofàgia, que és un mecanisme de mort cel·lular i que les cèl·lules canceroses han d'evadir per poder proliferar de forma aberrant, Figura 8.7.1a i Figura 8.7.1c.



**Figura 8.7.1:**

- Resultat del GSEA per la via regulació de l'autofàgia (Subramanian et al., 2005). La part inferior del gràfic correspon a la llista de proteïnes univers ordenades segons la probabilitat de generar el nombre de fallades observat o major. Per a la via estudiada, s'indiquen les posicions dels seus elements en la llista univers. Així, aquelles probabilitats a prop del zero corresponen a proteïnes que generen un nombre de fallades gran. S'observa que aquesta via tendeix a tenir les proteïnes que l'integren cap a la cua de proteïnes que generen més fallades que les esperades per atzar.
- Resultats de les vies integrades per proteïnes que generen un nombre de fallades baix. Es representa la distribució de les proteïnes del cicle cel·lular a mode d'exemple.
- Comprovació pel cas de la via regulació de l'autofàgia.

## 8.8 Robustesa topològica respecte la distància dels vèrtexs

Els gens mutats o els productes gènics diferencialment expressats associats al càncer tendeixen a estar més a prop dins de l'interactoma que d'altres gens no associats al càncer, és a dir a tenir una distància mitjana més petita que l'esperada per atzar (Hernández et al., 2007; Wachi et al., 2005; Syed et al., 2010; Dezso et al., 2009;

Rambaldi et al., 2015; Jonsson i Bates, 2006). Per analitzar com pot contribuir aquesta propietat en la xarxa de l'interactoma i la seva topologia, s'han construït conjunts de vèrtexs enriquits per elements a una distància  $d$  escollida entre ells i s'han estudiat les fallades que generen aquests conjunts.

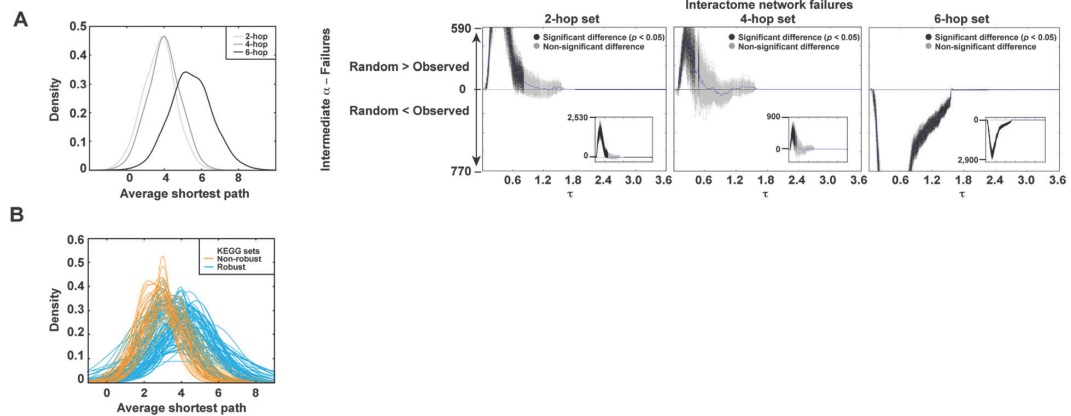
Per construir aquests conjunts  $V_d$ , s'ha fixat una distància  $d \in \{2, 4, 6\}$  i s'ha pres un vèrtex  $v$  a l'atzar de la xarxa. Per enriquir el conjunt  $V_d$  amb elements que estan a distància  $d$  entre ells, s'han seguit els passos següents:

1. es defineix  $V_d = \emptyset$
2. es defineix el nou conjunt  $V_d$  com  $V_d \cup \{v\}$
3. es defineix el conjunt de vèrtexs  $V'$  que estan a distància  $d$  de  $v$
4. es pren un vèrtex  $u$  a l'atzar de  $V'$
5. es redefineix  $v = u$

i s'han repetit els passos 2-5 fins que el conjunt conté 1 000 elements Figura 8.8.1a. S'han considerat els conjunts  $V_2$ ,  $V_4$  i  $V_6$  com a conjunts inicials per iniciar les casca-des de fallades i s'han comparat el nombre de fallades produïdes respecte a l'esperat per atzar, considerant conjunts aleatoris respectant la mida del conjunt i la distribució del grau, però no l'enriquiment de distàncies entre elements.

Amb aquest procediment, s'han obtingut els conjunts  $V_2$  (amb distància mitjana 3.43),  $V_4$  (amb distància mitjana 3.74) i  $V_6$  (amb distància mitjana 5.44). Els conjunts de vèrtexs enriquits per una distància petita i mitjana,  $V_2, V_4$  tendeixen a generar cascades que produeixen moltes menys fallades que les esperades per atzar, mentre que el conjunt  $V_6$  moltes més Figura 8.8.1a.

Per veure si aquests resultats podrien explicar els resultats de l'anàlisi de les 136 vies KEGG que han generat patrons diferents als patrons generats pels conjunts associats al desenvolupament i progressió del càncer, s'han calculat les distàncies entre els elements de totes les vies KEGG Figura 8.8.1b. La tendència, com en els conjunts  $V_2, V_4$  i  $V_6$ , és que les 136 vies KEGG —les quals revelaven patrons diferents o oposats als dels conjunts associats al càncer—tendeixen a tenir una distribució de les distàncies centrades en un valor gran. Per altra banda, les 35 vies KEGG que mostraven patrons similars als dels conjunts associats al càncer, tendeixen a tenir les distàncies centrades en un valor més petit, Figura 8.7.1b. Si bé aquesta



**Figura 8.8.1:**

- (a) A l'esquerra, distribucions de les distàncies entre els elements dels conjunts  $V_d$ ,  $d \in \{2, 4, 6\}$ . S'observa, a la dreta, que per  $d \in \{2, 4\}$  el nombre de fallades és menor que l'esperat per atzar, mentre que per  $d = 6$  el nombre de fallades és molt superior a l'esperat per atzar.
- (b) Distribucions de les distàncies entre els elements de les vies KEGG dins de la xarxa de l'interactoma.

és una tendència, s'observen algunes vies que no compleixen aquest comportament. Per exemple, els elements de la via *SNARE Interactions in Vesicular Transport* tenen una distància mitjana de 2.69 i, en canvi, han revelat patrons diferents als del conjunt associats al càncer. Aquest cas suggereix que, si bé la distància entre elements pot capturar part dels patrons en la xarxa de l'interactoma, hi pot haver altres condicions topològiques per explicar el comportament dels patrons.

També s'han calculat les distàncies entre les proteïnes associades a  $N \rightarrow HDA$  i  $CDI \rightarrow M$  i són 4.09 i 4.26, respectivament, amb concordança amb aquest anàlisi.

# 9

## Resultats II

### 9.1 L'activitat de la xarxa captura informació de la resposta de fàrmacs

Com que els gens estan funcionalment i topològicament organitzats en les xarxes complexes moleculars (Barabasi i Oltvai, 2004; Goh et al., 2007) i en les cèl·lules canceroses els processos biològics i vies de senyalització són robustos a perturbacions, una mesura de l'activitat de la xarxa podria capturar la sensibilitat d'una teràpia donada. Aquí, s'ha adreçat aquesta qüestió integrant les interaccions proteiques humanes (la xarxa de l'interactoma), la mesura de l'expressió basal gènica en centenars de línies cel·lulars del càncer que representen la diversitat dels càncers humans, i els valors de concentració inhibidòria de l'activitat cel·lular a la meitat ( $IC_{50}$ ) de 130 fàrmacs que han estat analitzats en estudis previs (Garnett et al., 2012).

Per calcular l'activitat de la xarxa  $NA$  per a cadascuna de les línies cel·lulars del càncer, s'ha calculat la matriu d'adjacència amb pesos  $W = (w_{ij})_{ij}$  a partir de les dades d'expressió basal de la línia, on  $w_{ij}$  correspon als pesos de l'aresta incident als vèrtexs  $i, j$  de la xarxa de l'interactoma

$$w_{ij} = x_i \frac{x_j}{\sum_{\ell \in \Gamma_i} x_\ell} + x_j \frac{x_i}{\sum_{\ell \in \Gamma_j} x_\ell}$$

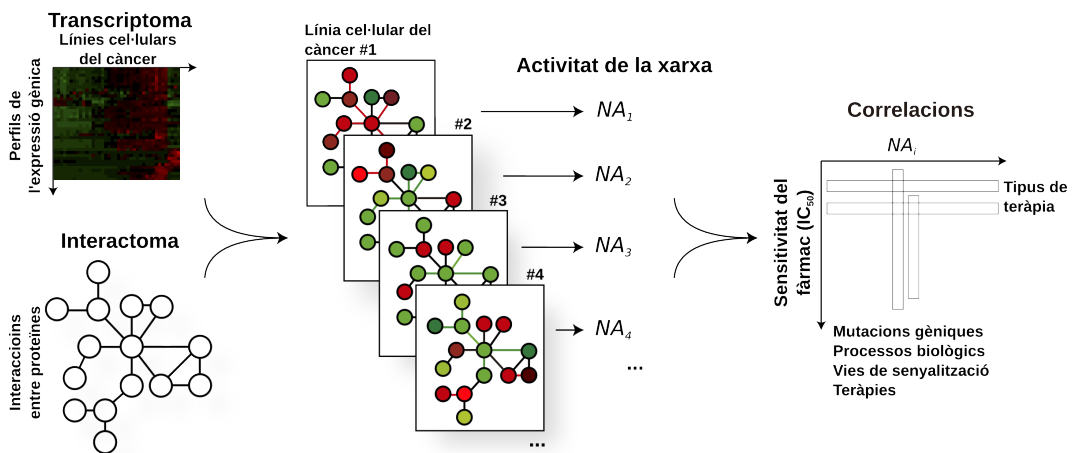
on  $x_i$  és l'expressió del gen associat al vèrtex  $i$  de la xarxa i  $\Gamma_i$  el conjunt dels veïns del vèrtex  $i$ . L'activitat de la xarxa s'ha definit com la suma de la centralitat sungraf dels seus vèrtexs (Estrada i Rodríguez-Velázquez, 2005; Estrada, 2000) basada en

la matriu de pesos  $W$ ,

$$C_G(i) = (e^W)_{ii} = \sum_{k \geq 0} \frac{W^k}{k!}$$

$$NA(W) = \sum_{i \in V} C_G(i)$$

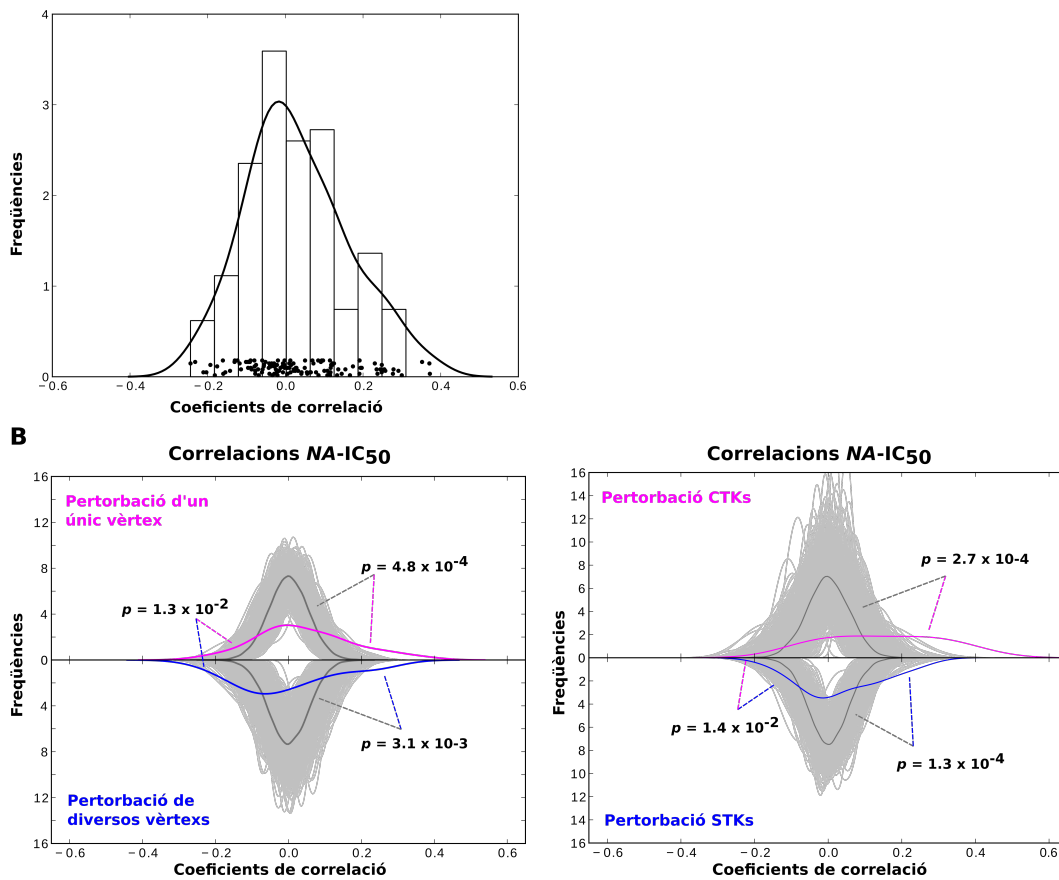
on  $G = (V, E)$  és la xarxa de l'interactoma. Així, per a cada línia cel·lular del càncer s'obté un valor d'activitat de la xarxa  $NA$  que, potencialment, captura els mecanismes associats als càncers i, que comparant-los amb les  $IC_{50}$ , es pot capturar informació sobre la sensibilitat de les cèl·lules respecte de les teràpies, Figura 9.1.1.



**Figura 9.1.1:** Estratègia usada per l'anàlisi. Per a cada línia cel·lular del càncer, s'han integrat les dades d'expressió gènica a la xarxa de l'interactoma i s'ha calculat l'activitat de la xarxa  $NA_i$ . Finalment, s'han calculat els coeficients de correlació de Pearson per a cada droga entre l'activitat de la xarxa i les  $IC_{50}$  respecte diverses condicions. Els coeficients de correlació s'han calculat o bé a través de totes les 595 línies del càncer, o bé respecte de les línies complint una certa condició, com per exemple l'estatus de mutació d'un gen. Per comparar els coeficients de correlació de cada fàrmac, s'han calculat les distribucions dels fàrmacs respecte diferents característiques que separen els fàrmacs, per exemple entre fàrmacs citotòxics o dirigits a teràpies.

S'han calculat i analitzat els coeficients de correlació respecte de diverses condicions, com per exemple respecte de mutacions gèniques, vies de senyalització o tipus de teràpies.

Per reflectir les possibles associacions entre l'activitat de la xarxa i la sensibilitat de les línies respecte a teràpies, s'han calculat els coeficients de correlació de Pearson



**Figura 9.1.2:**

- (a) Distribució dels coeficients de correlació de Pearson a través de les 595 línies cel·lulars del càncer i per a cada droga. Cada punt correspon al coeficient de distribució d'un fàrmac entre les IC<sub>50</sub> i les mesures de la xarxa NA.
- (b) A l'esquerra, distribució dels coeficients de correlació de Pearson a través de les 595 línies cel·lulars però només pels fàrmacs amb una sola diana (lila) o múltiples dianes (blau). Les diferències entre les distribucions de les dues classes de fàrmacs és significativa (test de Wolkocoxn  $p = 1.3 \times 10^{-2}$ ). Les distribucions de color gris corresponen a permutacions de les  $NA_i$  (a través de les línies cel·lulars), mostrant la distribució dels coeficients de correlació esperada per atzar. A la dreta, les distribucions dels coeficients de correlació entre les NA i les IC<sub>50</sub> de les famílies de proteïnes tirosina-cinases (CTKs) i serina/treonina cinases (STKs). Amb concordança amb el resultat anterior, les STKs tendeixen a està centrades a valors negatius, respecte de les CTKs. Aquest resultat podria reflectir la major pertorbació de les drogues que tenen per diana les proteïnes de les STKs ja que és una família molt més gran de proteïnes que la CTKs (Shi, 2009).

entre les  $NA$  i les  $IC_{50}$  per a cada fàrmac a través de les línies cel·lulars.

Per comparar els efectes dels fàrmacs en la xarxa dels càncers entre ells, s'ha calculat la distribució dels coeficients de correlació. Mentre que la distribució dels coeficients de correlació de tots els fàrmacs està centrada en el zero (Figura 9.1.2a), s'han trobat diferències significatives entre les distribucions dels coeficients de correlació respecte de les característiques dels diferents fàrmacs, Figura 9.1.2. S'observa que la  $IC_{50}$  dels fàrmacs que afecten a més processos, per exemple els usats en la quimioteràpia, tendeixen a tenir una correlació negativa amb l'activitat de la xarxa  $NA$ , Figura 9.1.2b i Taula D.0.1. Per altra banda, les  $IC_{50}$  dels fàrmacs que defineixen teràpies dirigides a una diana concreta tendeixen a correlacionar positivament amb les  $NA$ . A més, la diferència entre els dos comportaments és significativa, Figura 9.1.2b.

En general, s'han trobat diferències significatives entre les diferents classes de fàrmacs, diferents tipus de famílies de dianes i les diferents vies de senyalització a les que afecten, Taula E.0.1. Aquests resultats suggereixen que les mesures  $NA$  de la xarxa, captura informació biològica rellevant relacionada amb la sensibilitat terapèutica o els mecanismes d'acció.

## 9.2 Els processos biològics influeixen en les associacions entre l'activitat de la xarxa els fàrmacs

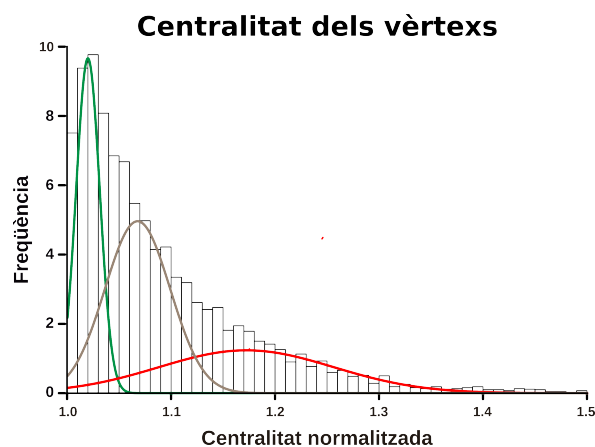
Com que l'activitat de la xarxa està definida respecte de la centralitat o activitat de cada gen (Equació 7.3), s'han considerat, només, els gens amb la centralitat més alta (tercil superior del conjunt de totes les centralitats, Figura 9.2.1). Primer, s'ha comprovat que aquests gens capturessin la major part de les mesures  $NA$ , Figura 9.2.2. Tot seguit, s'ha realitzat un test d'enriquiment d'aquest conjunt de gens de processos biològics (*Gene Ontology* (Ashburner et al., 2000)). Entre les processos biològics obtinguts, hi ha els de l'apoptosi, expressió gènica o resposta immune que estan relacionats en el desenvolupament i progressió del càncer (Hanahan i Weinberg, 2011), Taula 9.2.1.

Així, aquestes anàlisis suggereixen que diversos processos biològics estan associats a l'activitat de la xarxa del càncer  $NA$  i que poden contribuir entre les associacions



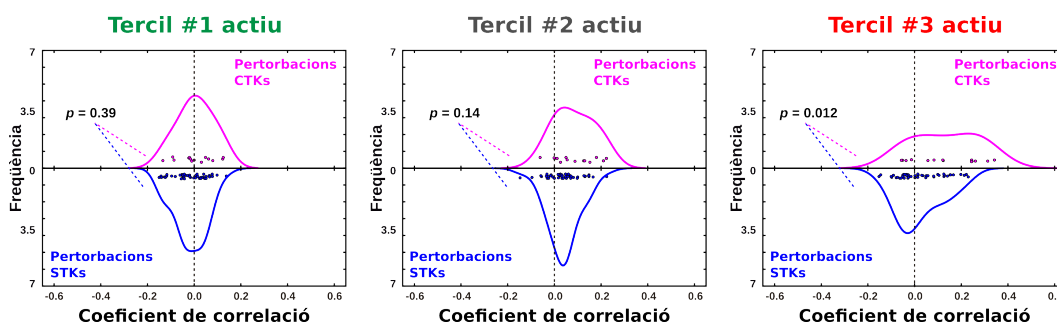
**Taula 9.2.1:** Taula dels termes *Gene Ontology* sobrerepresentats en el conjunt dels vèrtexs amb centralitat alta (FDR 5%), Figura 9.2.1

ID del terme	nom del terme
GO:0072594	Establishment of protein localization to organelle
GO:0009057	Macromolecule catabolic process
GO:0019058	Viral infectious cycle
GO:0046907	Intracellular transport
GO:0015031	Protein transport
GO:0019083	Viral transcription
GO:0044248	Cellular catabolic process
GO:0010467	Gene expression
GO:0070727	Cellular macromolecule localization
GO:0008104	Protein localization
GO:0022411	Cellular component disassembly
GO:0043933	Macromolecular complex subunit organization
GO:0061024	Membrane organization
GO:0007276	Gamete generation
GO:0007281	Germ cell development
GO:0071842	Cellular component organization at cellular level
GO:0045767	Regulation of anti-apoptosis



**Figura 9.2.1:** Per dividir el conjunt de les centralitats dels vèrtexs, s'ha ajustat la distribució de centralitats mitjançant tres distribucions normals (paquet mixtools del llenguatge R (Benaglia et al., 2009)) que han definit tres conjunts de vèrtexs amb centralitat petita, mitjana i gran segons si la mitjana de les distribucions és petita, mitjana o gran, respectivament.

de les *NA* i la sensibilitat de teràpies específiques.



**Figura 9.2.2:** Distributions dels coeficients de correlació de Pearson entre les  $NA$  (segons la centralitat dels gens) i les  $IC_{50}$  a través de les 595 línies del càncer. Es comparen les distribucions dels coeficients entre les famílies CTKs (lila) i STKs (blau). Per calcular les mesures  $NA$ , els gens d'un dels tres conjunts definits per la centralitat han conservat els valors d'expressió, pels altres dos se'ls hi ha assignat l'expressió esperada de l'experiment. Després, s'ha tornat a calcular la matriu de pesos i les mesures  $NA$ . D'esquerra a dreta, gràfics corresponent al conjunt de gens amb centralitat baixa, mitjana i alta. Tal i com s'esperava, els gens amb major centralitat capturen la proporció més gran d'informació.

## 9.3 Les mutacions conductores del càncer influeixen en les associacions

### 9.3.1 Diferències respecte de l'estatus de mutació dels gens conductors del càncer

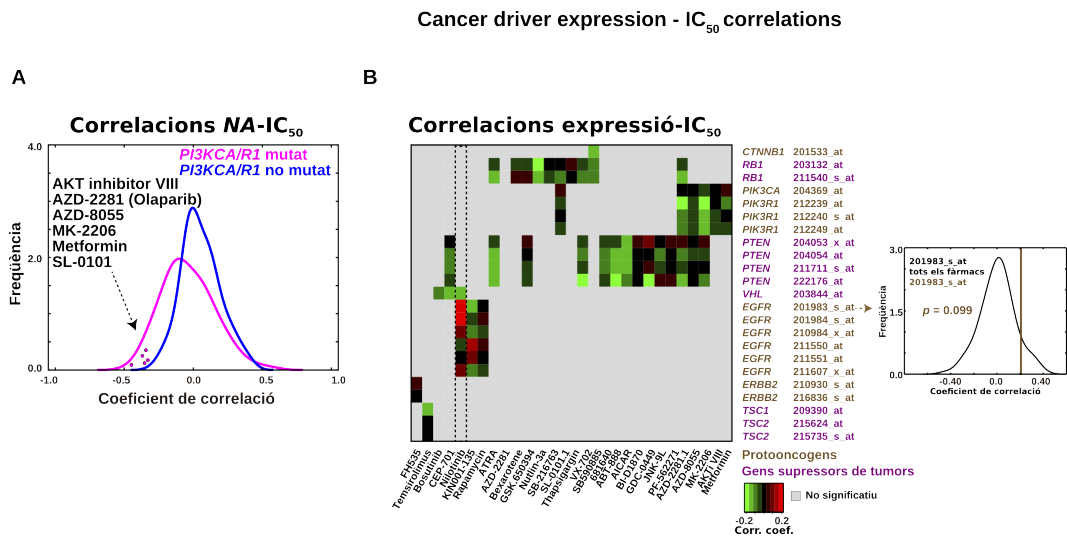
Com que l'activitat de la xarxa  $NA$  distingeix entre els diferents tipus de fàrmacs i diversos processos biològics poden influir en les diferències observades, s'han analitzat les associacions segons les mutacions dels protooncogenes i supressors de tumors per veure si també podrien explicar associacions o diferències. S'han calculat els coeficients de correlació per a cada droga i a través de les línies portadores dels gens mutats i de les línies no portadores dels gens mutats. La comparació de les distribucions dels coeficients de correlació de Pearson entre l'estatus de mutació dels gens mostren diferències significatives entre les distribucions dels coeficients de correlació entre les  $NA$  i les  $IC_{50}$  dels fàrmacs pels protooncogenes *CTNNB1*, *EGFR*, *ERBB2* i *PIK3CA* (conjuntament amb la mutació *PIK3R1*) i pels supressors de tumors *PTEN*, *RB1*, *TSC1/2* i *VHL*, Figura 9.3.1a i Taula G.0.1.

Per identificar els fàrmacs associats a l'estat mutat dels gens, s'han comparat els coeficients de correlació de Pearson de cada fàrmac a través segons l'estatus de mutació gens mutats i s'han considerat aquells amb els coeficients de correlació més

extremes (respecte dels no mutats),

$$F_{\text{mut}} = \{i \mid \rho_i < \min_{\rho' \in P'} \rho', \rho_i \in P\} \tag{9.1}$$

on  $P$  és el conjunt dels coeficients de correlació de Pearson a través de les línies portadores del gen mutat i  $P'$  els de les línies no portadores del gen mutat,  $\rho_i$  és el coeficient de correlació del fàrmac  $i$ . S'ha escollit aquest criteri per detectar els casos més extrems.



**Figura 9.3.1:**

- (a) Distribucions dels coeficients de correlació de Pearson de les drogues a través de les línies cel·lulars del càncer segons si son portadores del gen *PIK3CA* mutat (lila) o no (blau). Els fàrmacs amb coeficients més extrems de les cèl·lules portadores del gen mutat i més petits que els coeficients respecte de les no mutades s'han considerat pels anàlisis posteriors.
- (b) A l'esquerra, *heatmap* mostrant el resultat de les correlacions de Pearson entre l'expressió gènica i les IC<sub>50</sub> per a les drogues associades a l'estatus mutat per algun dels gens conductors del càncer. Es mostren només les correlacions significatives amb nivell del 5%. A la dreta, distribució dels coeficients de correlació entre l'expressió de *EGFR* i les IC<sub>50</sub> de tots els fàrmacs. Amb marró s'indica el cas corresponent al fàrmac nolitnib. Encara que la correlació entre *EGFR* i nolitnib és marcadament positiva, no està entre les més significatives ( $p = 0.099$ ), suggerint que les *NA* han capturat aquella part d'activitat que podria explicar que aquesta correlació no fos més gran.

### 9.3.2 Concordança entre l'activitat i els perfils d'expressió gènica

Un 93% dels fàrmacs (26/28) capturats per les diferències que s'han observat mitjançant la correlació de les *NA* i la  $IC_{50}$  a través de les mutacions dels protooncogenes o supressors de tumors, concorden amb la corresponent correlació entre l'expressió gènica i la  $IC_{50}$ , Figura 9.3.1b-esquerra. Tanmateix, les diferències entre l'activitat de la xarxa i l'expressió gènica poden explicar-se mitjançant la integració de les dades de l'interactoma a l'hora de calcular les *NA*, capturant la influència de processos biològics. Es pot pensar doncs, que la mesura de les *NA* captura l'impacte global d'una mutació gènica conductora del càncer. Per exemple, entre els fàrmacs considerats, s'ha observat que la  $IC_{50}$  del nilotinib té una forta correlació positiva amb l'expressió del gen *EGFR*. Tanmateix, s'ha observat que la mesura de les *NA* correlaciona negativament amb la  $IC_{50}$  quan *EGFR* està mutat i hi ha evidències preclíniques d'un efecte sinèrgic de la combinació del nilotinib amb teràpies dirigides contra el gen *EGFR* (Smaglo et al., 2013). A més, corroborant la influència de diversos processos biològics en l'activitat de la xarxa, els fàrmacs identificats usant les *NA* no corresponen a les expressions més altes entre l'expressió gènica i la  $IC_{50}$ , Figura 9.3.1b-dreta.

## 9.4 Coherència amb les vies de senyalització i processos biològics

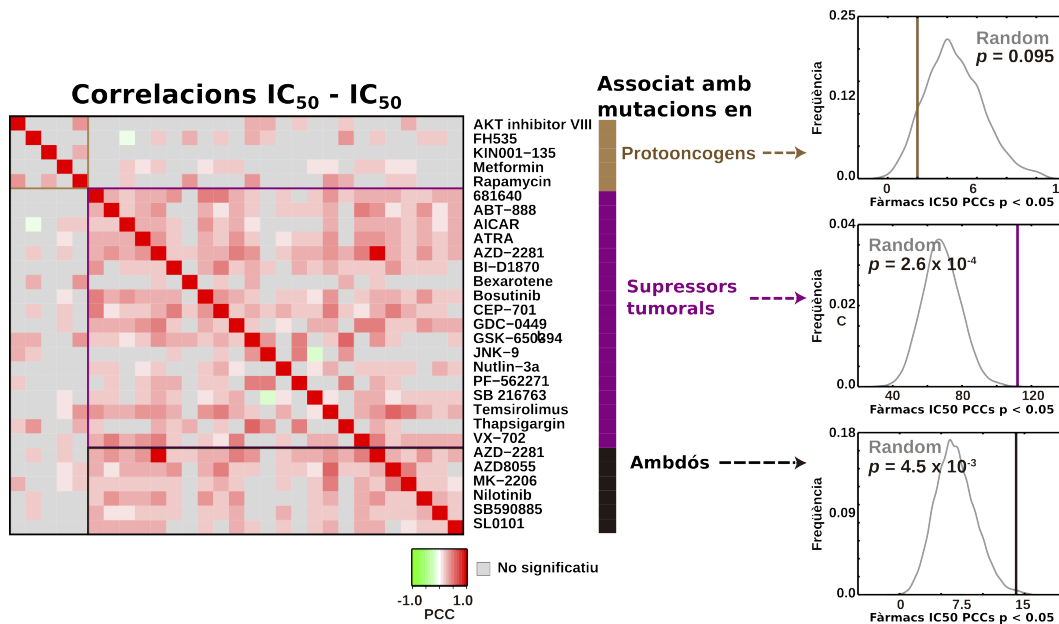
Les *NA* capturen informació entre les línies cel·lulars portadores de les mutacions *PIK3CA/R1* o *PTEN*. Tot i que no s'observen correlacions a través de l'estatus de mutació dels gens *PIK3CA/R1* i *PTEN* (coeficient de correlació de Pearson -0.05), s'han identificat dos fàrmacs associats a les mutacions dels dos gens.

- AZD-2281 (olaparib), que té com a diana PARP1/2, associada a la reparació de l'ADN,
- AZD-8055, que té com a diana mTORC1/2.

A més, els fàrmacs associats a l'estatus mutat dels protooncogenes tendeixen a alterar el procés biològic Metabolisme (test de Fisher  $p = 0.014$ ) mentre que els associats a l'estatus mutat dels gens supressors de tumors, tendeixen a alterar el procés biològic Transcripció ( $p = 0.001$ ).

### 9.4.1 Concordança entre els fàrmacs respecte de l'estatus de mutació dels gens

Els fàrmacs associats a mutacions dels gens de supressors de tumors tendeixen a estar correlacionats entre ells a través de les  $IC_{50}$  ( $p = 2.6 \times 10^{-4}$ ), mentre que els fàrmacs que tenen per diana els protooncogenes mostren un patró oposat ( $p = 0.95$ ), Figura 9.4.1.

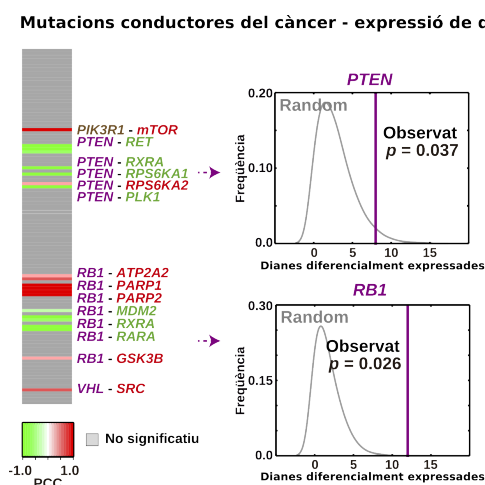


**Figura 9.4.1:** A l'esquerra, *heatmap* de les correlacions de Pearson entre les  $IC_{50}$  de fàrmacs associats (mitjançant *NA*) a l'estatus mutat dels gens conductors del càncer. A la dreta, s'ha generat una distribució del nombre de fàrmacs correlacionats, prenent 29 fàrmacs a l'atzar en cada iteració. La línia vertical mostra el nombre de fàrmacs correlacionats observats. Tal i com mostren els gràfics, els protooncogenes tendeixen a no està relacionats entre ells, mentre que els supressors de tumors sí.

### 9.4.2 Expressió de les dianes dels fàrmacs

L'anàlisi de les *NA* i les  $IC_{50}$  segons els estatus de mutació dels gens *PTEN* i *R1* captura el nombre més gran de fàrmacs associats a l'estat mutat d'aquests gens (correlació negativa entre *NA* i  $IC_{50}$ ), Taula G.0.1. Per veure quina activitat tenen les dianes dels fàrmacs identificats, s'han calculat les diferències d'expressió entre les línies cel·lulars portadores de la mutació i les no portadores. Els resultats mostren

que aquestes dianes tendeixen a estar diferencialment expressades, Figura 9.4.2



**Figura 9.4.2:** A l'esquerra, panell mostrant les dianes diferencialment expressades dels fàrmacs associats (mitjançant *NA*) a l'estat mutat dels gens conductors del càncer. El vermell indica sobreexpressió i el verd infraexpressió en els corresponent estatus de mutació del gen. A la dreta, el nombre de dianes diferencialment expressades identificades mitjançant l'estatus dels gens *PTEN* i *R1* és més alt que el nombre esperat. Els gràfics mostren les distribucions del nombre de dianes diferencialment expressades. Per construir la distribució, s'han agafat el mateix nombre de fàrmacs a l'atzar i s'han comptat la quantitat d'elements diferencialment expressats. La línia vertical indica la quantitat observada.

## 9.5 Predicció de sinergies mitjançant l'activitat

En els resultats mostrats fins ara, les mesures *NA* distingeixen tipus de fàrmacs segons les seves característiques i també segons l'estatus de mutació dels gens conductors del càncer. Així doncs, les mesures *NA* poden ser útils per identificar fàrmacs que maximitzin la pertorbació de la xarxa del càncer. Com que les mesures *NA* semblen capturar un efecte produït pels processos biològics, poden predir sinergies entre les combinacions dels fàrmacs identificats. Tanmateix, els fàrmacs identificats poden coincidir en els mecanismes cel·lulars als quals afecten. Per minimitzar aquest efecte, s'han considerat aquells fàrmacs tals que les seves  $IC_{50}$  no estan correlacionades. Amb aquesta hipòtesi, s'assumeix que dos fàrmacs amb perfils  $IC_{50}$  no relacionats, afecten a mecanismes diferents a través de les línies cel·lulars.

L'estudi s'ha centrat en la via de senyalització PI3KT-AKT, que té un rol important en la carcinogènesi i la resposta a les teràpies, en particular en el càncer de mama

(Lai et al., 2015). Aquesta via està associada a diverses característiques bàsiques del càncer, com el cycle cel·lular, la supervivència, la inestabilitat genòmica o l'angiogènesi (Fruman i Rommel, 2014) i, en els últims anys, ha emergit com a diana terapèutica de diversos estudis (Juvekar et al., 2012; Janku et al., 2012; Holohan et al., 2013). Un dels mecanismes d'activació d'aquesta via és mitjançant la mutació del protooncogen *PIK3CA*.

Per realitzar els experiments, s'han escollit quatre línies cel·lulars, dues portadores de la mutació *PIK3CA* i dues sense el gen mutat:

- portadores de la mutació: MDA-MB-453 i HCC1954,
- no portadores de la mutació: MDA-MB-231 i SKBR3

Per seleccionar els fàrmacs, s'han considerat aquells que estan associats a l'estatus mutat del gen *PIK3CA/R1* (identificats mitjançant la correlació entre les *NA* i les  $IC_{50}$ ) Taula G.0.1. D'entre aquests, s'han considerat els que les  $IC_{50}$  no estan correlacionades a través de les línies cel·lulars del càncer. Amb aquestes dues condicions, s'ha escollit la metformina (diana AMPK) per realitzar combinacions amb AZD-8055, olaparib o SL-0101 (amb dianes mTORC1/2, AURKB i PARP1/2 respectivament), ja que no està correlacionada amb cap d'ells.

Les combinacions de cadascun d'aquests fàrmacs amb la metformina revelen una sinergia significativa en 9 dels 12 experiments quan es comparen amb l'efecte dels fàrmacs per separat (Chou-Talalay index menor de 0.85), Figura 9.5.1a. S'han realitzat tres assajos independents per a cada combinació i s'han obtingut resultats equivalents. Per confirmar i comprovar la interacció detectada, s'ha ajustat la resposta de l'activitat cel·lular a les dosis dels fàrmacs mitjançant el mètode de Newton-Raphson de dos paràmetres lliures.

Per provar la hipòtesi de que els fàrmacs no correlacionats poden predir sinergies, s'han realitzat experiments amb les mateixes línies cel·lulars i fàrmacs però considerant les combinacions entre els fàrmacs amb les  $IC_{50}$  positivament correlacionades. Les combinacions d'aquests fàrmacs no han mostrat sinergia i 4 de 8 combinacions/línies cel·lulars han revelat un efecte antagonista, Figura 9.5.1b.

Finalment, com que la metformina i l'olaparib han estat aprovats per diverses patologies, s'ha analitzat la seva combinació en un assaig *ex vivo* de cèl·lules d'un xenoempelt derivat de pacients de càncer de mama (PDX) sensibles a la inhibició de PI3K (PDX-156) (Serra et al., 2013). S'han realitzat experiments amb dues ràtios entre les concentracions dels fàrmacs diferents, Figura 9.5.1c. Ambdues combina-

cions revelen un efecte de la reducció de la viabilitat cel·lular significatiu respecte dels tractaments individuals. Com a control, s'ha usat la línia cel·lular MCF10A de mama no cancerosa que no ha mostrat cap efecte en la combinació dels fàrmacs, Figura 9.5.1d.

Aquests resultats proven que la mesura *NA* és útil i viable en la identificació de combinacions de fàrmacs per la teràpia del càncer.



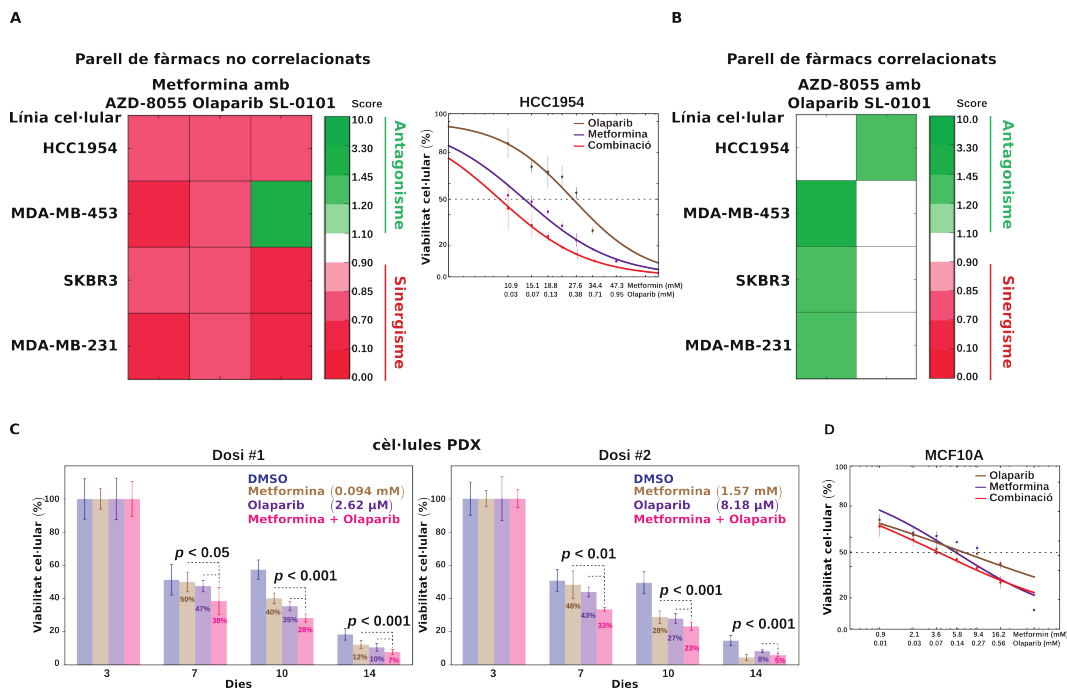


Figura 9.5.1:

- A l'esquerra, panell dels índexs Chou-Talalay obtinguts amb la combinació de la metformina i els fàrmacs no correlacionats en les quatre línies cel·lulars del càncer de mama, les dues superiors portadores de la mutació *PIK3CA/R1* mentre que les dues inferiors no. A la dreta, gràfic de les corbes de la resposta cel·lular a les dosis de fàrmac pel cas HCC1954 amb la combinació metformina i olaparib i individuals.
- Panell dels índexs Chou-Talalay obtinguts mitjançant combinacions del AZD-8055 amb fàrmacs positivament correlacionats.
- Gràfics del tractament amb metformina i/o olaparib de cultius cel·lulars de la PDX sensible a la inhibició de la senyalització de PI3K. Es mostren les diferències significatives entre el tractament amb la combinació de fàrmacs i els individuals i la corresponent resposta cel·lular. S'observen diferències en les dues dosis per a tots els dies excepte per la dosi #2 i el dia 14. Tanmateix, com que els percentatges de la viabilitat cel·lular són molt petits, les estimacions poden contenir un error relatiu gran.
- Gràfic de les corbes dels tractaments individuals i combinació de metformina i olaparib en les cèl·lules control MCF10A no canceroses.



Part V

Discussió



# 10

## Discussió I

### 10.1 Les proteïnes del càncer en l'interactoma

En els darrers anys, la tecnologia dels bioxips ha permès interrogar milers de gens alhora, obtenint una fotografia dels seus nivells i inferir possibles relacions entre la seva activitat. Tanmateix, encara no s'ha pogut interrogar tota la complexitat del sistema cel·lular. Les malalties complexes, com són els càncers, es desenvolupen i progressen mitjançant mecanismes que formen part de la complexitat cel·lular. Per entendre aquestes malalties, és necessari desxifrar la complexitat cel·lular i realitzar estudis dels diferents sistemes cel·lulars.

Mitjançant la integració de dades d'expressió gènica i d'interaccions proteïques, diversos estudis han vist que els gens associats al desenvolupament o progressió d'un càncer tendeixen a ocupar posicions més centrals en la xarxa de l'interactoma que altres gens no associats al càncer (Hernández et al., 2007; Wachi et al., 2005; Jonsson i Bates, 2006; Dezsó et al., 2009). Les proteïnes codificades del càncer tendeixen a tenir més interaccions que altres proteïnes, suggerint que poden influir en més processos biològics o mecanismes cel·lulars. A més de tenir més interaccions, aquestes proteïnes tendeixen a ocupar posicions rellevants entre els camins que uneixen dues proteïnes qualssevol en la xarxa de l'interactoma —és a dir, a tenir una centralitat *betweenness* més alta (Hernández et al., 2007)—, posicions que uneixen clústers (Jonsson i Bates, 2006) o posicions estables quan la xarxa és atacada eliminant interaccions (Wachi et al., 2005).

Tot i el nombre d'anàlisis integratius per aprofundir en el coneixement entre el càncer i els sistemes cel·lulars, encara s'està lluny d'obtenir un coneixement exhaustiu de com els sistemes cel·lulars funcionen i com els càncers i la seva dinàmica hi poden influir. A mesura que s'aconsegueixi descriure exhaustivament els diferents sistemes i

la complexitat del càncer, millor s'entendrà aquesta relació. Mentrestant, les anàlisis usant els recursos disponibles poden ajudar no sols avançar cap a aquest coneixement profund del càncer sinó també a identificar elements centrals pel seu diagnòstic, pronòstic o, inclús, tractament.

## 10.2 Anàlisi topològica de la xarxa de l'interactoma

L'estudi que s'ha presentat en aquest treball és un intent d'aprofundir en la complexitat cel·lular i entendre una mica millor el càncer a través de les xarxes complexes. El principal objectiu del treball és estudiar l'efecte que té el desenvolupament i progressió del càncer en la topologia de la xarxa de l'interactoma humà. Per dur a terme aquest objectiu, s'ha implementat un model de cascada de fallades (Wang i Rong, 2009). El mètode de cascada de fallades de Wang i Rong, modelitza la transmissió de càrrega elèctrica entre les centraletes de la xarxa elèctrica dels Estats Units per analitzar-ne la seva robustesa. Aquí, s'ha usat una lleugera modificació d'aquest model per modelitzar la transmissió de càrrega proteica en la xarxa de l'interactoma.

El model implementat defineix la càrrega i la capacitat dels vèrtexs en funció del nombre d'interactors que té una proteïna. Aquesta definició s'ha aplicat assumint que les proteïnes més centrals participen en més processos biològics o mecanismes cel·lulars i que un mal funcionament d'una proteïna central pot tenir un impacte major en el fenotip cel·lular (Barabasi i Oltvai, 2004; Barabasi et al., 2011; Goh et al., 2007). En aquest sentit, s'ha definit la càrrega de cada vèrtex en funció del seu grau, és a dir del nombre d'interaccions que té la proteïna que representa, i del grau total del seu veïnatge. Així, aquelles proteïnes que tenen més interaccions i que els seus veïns tenen més interaccions estan associades a vèrtexs amb una càrrega major.

Per definir la capacitat s'ha assumit que un gen amb una variabilitat més gran en diferents condicions reflecteix un gen que, en general, té uns nivells d'expressió més variables i que la cèl·lula permet més variabilitat en els nivells del seu producte proteic. Per reflectir aquest comportament, s'ha introduït el paràmetre  $r$  en la capacitat dels vèrtexs. D'aquesta manera, aquelles proteïnes amb nivells més variables es tradueixen en vèrtexs amb més capacitat. Cal observar però, que s'assumeix que els nivells del producte gènic expressat i els nivells de proteïna estan positivament correlacionats (Tian et al., 2004; Maier et al., 2009). Cal tenir en compte també

que el paràmetre  $r$  pot capturar el soroll de cada experiment a l'hora de mesurar els nivells d'expressió gènica i introduir-los en el mètode.

Aquest model de cascada de fallades es genera a partir d'un vèrtex inicial, o un conjunt de vèrtexs inicials, que és forçat a transmetre la seva càrrega i se'l suprimeix de la xarxa, generant la cascada de fallades. Els seus veïns absorbeixen la seva càrrega proporcionalment segons la seva càrrega en el veïnatge. Si l'increment que pateixen és superior a la capacitat, llavors pateixen una fallada i distribueixen la càrrega als seus veïns. I així successivament fins que la xarxa assoleix un estat estacionari en el qual ja no es produeixen més fallades.

Per analitzar la topologia de la xarxa respecte de la dinàmica del càncer, s'han considerat les proteïnes del càncer com a conjunts inicials. Tot seguit s'han calculat les fallades en la xarxa de l'interactoma respecte d'aquests conjunts i s'han analitzat els patrons que han generat. Anàlogament, s'han analitzat els patrons de fallades en els estadis del càncer subsegüents a l'estadi inicial respecte d'aquest mateix. Tots els resultats han revelat patrons de robustesa de la xarxa. Aquesta robustesa suggereix que l'impacte de les proteïnes del càncer en la xarxa de l'interactoma proteic ve donat per un nombre de fallades menor que el nombre de fallades esperat per atzar. En aquest sentit, es pot pensar que les posicions que ocupen les proteïnes del càncer en la xarxa de l'interactoma és tal que la mateixa topologia de la xarxa atenua l'impacte que puguin generar. A més, les proteïnes del càncer solen ocupar posicions robustes de la xarxa, si més no respecte de l'estadi inicial del càncer de mama.

Aquesta conclusió sembla estar reforçada pels resultats que revelen una topologia més feble respecte aquelles proteïnes associades a mecanismes cel·lulars que es poden pensar com a contraris al càncer, com per exemple les dues vies identificades en les anàlisis: regulació de l'autofàgia i SNARE *interactions vesicular transport* (Moreau et al., 2013; Kuballa et al., 2012; Bindea et al., 2010). L'autofàgia és un mecanisme de mort cel·lular que es pot activar quan es detecten danys importants en una cèl·lula —per exemple mutacions gèniques que tenen un efecte danyós en els mecanismes de replicació de l'ADN. Ambdues vies estan relacionades mitjançant les proteïnes de la família SNARE que estan involucrades en l'autofàgia (Moreau et al., 2013). En aquest sentit, les vies identificades relacionades amb el sistema immunològic o el procés biològic de la resposta immune que generen patrons de robustesa oposats als del càncer, també reforcen la conclusió de la robustesa de la xarxa respecte de les proteïnes del càncer. Si bé el rol del sistema immunològic respecte del càncer

és un tema de controvèrsia, segons si té un paper supressor tumoral o estimula les cèl·lules canceroses (Hanahan i Weinberg, 2011; Bindea et al., 2010), sí que és cert que, en algunes de les seves formes, té un paper protector o supressor.

Per esbrinar quines propietats topològiques podrien explicar els patrons trobats respecte de les proteïnes considerades, s'han analitzat el nombre de fallades respecte de conjunts enriquits amb elements a una distància  $d \in \{2, 4, 6\}$ . Els resultats mostren que per  $d \in \{2, 4\}$ , el nombre de fallades observades és menor que les esperades per atzar. Pel cas  $d = 6$ , el patró s'inverteix i el nombre de fallades és major que les esperades per atzar. Tanmateix, algunes de les vies de senyalització analitzades mostren patrons contraris, una distància petita i un nombre de fallades generades més grans que les esperades per atzar, com per exemple la via *SNARE interactions in vesicular transport*. Així, la distància entre vèrtexs podria capturar parcialment les causes dels patrons observats.

### 10.3 La xarxa de l'interactoma

Per construir la xarxa de l'interactoma s'han usat les dades del repositori *Human Protein Reference Database* i *IntAct Molecular Interaction Database*. Aquestes bases de dades recopilen les interaccions proteïques a través de la literatura científica, és a dir, amb alguna publicació que descriu el mètode experimental per identificar la interacció. El fet, doncs, que cada interacció tingui una publicació experimental és una manera de reduir els falsos positius —en contrast, per exemple, d'aquelles que s'infereixen *in silico*. Tanmateix, la construcció de les bases de dades mitjançant la recopilació introdueixen un altre tipus de biaix. L'interès per l'anàlisi de certs gens i proteïnes pot esbiaixar la distribució d'interaccions recopilades, descrivint una proporció d'interaccions més grans per aquests gens/proteïnes que per la resta (Rolland et al., 2014). Aquest biaix es pot reflectir en la representació de la xarxa de l'interactoma, concretament en la seva topologia. Per exemple, podria passar que alguns gens poc centrals en la xarxa de l'interactoma representada fossin més centrals del que s'han descrit. Algunes anàlisis han provat el biaix d'aquest tipus de bases de dades escollint metòdicament mostres en l'espai de l'interactoma (Rolland et al., 2014).

En la xarxa de l'interactoma cada vèrtex correspon a un gen/proteïna. És a dir, s'ha identificat el producte proteic d'un gen en un sol element. Els gens però, poden expressar-se en diferents isoformes d'una proteïna —per exemple, en el moment del



tall i unió dels exons i introns es poden prendre combinacions d'exons diferents per la transcripció, transcrivint diferents seqüències d'ARN missatger i traduint-se en cadenes d'aminoàcids diferents. Així, dues isoformes proteiques diferents codificades pel mateix gen poden tenir una estructura diferent i, per tant, poden interaccionar amb proteïnes diferents i participar en processos biològics o mecanismes diferents (Corominas et al., 2014). Com que la diversitat de les isoformes i les seves interaccions és una informació que encara no es coneix, les interaccions descrites en la xarxa de l'interactoma corresponen a una, o poques, isoformes del producte gènic. L'omissió d'aquesta informació segurament restringeix la construcció de la xarxa, la seva topologia i, per tant, els resultats que se'n puguin inferir.

Una altra consideració respecte de la construcció de la xarxa de l'interactoma, és la localització de l'expressió de les proteïnes. Els processos biològics poden produir-se a localitzacions subcel·lulars diferents. La localització d'un procés pot estar determinada pel mateix procés —per exemple la replicació de l'ADN es produeix al nucli cel·lular— i és una manera de regular el procés en si mateix i els elements que hi intervenen. Les proteïnes poden expressar-se a diferents localitzacions cel·lulars i, segons aquesta localització, poden tenir funcions diferents. Les bases de dades usades per construir la xarxa de l'interactoma, ni contenen tota la informació de la localització de les proteïnes ni les interaccions estan determinades tenint en compte la localització a on podrien expressar-se. D'aquesta manera, una interacció entre dues proteïnes detectada experimentalment podria no tenir sentit biològic i que en la cèl·lula no es produís (Veres et al., 2014). Així, la xarxa de l'interactoma construïda s'ha d'entendre com una recopilació d'interaccions possibles respecte de l'estructura de les proteïnes, independentment de la seva localització i condicions biològiques. Aquest fet també pot condicionar l'estructura de la xarxa ja que algunes interaccions podrien no produir-se.

Les interaccions usades per la construcció de la xarxa de l'interactoma tampoc conté informació del moment en que es produeix. És a dir, entre totes les interaccions no es coneix quines poden precedir quines. Aquest podria ser un factor important perquè quan dues proteïnes interaccionen poden canviar la seva estructura i limitar les interaccions posteriors. A més a més, tampoc se sap en quina direcció es transmet la informació quan es produeix la interacció, és a dir que només es representa l'existència de la interacció.

Totes aquestes observacions poden limitar les anàlisis com les que s'han presentat en aquest treball. Com que la topologia de la xarxa de l'interactoma podria ser

---

diferent de la representació que es coneix actualment, les anàlisis podrien donar resultats diferents. Tanmateix, les dades actuals semblen capturar informació real, ja que diversos estudis que integren dades de diferents nivells del sistema han resultat ser útils per predir gens específics de malalties o tractaments (Dezso et al., 2009; Barabasi et al., 2011).

# 11

## Discussió II

### 11.1 Activitat de la xarxa

Per a poder combatre les malalties malignes complexes com els càncers, és necessari entendre la complexitat cel·lular, com les vies de senyalització formen estructures complexes i com s'alteren respecte de la dinàmica del càncer i de les teràpies. Comprendre com es desregulen i s'orquestren les vies que governen els mecanismes de la cèl·lula cancerosa i com els seus patrons d'activitat responen a les teràpies, és necessari per combatre el càncer i trobar tractaments beneficiosos per als pacients, reduint-ne la toxicitat i els efectes secundaris (Kreeger i Lauffenburger, 2010; Holohan et al., 2013). Com que el desenvolupament i progressió d'un càncer depèn dels patrons complexos de les vies de senyalització i dels processos biològics, la comprensió de la dinàmica del càncer i la resposta terapèutica de les cèl·lules canceroses pot millorar-se integrant dades de diferents nivells, com el molecular, per reflectir l'activitat de les molècules que influeixen en les vies o processos biològics (Werner et al., 2014; Kreeger i Lauffenburger, 2010).

En aquest estudi, s'han integrat les dades d'expressió gènica corresponents a diverses línies cel·lulars representant la diversitat de tots els càncers humans (Garnett et al., 2012) amb les dades de l'interactoma proteic humà (Keshava Prasad et al., 2009) i s'ha definit una mesura per capturar l'activitat de la xarxa de l'interactoma respecte de cadascuna de les línies cel·lulars canceroses. La mesura de l'activitat de la xarxa s'ha basat en la centralitat subgraf i l'índex d'Estrada de xarxes complexes (Estrada i Rodríguez-Velázquez, 2005; Estrada, 2000). A diferència d'altres mesures que només es basen en els camins més curts, la centralitat subgraf d'un vèrtex en una xarxa està basada en tots els camins tancats amb origen al vèrtex considerat, penalitzant aquells camins més llargs assumint que tenen menys influència. En aquest treball s'ha construït la matriu de pesos mitjançant l'expressió dels gens associats a cada

vèrtex. L'expressió d'un gen s'ha distribuït entre els vèrtexs veïns corresponents, considerant també l'expressió dels veïns. S'ha calculat la centralitat subgraf de cada vèrtex usant la matriu d'adjacència amb pesos. Aquesta mesura representa l'activitat de tots els camins tancats amb origen els vèrtexs de la xarxa. L'activitat de la xarxa s'ha definit a partir de la centralitat subgraf dels vèrtexs. Les mesures obtingudes (*NA*) per a cada línia cel·lular s'han comparat amb les dosis de fàrmac necessàries per reduir l'activitat cel·lular a la meitat ( $IC_{50}$ ) de més d'un centenar de fàrmacs diferents, per capturar patrons entre l'activitat de la xarxa i la sensibilitat cel·lular (Garnett et al., 2012). Comparant l'activitat de la xarxa amb les  $IC_{50}$  dels fàrmacs, a través de les línies cel·lulars dels càncers, s'ha observat que la mesura de l'activitat dels càncers distingeix els fàrmacs segons les seves característiques —per exemple si ataquen a una diana o diverses dianes. S'ha observat que moltes d'aquestes diferències s'expliquen per la quantitat de dianes que afecta el fàrmac, com és el cas de les famílies de proteïnes tirosina-cinases (CTKs) i serina/treonina cinases (STKs), que estan relacionades a la transducció de senyals que regulen la majoria de vies de senyalització cel·lulars, essent la família STKs major que la CTKs. Tanmateix, aquestes diferències també es podrien explicar pels tipus de processos biològics o vies de senyalització involucrats.

Les mesures *NA* també capturen els processos biològics associats als càncers. Com que la mesura de l'activitat de la xarxa està basada en la centralitat o activitat de cada gen, aquesta anàlisi s'ha realitzat respecte dels gens més centrals o actius de la xarxa. Considerant només aquests gens, poden descartar-se possibles processos biològics que podrien està presents mitjançant l'aportació addicional dels gens menys actius. S'han identificat doncs, els processos biològics associats als gens més actius, entre d'altres, s'han identificat processos associats a l'apoptosi i expressió gènica, relacionats amb la tumorigènesi (Hanahan i Weinberg, 2011).

Com que aquestes anàlisis suggereixen que les mesures *NA* capturen informació dels fàrmacs en funció de les seves característiques i processos biològics afectats, també s'han usat per capturar l'efecte dels fàrmacs respecte de l'estat de mutació de gens conductors del càncer. S'han identificat doncs, aquells fàrmacs que tenien un impacte major en les cèl·lules portadores dels gens mutats que en les portadores dels gens no mutats, és a dir correlacions més negatives enter les  $NA-IC_{50}$  per les línies mutades. S'han considerat els fàrmacs amb coeficient de correlació negatiu entre les línies cel·lulars portadores dels gens mutats més petit que qualsevol coeficient de les línies no portadores dels gens mutats, Equació 9.1. S'ha escollit aquest

criteri tant conservador per capturar els casos més extrems. En aquesta anàlisi, integrant dades a nivell molecular, el comportament de les mesures *NA* concorda amb el comportament de l'expressió, encara que les mesures *NA* semblen capturar més informació. La majoria dels fàrmacs identificats amb la correlació negativa entre les *NA* i les  $IC_{50}$  mostren un patró similar respecte de l'expressió gènica. Tanmateix, les mesures *NA* semblen capturar altres mecanismes o processos biològics que no es capturen a través de l'expressió, capturant un comportament més global dels gens conductors del càncer. En aquest mateix sentit, els fàrmacs identificats tendeixen a tenir patrons similars de les  $IC_{50}$  respecte del tipus de gen. En el cas dels supressors de tumors, les  $IC_{50}$  tendeixen a està correlacionades mentre que en el cas dels protooncogens tendeixen a no estar-ho.

Així doncs, la mesura *NA* de l'activitat de la xarxa captura els patrons d'activitat genètica associats als càncers a través de la xarxa molecular d'interaccions proteiques. Tanmateix, aquest estudi podria estar limitat pel tipus de dades d'expressió gènica usades. Aquestes dades contenen l'expressió basal de cada gen en cada línia cel·lular, és a dir un sol punt d'expressió per a cada línia. Si enlloc de dades d'expressió basal en els càncers, s'usessin dades quantitatives entre condicions diferents —per exemple entre teixit normal i cancerós o tractat—, la integració de dades podria reflectir amb més fidelitat l'activitat gènica en les condicions desitjades, ja que es podrien detectar canvis significatius en els nivells d'expressió gènica, independentment de si són grans o petits. Tanmateix, les dades representen adequadament els nivells de l'expressió gènica ja que considerant una gran quantitat de línies cel·lulars de diferents càncers capturen la variabilitat en els seus nivells.

La construcció de la xarxa també pot influir en els resultats. La topologia de la xarxa pot està esbiaixada pel fet de que està construïda recopilant interaccions de la literatura científica. A més, el fet de col·lapsar totes les isoformes d'un gen en un sol vèrtex, modifica lleugerament la topologia de la xarxa ja que el vèrtex representatiu té totes les interaccions de les isoformes, mentre que no totes les isoformes han d'interaccionar amb els mateixos elements. A aquestes dues observacions, cal afegir-hi la localització subcel·lular de les proteïnes. La xarxa de l'interactoma està formada per a totes les interaccions possibles observades experimentalment. Tanmateix, a causa de la localització de la proteïna en la cèl·lula, algunes d'aquestes interaccions podrien no tenir sentit biològic (Rolland et al., 2014).

## 11.2 Combinació de fàrmacs

Les cèl·lules canceroses solen mostrar resistència a les teràpies de fàrmacs, ja sigui intrínseca —a causa de l'heterogeneïtat de la cèl·lula cancerosa— o adquirida. Una manera de combatre la resistència de la cèl·lula cancerosa és mitjançant la combinació de fàrmacs que alterin els mecanismes necessaris per induir la cèl·lula maligna a la mort cel·lular. L'objectiu de la combinació de fàrmacs, és trobar teràpies amb un efecte additiu o sinèrgic dirigits als mecanismes cel·lulars a través de dianes específiques, produint la menor toxicitat i efectes secundaris possibles. Les combinacions de fàrmacs amb un efecte sinèrgic poden resultar més específiques en un context cel·lular particular, tanmateix és un problema complex, ja que també pot provocar sinergia dels efectes secundaris (Lehar et al., 2009). Per la gran quantitat de fàrmacs que hi ha, aprovats o en desenvolupament, estudiar-ne les combinacions entre dos, tres, etc. no és un problema fàcil. Una manera de reduir dràsticament el nombre de possibles combinacions és mitjançant la integració de dades de diferents nivells del sistema cel·lular. La integració de dades d'expressió gènica, processos biològics, vies de senyalització, xarxes metabòliques, etc. pot ser útil per identificar aquells mecanismes que resulten vitals per la cèl·lula cancerosa o pels quals adquireix resistència. En aquest sentit, s'han començat projectes en la comunitat científica per predir l'activitat de combinacions mitjançant mètodes computacionals (Bansal et al., 2014). Per poder avançar en aquest camp però, cal desenvolupar més projectes i crear repositoris amb dades de teràpies dirigides a dianes específiques i els fenotips de les cèl·lules canceroses. En aquest marc, la mesura de l'activitat de la xarxa *NA* definida en aquest estudi podria ser útil per identificar combinacions amb un caràcter sinèrgic.

En les anàlisis realitzades entre les mesures *NA* i les  $IC_{50}$  dels fàrmacs respecte de l'estatus de mutació dels gens conductors del càncer, s'han identificat fàrmacs amb un comportament diferenciat entre les línies portadores de mutacions respecte de les que no. Tenint en compte això i que les *NA* capturen informació de l'activitat de les línies cel·lulars del càncer, les mesures *NA* podrien identificar fàrmacs que maximitzessin la pertorbació en les xarxes del càncer. Com que les *NA* també capturen processos biològics i mecanismes involucrats en les xarxes del càncer, la combinació dels fàrmacs identificats mitjançant les *NA* podrien generar un efecte sinèrgic.

Per comprovar aquesta hipòtesi, s'han escollit quatre fàrmacs per estudiar la respos-

ta cel·lular respecte de les seves combinacions. Basat en els resultats, l'estudi s'ha centrat en l'activitat de la via PI3K-AKT. Els fàrmacs s'han escollit segons l'estatus de mutació del protooncogen *PIK3CA/R1*, activador de la via PI3K-AKT, però també maximitzant els seus efectes mitjançant l'elecció d'aquells fàrmacs relacionats a diferents mecanismes. Per detectar els fàrmacs amb mecanismes diferents s'han calculat les correlacions entre les  $IC_{50}$  dels fàrmacs a través de les línies cel·lulars del càncer, assumint que la resposta de les línies cel·lulars al tractament ve determinada pels mecanismes concrets associats al càncer de cada línia cel·lular, i s'han considerat aquells fàrmacs amb la  $IC_{50}$  no correlacionada. Els fàrmacs identificats estan associats a les línies mutades (*PIK3CA/R1*) i tenen mecanismes d'acció diferents. Així, la combinació de fàrmacs amb aquestes dues condicions poden generar un efecte sinèrgic en la cèl·lula.

Un valor afegit a l'hora d'escollir la via PI3K-AKT-mTOR, a més dels resultats obtinguts, és que aquesta via està associada a les característiques bàsiques del càncer, com per exemple cicle cel·lular o a la inestabilitat genòmica. Aquesta via és una de les més alterades en els càncers i la seva activació, ja sigui per causes directes com la mutació d'algun dels seus elements o per causes indirectes, ha estat associada a diversos càncers. En els darrers anys, aquesta via s'ha consolidat com a diana terapèutica contra als càncers (Fruman i Rommel, 2014). A més, els tractaments contra al càncer inhibint aquesta via han augmentat l'efectivitat respecte de tractaments estàndards en pacients de càncer de mama (Janku et al., 2012) i també s'ha observat un efecte sinèrgic entre fàrmacs inhibidors d'aquesta via i de la diana PARP (Juvekar et al., 2012).

Amb aquestes consideracions, s'han estudiat les combinacions entre quatre fàrmacs obtinguts en quatre línies cel·lulars, dues portadores dels gens *PI3KCA/R1* mutats i dues no portadores dels gens mutats. Comparant els efectes de les dosis pels experiments amb els fàrmacs individuals i amb la combinació, s'ha capturat un efecte sinèrgic en la majoria de les combinacions. Per provar la hipòtesi de que les combinacions de fàrmacs amb  $IC_{50}$  no correlacionades suggereixen la implicació de mecanismes diferents, s'han estudiat les combinacions dels fàrmacs amb correlacions entre les  $NA$  i les  $IC_{50}$  positives, suggerint la involucració de mecanismes d'acció similars, i s'ha obtingut que cap de les combinacions genera sinergia mentre que quatre de les vuit combinacions generen antagonisme.

Finalment, com que la metformina i l'olaparib són fàrmacs aprovats individualment, per diverses patologies, s'ha comprovat l'efecte de la seva combinació mitjançant

assaigs *ex vivo* en cèl·lules sensibles a la inhibició de PI3K. Els resultats mostren un efecte sinèrgic de la combinació dels fàrmacs considerats.

Algunes consideracions per aquest anàlisi s'haurien de tenir en compte. Com abans, el tipus de dades d'expressió gènica i per la construcció de la xarxa podrien condicionar aquests resultats. Amb unes dades de fàrmacs més acurada també es podrien obtenir resultats més acurtats. Per exemple, tot i que les  $IC_{50}$  a través de les línies cel·lulars del càncer capturen els mecanismes d'acció dels fàrmacs, no hi ha informació de quins mecanismes són aquests. Ampliant aquesta informació, es podrien acotar el nombre de combinacions específiques atacant mecanismes concrets en les cèl·lules canceroses. Tanmateix, si es tingués aquesta informació tampoc es podrien descartar aquells fàrmacs amb mecanismes semblants, ja que, per exemple, dos fàrmacs dirigits a una mateixa via amb dianes diferents, coincidint en la inhibició de la via comuna però amb efectes col·laterals diferents en la intricada xarxa que formen les vies de senyalització, podrien generar també efectes sinèrgics. Per dur estudis a aquest nivell de detall però, calen repositoris que continguin informació d'experiments que es puguin consultar i compartir.

Amb tot, la mesura de la xarxa definida sembla capturar els patrons d'activitat gènica a nivell del sistema molecular que separa fàrmacs respecte de les seves característiques i respecte de l'estatus de mutacions dels gens conductors del càncer. Combinant aquesta informació i la  $IC_{50}$  ha estat útil per detectar sinergies entre fàrmacs.

Una altra consideració a tenir en compte per detectar sinergies, és el model usat per mesurar l'efecte de la combinació. Els models emprats deriven de l'equació *median-effect* (Chou, 2006) i tenen una expressió algebraica per ajustar les respostes respecte de les dosis (Chou, 2006). En aquest model, s'assumeix que la resposta és monòtona i sigmoïdal. Tanmateix, les respostes cel·lulars a fàrmacs poden tenir comportaments inesperats en el seu creixement o en el comportament sigmoïdal. Per capturar comportaments més generals és necessari desenvolupar altres models no paramètrics per reflectir millor les respostes i no discriminar informació biològica que podria ser significativa (Zhang et al., 2013; Boik et al., 2008).

Amb tot, la mesura de la xarxa definida captura patrons d'activitat gènica a nivell del sistema molecular que separa els fàrmacs respecte de les seves característiques i de l'estatus de mutació dels gens conductors del càncer. Combinant aquesta informació i la  $IC_{50}$  ha estat útil per detectar sinergies entre fàrmacs. És necessari



realitzar aquest tipus de d'estudis, integrant diferents tipus de dades, per identificar possibles tractaments del càncer augmentant la seva efectivitat, especificitat i reduint la toxicitat.



**Part VI**

**Conclusions**



# 12

## Conclusions I

En aquest estudi, s'han integrat dades d'expressió gènica associades al càncer de mama i colorectal en la xarxa de l'interactoma proteic humà per analitzar l'impacte dels canvis d'expressió a través de la topologia de l'interactoma i la seva robustesa. S'ha analitzat la robustesa de l'interactoma respecte dels estadis del càncer, la pertorbació de vèrtexs en la xarxa de l'interactoma respecte de la pertorbació gènica i del tractament. Amb les anàlisis realitzades, es conclou que

En general, la xarxa mostra robustesa associada al càncer:

- la xarxa de l'interactoma és robusta respecte als estadis del càncer,
- els conjunts de proteïnes dels estadis del càncer en la xarxa de l'interactoma són robusts respecte l'estadi inicial del càncer,
- els conjunts de proteïnes dels estadis del càncer en la xarxa de l'interactoma són robusts respecte del tractament,
- els conjunts de proteïnes dels estadis del càncer en la xarxa de l'interactoma són robusts respecte de la pertorbació del gen *SNCA* central i desregulat en el càncer.

I no mostra robustesa en mecanismes oposats al càncer:

- la xarxa de l'interactoma no és robusta respecte de les proteïnes involucrades en el procés biològic resposta immune,
- la xarxa de l'interactoma no és robusta respecte de les proteïnes involucrades en el procés biològic autofàgia.



# 13

## Conclusions II

En aquesta part de la tesis, s'ha definit una mesura de l'activitat de la xarxa dels càncers integrant dades d'interaccions proteiques, d'expressió gènica i de resposta a fàrmacs. Mitjançant les anàlisis realitzades, es pot concloure que

- la mesura definida captura l'activitat en la xarxa de l'interactoma respecte dels càncers i els mecanismes involucrats,
- la mesura definida captura l'impacte dels fàrmacs en la xarxa de l'interactoma respecte dels càncers segons les seves característiques,
- la mesura definida captura l'impacte dels fàrmacs en la xarxa de l'interactoma respecte dels càncers segons de l'estatus de mutació dels gens conductors del càncer,
- la mesura definida captura els efectes dels fàrmacs en la xarxa de l'interactoma respecte dels càncers segons de l'estatus de mutació dels gens conductors del càncer que combinats generen sinergia.





**Part VII**

**Bibliografia**



## Bibliografia

Mark D. Adams, Susan E. Celniker, Robert A. Holt, Cheryl A. Evans, Jeannine D. Gocayne, Peter G. Amanatides, Steven E. Scherer, Peter W. Li, Roger A. Hoskins, Richard F. Galle, Reed A. George, Suzanna E. Lewis, Stephen Richards, Michael Ashburner, Scott N. Henderson, Granger G. Sutton, Jennifer R. Wortman, Mark D. Yandell, Qing Zhang, Lin X. Chen, Rhonda C. Brandon, Yu-Hui C. Rogers, Robert G. Blazej, Mark Champe, Barret D. Pfeiffer, Kenneth H. Wan, Clare Doyle, Evan G. Baxter, Gregg Helt, Catherine R. Nelson, George L. Gabor, Miklos, Josep F. Abril, Anna Agbayani, Hui-Jin An, Cynthia Andrews-Pfannkoch, Danita Baldwin, Richard M. Ballew, Anand Basu, James Baxendale, Leyla Bayraktaroglu, Ellen M. Beasley, Karen Y. Beeson, P. V. Benos, Benjamin P. Berman, Deepali Bhandari, Slava Bolshakov, Dana Borkova, Michael R. Botchan, John Bouck, Peter Brokstein, Phillipe Brottier, Kenneth C. Burtis, Dana A. Busam, Heather Butler, Edouard Cadieu, Angela Center, Ishwar Chandra, J. Michael Cherry, Simon Cawley, Carl Dahlke, Lionel B. Davenport, Peter Davies, Beatriz de Pablos, Arthur Delcher, Zuoming Deng, Anne Deslattes Mays, Ian Dew, Suzanne M. Dietz, Kristina Dodson, Lisa E. Doup, Michael Downes, Shannon Dugan-Rocha, Boris C. Dunkov, Patrick Dunn, Kenneth J. Durbin, Carlos C. Evangelista, Concepcion Ferraz, Steven Ferriera, Wolfgang Fleischmann, Carl Foster, Andrei E. Gabrielian, Neha S. Garg, William M. Gelbart, Ken Glasser, Anna Glodek, Fangcheng Gong, J. Harley Gorrell, Zhiping Gu, Ping Guan, Michael Harris, Nomi L. Harris, Damon Harvey, Thomas J. Heiman, Judith R. Hernandez, Jarrett Houck, Damon Hostin, Kathryn A. Houston, Timothy J. Howland, Ming-Hui Wei, Chinyere Ibegwam, Mena Jalali, Francis Kalush, Gary H. Karpen, Zhaoxi Ke, James A. Kennison, Karen A. Ketchum, Bruce E. Kimmel, Chinnappa D. Kodira, Cheryl Kraft, Saul Kravitz, David Kulp, Zhongwu Lai, Paul Lasko, Yiding Lei, Alexander A. Levitsky, Jiayin Li, Zhenya Li, Yong Liang, Xiaoying Lin, Xiangjun Liu, Bettina Mattei, Tina C. McIntosh, Michael P. McLeod, Duncan McPherson, Gennady Merkulov, Natalia V. Milshina, Clark Mobarry, Joe Morris, Ali Moshrefi, Stephen M. Mount, Mee Moy, Brian Murphy, Lee Murphy, Donna M. Muzny, David L. Nelson, David R. Nelson, Keith A. Nelson, Katherine Nixon, Deborah R. Nusskern, Joanne M. Pacleb, Michael Palazzolo, Gjange S. Pittman, Sue Pan, John Pollard, Vinita Puri, Martin G. Reese, Knut Reinert, Karin Remington, Robert D. C. Saunders, Frederick Scheeler, Hua Shen, Bixiang Christopher Shue, Inga Sidén-Kiamos, Michael Simpson, Marian P. Skupski, Tom Smith, Eugene Spier, Allan C. Spradling, Mark Stapleton, Renee Strong, Eric Sun, Robert Svirskas, Cyndee Tector, Russell Turner, Eli Venter, Aihui H.

- Wang, Xin Wang, Zhen-Yuan Wang, David A. Wassarman, George M. Weinstock, Jean Weissenbach, Sherita M. Williams, Trevor Woodage, Kim C. Worley, David Wu, Song Yang, Q. Alison Yao, Jane Ye, Ru-Fang Yeh, Jayshree S. Zaveri, Ming Zhan, Guangren Zhang, Qi Zhao, Liansheng Zheng, Xiangqun H. Zheng, Fei N. Zhong, Wenyan Zhong, Xiaojun Zhou, Shiaoping Zhu, Xiaohong Zhu, Hamilton O. Smith, Richard A. Gibbs, Eugene W. Myers, Gerald M. Rubin, i J. Craig Venter. The genome sequence of *Drosophila melanogaster*. *Science*, 287(5461): 2185–2195, 2000. doi: 10.1126/science.287.5461.2185.
- Réka Albert i Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, Jan 2002. doi: 10.1103/RevModPhys.74.47. URL <http://link.aps.org/doi/10.1103/RevModPhys.74.47>.
- Reka Albert, Hawoong Jeong, i Albert-Laszlo Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, July 2000. ISSN 1476-4687. doi: 10.1038/35019019.
- B. alberts. The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell*, 92(3):291–294, February 1998. ISSN 00928674. doi: 10.1016/s0092-8674(00)80922-8. URL [http://dx.doi.org/10.1016/s0092-8674\(00\)80922-8](http://dx.doi.org/10.1016/s0092-8674(00)80922-8).
- Ash A. Alizadeh, Michael B. Eisen, R. Eric Davis, Chi Ma, Izidore S. Lossos, Andreas Rosenwald, Jennifer C. Boldrick, Hajeer Sabet, Truc Tran, Xin Yu, John I. Powell, Liming Yang, Gerald E. Marti, Troy Moore, James Hudson, Lisheng Lu, David B. Lewis, Robert Tibshirani, Gavin Sherlock, Wing C. Chan, Timothy C. Greiner, Dennis D. Weisenburger, James O. Armitage, Roger Warnke, Ronald Levy, Wyndham Wilson, Michael R. Grever, John C. Byrd, David Botstein, Patrick O. Brown, i Louis M. Staudt. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769):503–511, February 2000. ISSN 0028-0836. doi: 10.1038/35000501. URL <http://dx.doi.org/10.1038/35000501>.
- Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Karsarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, i Gavin Sherlock. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29, May 2000. ISSN 1061-4036. doi: 10.1038/75556. URL <http://dx.doi.org/10.1038/75556>.
- Mukesh Bansal, Vincenzo Belcastro, Alberto Ambesi-Impiombato, i Diego di Bernardo. How to infer gene networks from expression profiles. *Molecular Systems Biology*, 3(1), 2007. ISSN 1744-4292. doi: 10.1038/msb4100158.
- Mukesh Bansal, Jichen Yang, Charles Karan, Michael P. Menden, James C. Costello, Hao Tang, Guanghua Xiao, Yajuan Li, Jeffrey Allen, Rui Zhong, Beibei Chen, Minsoo Kim, Tao Wang, Laura M. Heiser, Ronald Realubit, Michela Mattioli, Mariano J. Alvarez, Yao Shen, Nci-Dream Community, Daniel Gallahan, Dinah

- Singer, Julio Saez-Rodriguez, Yang Xie, Gustavo Stolovitzky, i Andrea Califano. A community computational challenge to predict the activity of pairs of compounds. *Nature Biotechnology*, 32(12):1213–1222, November 2014. ISSN 1087-0156. doi: 10.1038/nbt.3052. URL <http://dx.doi.org/10.1038/nbt.3052>.
- Albert-Laszlo Barabasi i Zoltan N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, February 2004. ISSN 1471-0056. doi: 10.1038/nrg1272. URL <http://dx.doi.org/10.1038/nrg1272>.
- Albert-Laszlo Barabasi, Natali Gulbahce, i Joseph Loscalzo. Network medicine: a network-based approach to human disease. *Nat Rev Genet*, 12(1):56–68, 2011. ISSN 1471-0056. doi: 10.1038/nrg2918. URL <http://dx.doi.org/10.1038/nrg2918>.
- Albert-László Barabási i Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. doi: 10.1126/science.286.5439.509. URL <http://www.sciencemag.org/content/286/5439/509.abstract>.
- Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L. Robertson, Nadezhda Serova, Sean Davis, i Alexandra Soboleva. Ncbi geo: archive for functional genomics data sets—update. *Nucleic Acids Research*, 41(D1):D991–D995, 2013. doi: 10.1093/nar/gks1193.
- Tatiana Benaglia, Didier Chauveau, David R. Hunter, i Derek Young. mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software*, 32(6):1–29, 2009. URL <http://www.jstatsoft.org/v32/i06/>.
- N. Biggs. *Algebraic Graph Theory*. Cambridge Mathematical Library. Cambridge University Press, 1993. ISBN 9780521458979. URL <https://books.google.es/books?id=6TasRmIF0xQC>.
- Gabriela Bindea, Bernhard Mlecnik, Wolf-Herman Fridman, Franck Pagès, i Jérôme Galon. Natural immunity to cancer in humans. *Current Opinion in Immunology*, 22(2):215 – 222, 2010. ISSN 0952-7915. doi: <http://dx.doi.org/10.1016/j.coi.2010.02.006>. URL <http://www.sciencedirect.com/science/article/pii/S0952791510000385>. Lymphocyte development Tumour immunology.
- Frederick R. Blattner, Guy Plunkett, Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, i Ying Shao. The complete genome sequence of escherichia coli k-12. *Science*, 277(5331):1453–1462, 1997. doi: 10.1126/science.277.5331.1453.
- John C. Boik, Robert A. Newman, i Robert J. Boik. Quantifying synergism/antagonism using nonlinear mixed-effects modeling: A simulation study.

- Statistics in Medicine*, 27(7):1040–1061, 2008. ISSN 1097-0258. doi: 10.1002/sim.3005. URL <http://dx.doi.org/10.1002/sim.3005>.
- Béla Bollobás. *Random graphs*. Cambridge University Press, 2 edition, October 2001. ISBN 0521797225.
- Dylan Breikreutz, Lynn Hlatky, Edward Rietman, i Jack A. Tuszynski. Molecular signaling network complexity is correlated with cancer patient survivability. *Proceedings of the National Academy of Sciences*, 109(23):9209–9212, 2012. doi: 10.1073/pnas.1201416109. URL <http://www.pnas.org/content/109/23/9209.abstract>.
- Bruce A. Chabner i Thomas G. Roberts. Chemotherapy and the war on cancer. *Nature Reviews Cancer*, 5(1):65–72, January 2005. ISSN 1474-175X. doi: 10.1038/nrc1529. URL <http://dx.doi.org/10.1038/nrc1529>.
- Jeffrey T. Chang i Joseph R. Nevins. Gather: a systems approach to interpreting genomic signatures. *Bioinformatics*, 22(23):2926–2933, 2006. doi: 10.1093/bioinformatics/btl483.
- Koei Chin, Sandy DeVries, Jane Fridlyand, Paul T. Spellman, Ritu Roydasgupta, Wen-Lin Kuo, Anna Lapuk, Richard M. Neve, Zuwei Qian, Tom Ryder, Fanqing Chen, Heidi Feiler, Taku Tokuyasu, Chris Kingsley, Shanaz Dairkee, Zhenhang Meng, Karen Chew, Daniel Pinkel, Ajay Jain, Britt M. Ljung, Laura Esserman, Donna G. Albertson, Frederic M. Waldman, i Joe W. Gray. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell*, 10(6):529–541, December 2006. ISSN 15356108. doi: 10.1016/j.ccr.2006.10.009. URL <http://dx.doi.org/10.1016/j.ccr.2006.10.009>.
- Raymond J. Cho, Michael J. Campbell, Elizabeth A. Winzeler, Lars Steinmetz, Andrew Conway, Lisa Wodicka, Tyra G. Wolfsberg, Andrei E. Gabrielian, David Landsman, David J. Lockhart, i Ronald W. Davis. A Genome-Wide Transcriptional Analysis of the Mitotic Cell Cycle. *Molecular Cell*, 2(1):65–73, July 1998. ISSN 10972765. doi: 10.1016/s1097-2765(00)80114-8. URL [http://dx.doi.org/10.1016/s1097-2765\(00\)80114-8](http://dx.doi.org/10.1016/s1097-2765(00)80114-8).
- Ting-Chao Chou. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacological Reviews*, 58(3):621–681, 2006. doi: 10.1124/pr.58.3.10. URL <http://pharmrev.aspetjournals.org/content/58/3/621.abstract>.
- Han-Yu Chuang, Eunjung Lee, Yu-Tsueng Liu, Doheon Lee, i Trey Ideker. Network-based classification of breast cancer metastasis. *Molecular Systems Biology*, 3(1), 2007. ISSN 1744-4292. doi: 10.1038/msb4100180.
- F. Chung. The Diameter of Sparse Random Graphs. *Advances in Applied Mathematics*, 26(4):257–279, May 2001. ISSN 01968858. doi: 10.1006/aama.2001.0720. URL <http://dx.doi.org/10.1006/aama.2001.0720>.
- Aaron Clauset, M. E. J. Newman, i Cristopher Moore. Finding community structure in very large networks. *Physical Review E*, 70(6):066111+, December 2004.

- ISSN 1539-3755. doi: 10.1103/physreve.70.066111. URL <http://dx.doi.org/10.1103/physreve.70.066111>.
- Roser Corominas, Xiping Yang, Guan Ning Lin, Shuli Kang, Yun Shen, Lila Gham-sari, Martin Broly, Maria Rodriguez, Stanley Tam, Shelly A. Trigg, Changyu Fan, Song Yi, Murat Tasan, Irma Lemmens, Xingyan Kuang, Nan Zhao, Dheeraj Mal-hotra, Jacob J. Michaelson, Vladimir Vacic, Michael A. Calderwood, Frederick P. Roth, Jan Tavernier, Steve Horvath, Kouros Salehi-Ashtiani, Dmitry Korkin, Jonathan Sebat, David E. Hill, Tong Hao, Marc Vidal, i Lilia M. Iakoucheva. Protein interaction network of alternatively spliced isoforms from brain links ge-netic risk factors for autism. *Nat Commun*, 5, 2014. doi: 10.1038/ncomms4650. URL <http://dx.doi.org/10.1038/ncomms4650>.
- Jonathan J Crofts i Desmond J Higham. A weighted communicability measure applied to complex brain networks. *Journal of The Royal Society Interface*, 6 (33):411–414, 2009. ISSN 1742-5689. doi: 10.1098/rsif.2008.0484.
- Alan D. D’Andrea. Susceptibility pathways in fanconi’s anemia and breast cancer. *New England Journal of Medicine*, 362(20):1909–1919, 2010. doi: 10.1056/NEJMra0809889. URL <http://dx.doi.org/10.1056/NEJMra0809889>. PMID: 20484397.
- Alan D. D’Andrea i Markus Grompe. The Fanconi anaemia/BRCA pathway. *Nature reviews. Cancer*, 3(1):23–34, January 2003. ISSN 1474-175X. doi: 10.1038/nrc970. URL <http://dx.doi.org/10.1038/nrc970>.
- Joseph L. DeRisi, Vishwanath R. Iyer, i Patrick O. Brown. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278(5338):680–686, 1997. doi: 10.1126/science.278.5338.680. URL <http://www.sciencemag.org/content/278/5338/680.abstract>.
- Zoltan Dezso, Yuri Nikolsky, Tatiana Nikolskaya, Jeremy Miller, David Cherba, Craig Webb, i Andrej Bugrim. Identifying disease-specific genes based on their topological significance in protein networks. *BMC Systems Biology*, 3(1):36+, 2009. ISSN 1752-0509. doi: 10.1186/1752-0509-3-36. URL <http://dx.doi.org/10.1186/1752-0509-3-36>.
- Radu Dobrin, Qasim Beg, Albert L. Barabasi, i Zoltan Oltvai. Aggregation of topological motifs in the Escherichia coli transcriptional regulatory network. *BMC Bioinformatics*, 5(1):10+, 2004. ISSN 1471-2105. doi: 10.1186/1471-2105-5-10. URL <http://dx.doi.org/10.1186/1471-2105-5-10>.
- Mitch Dowsett, Elizabeth Folkard, Deborah Doody, i Ben Haynes. The biology of steroid hormones and endocrine treatment of breast cancer. *The Breast*, 14(6): 452–457, 2005. doi: 10.1016/j.breast.2005.08.019.
- Ian Dunham, Anshul Kundaje, Shelley F. Aldred, Patrick J. Collins, Carrie A. Da-vis, Francis Doyle, Charles B. Epstein, Seth Fretze, Jennifer Harrow, Rajinder Kaul, Jainab Khatun, Bryan R. Lajoie, Stephen G. Landt, Bum-Kyu Lee, Floren-cia Pauli, Kate R. Rosenbloom, Peter Sabo, Alexias Safi, Amartya Sanyal, Noam

Shoresh, Jeremy M. Simon, Lingyun Song, Nathan D. Trinklein, Robert C. Altschuler, James B. Brown, Chao Cheng, Sarah Djebali, Xianjun Dong, Jason Ernst, Terrence S. Furey, Belinda Giardine, Melissa Greven, Robert S. Harris, Javier Hertero, Michael M. Hoffman, Sowmya Iyer, Pouya Kheradpour, Timo Lassmann, Qunhua Li, Xinying Lin, Georgi K. Marinov, Angelika Merkel, Ali Mortazavi, Stephen C. J. Parker, Timothy E. Reddy, Joel Rozowsky, Felix Schlesinger, Robert E. Thurman, Jie Wang, Lucas D. Ward, Troy W. Whitfield, Steven P. Wilder, Weisheng Wu, Hualin S. Xi, Kevin Y. Yip, Jiali Zhuang, Chris Gunter, Michael J. Pazin, Rebecca F. Lowdon, Laura A. L. Dillon, Leslie B. Adams, Caroline J. Kelly, Julia Zhang, Judith R. Wexler, Peter J. Good, Elise A. Feingold, Yanbao Yu, John Wrobel, Brian A. Risk, Harsha P. Gunawardena, Heather C. Kuiper, Christopher W. Maier, Ling Xie, Xian Chen, Tarjei S. Mikkelsen, Shawn Gillespie, Alon Goren, Oren Ram, Xiaolan Zhang, Li Wang, Robbyn Issner, Michael J. Coyne, Timothy Durham, Manching Ku, Thanh Truong, Matthew L. Eaton, Alex Dobin, Andrea Tanzer, Julien Lagarde, Wei Lin, Chenghai Xue, Brian A. Williams, Chris Zaleski, Maik Röder, Felix Kokocinski, Rehab F. Abdelhamid, Tyler Alioto, Igor Antoshechkin, Michael T. Baer, Philippe Batut, Ian Bell, Kimberly Bell, Sudipto Chakraborty, Jacqueline Chrast, Joao Curado, Thomas Derrien, Jorg Drenkow, Erica Dumais, Jackie Dumais, Radha Dutttagupta, Megan Fastuca, Kata Fejes-Toth, Pedro Ferreira, Sylvain Foissac, Melissa J. Fullwood, Hui Gao, David Gonzalez, Assaf Gordon, Cédric Howald, Sonali Jha, Rory Johnson, Philipp Kapranov, Brandon King, Colin Kingswood, Guoliang Li, Oscar J. Luo, Eddie Park, Jonathan B. Preall, Kimberly Presaud, Paolo Ribeca, Daniel Robyr, Xiaolan Ruan, Michael Sammeth, Kuljeet Singh Sandhu, Lorain Schaeffer, Lei-Hoon See, Atif Shahab, Jorgen Skancke, Ana Maria Suzuki, Hazuki Takahashi, Hagen Tilgner, Diane Trout, Nathalie Walters, Huaien Wang, Yoshihide Hayashizaki, Alexandre Reymond, Stylianos E. Antonarakis, Gregory J. Hannon, Yijun Ruan, Piero Carninci, Cricket A. Sloan, Katrina Learned, Venkat S. Malladi, Matthew C. Wong, Galt P. Barber, Melissa S. Cline, Timothy R. Dreszer, Steven G. Heitner, Donna Karolchik, Vanessa M. Kirkup, Laurence R. Meyer, Jeffrey C. Long, Morgan Maddren, Brian J. Raney, Linda L. Grasmeyer, Paul G. Giresi, Anna Battenhouse, Nathan C. Sheffield, Kimberly A. Showers, Darin London, Akshay A. Bhinge, Christopher Shestak, Matthew R. Schaner, Seul Ki Kim, Zhuzhu Z. Zhang, Piotr A. Mieczkowski, Joanna O. Mieczkowska, Zheng Liu, Ryan M. McDaniell, Yunyun Ni, Naim U. Rashid, Min Jae Kim, Sheera Adar, Zhancheng Zhang, Tianyuan Wang, Deborah Winter, Damian Keefe, Vishwanath R. Iyer, Meizhen Zheng, Ping Wang, Jason Gertz, Jost Vielmetter, E. Christopher Partridge, Katherine E. Varley, Clarke Gasper, Anita Bansal, Shirley Pepke, Preti Jain, Henry Amrhein, Kevin M. Bowling, Michael Anaya, Marie K. Cross, Michael A. Muratet, Nicholas S. Davis, Kenneth McCue, Tracy Eggleston, Katherine I. Fisher-Aylor, Gilberto DeSalvo, Sarah K. Meadows, Sreeram Balasubramanian, Amy S. Nesmith, J. Scott Newberry, Kimberly M. Newberry, Stephanie L. Parker, Barbara Pusey, Shawn E. Levy, Devin M. Absher, Wing H. Wong, Matthew J.



Blow, Axel Visel, Len A. Pennachio, Hanna M. Petrykowska, Alexej Abyzov, Bronwen Aken, Daniel Barrell, Gemma Barson, Andrew Berry, Alexandra Big-nell, Veronika Boychenko, Giovanni Bussotti, Claire Davidson, Gloria Despacio-Reyes, Mark Diekhans, Iakes Ezkurdia, Adam Frankish, James Gilbert, Jose Manuel Gonzalez, Ed Griffiths, Rachel Harte, David A. Hendrix, Toby Hunt, Irwin Jungreis, Mike Kay, Ekta Khurana, Jing Leng, Michael F. Lin, Jane Loveland, Zhi Lu, Deepa Manthravadi, Marco Mariotti, Jonathan Mudge, Gaurab Mukherjee, Cedric Notredame, Baikang Pei, Jose Manuel Rodriguez, Gary Saunders, Andrea Sboner, Stephen Searle, Cristina Sisu, Catherine Snow, Charlie Steward, Elec-tra Tapanari, Michael L. Tress, Marijke J. van Baren, Stefan Washietl, Laurens Wilming, Amonida Zadissa, Zhengdong Zhang, Michael Brent, David Haussler, Alfonso Valencia, Nick Addleman, Roger P. Alexander, Raymond K. Auerbach, Suganthi Balasubramanian, Keith Bettinger, Nitin Bhardwaj, Alan P. Boyle, Ali-na R. Cao, Philip Cayting, Alexandra Charos, Yong Cheng, Catharine Eastman, Ghia Euskirchen, Joseph D. Fleming, Fabian Grubert, Lukas Habegger, Manoj Hariharan, Arif Harmanaci, Sushma Iyengar, Victor X. Jin, Konrad J. Karczewski, Maya Kasowski, Phil Lacroute, Hugo Lam, Nathan Lamarre-Vincent, Jin Lian, Marianne Lindahl-Allen, Renqiang Min, Benoit Miotto, Hannah Monahan, Zar-mik Moqtaderi, Xinmeng J. Mu, Henriette O'Geen, Zhengqing Ouyang, Dorrelyn Patacsil, Debasish Raha, Lucia Ramirez, Brian Reed, Minyi Shi, Teri Slifer, Heat-her Witt, Linfeng Wu, Xiaoqin Xu, Koon-Kiu Yan, Xinqiong Yang, Kevin Struhl, Sherman M. Weissman, Luiz O. Penalva, Subhradip Karmakar, Raj R. Bhanvadia, Alina Choudhury, Marc Domanus, Lijia Ma, Jennifer Moran, Alec Victorsen, Tho-mas Auer, Lazaro Centanin, Michael Eichenlaub, Franziska Gruhl, Stephan Heer-mann, Burkhard Hoekendorf, Daigo Inoue, Tanja Kellner, Stephan Kirchmaier, Claudia Mueller, Robert Reinhardt, Lea Schertel, Stephanie Schneider, Rebecca Sinn, Jochen Wittbrodt, Beate Wittbrodt, Gaurav Jain, Gayathri Balasundaram, Daniel L. Bates, Rachel Byron, Theresa K. Canfield, Morgan J. Diegel, Douglas Dunn, Abigail K. Ebersol, Tristan Frum, Kavita Garg, Erica Gist, R. Scott Han-sen, Lisa Boatman, Eric Haugen, Richard Humbert, Audra K. Johnson, Ericka M. Johnson, Tattyana V. Kutyaev, Kristen Lee, Dimitra Lotakis, Matthew T. Mau-rano, Shane J. Neph, Fiedencio V. Neri, Eric D. Nguyen, Hongzhu Qu, Alex P. Reynolds, Vaughn Roach, Eric Rynes, Minerva E. Sanchez, Richard S. Sands-trom, Anthony O. Shafer, Andrew B. Stergachis, Sean Thomas, Benjamin Vernot, Jeff Vierstra, Shinny Vong, Hao Wang, Molly A. Weaver, Yongqi Yan, Miaohua Zhang, Joshua M. Akey, Michael Bender, Michael O. Dorschner, Mark Groudine, Michael J. MacCoss, Patrick Navas, George Stamatoyannopoulos, Kathryn Beal, Alvis Brazma, Paul Flicek, Nathan Johnson, Margus Lukk, Nicholas M. Luscom-be, Daniel Sobral, Juan M. Vaquerizas, Serafim Batzoglou, Arend Sidow, Nadi-ne Hussami, Sofia Kyriazopoulou-Panagiotopoulou, Max W. Libbrecht, Marc A. Schaub, Webb Miller, Peter J. Bickel, Balazs Banfai, Nathan P. Boley, Haiyan Huang, Jingyi Jessica Li, William Stafford Noble, Jeffrey A. Bilmes, Orion J. Buske, Avinash D. Sahu, Peter V. Kharchenko, Peter J. Park, Dannon Baker,

- James Taylor, Lucas Lochovsky, Bradley E. Bernstein, Gregory E. Crawford, Job Dekker, Laura Elnitski, Peggy J. Farnham, Mark Gerstein, Morgan C. Giddings, Thomas R. Gingeras, Eric D. Green, Roderic Guigó, Ross C. Hardison, Timothy J. Hubbard, Manolis Kellis, W. James Kent, Jason D. Lieb, Elliott H. Margulies, Richard M. Myers, Michael Snyder, John A. Stamatoyannopoulos, Scott A. Tenenbaum, Zhiping Weng, Kevin P. White, Barbara Wold, i Ewan Birney. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, September 2012. ISSN 1476-4687. doi: 10.1038/nature11247. URL <http://dx.doi.org/10.1038/nature11247>.
- Michael B. Eisen, Paul T. Spellman, Patrick O. Brown, i David Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95(25):14863–14868, December 1998. ISSN 1091-6490. doi: 10.1073/pnas.95.25.14863. URL <http://dx.doi.org/10.1073/pnas.95.25.14863>.
- The C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, 282(5396):2012–2018, 1998. doi: 10.1126/science.282.5396.2012. URL <http://www.sciencemag.org/content/282/5396/2012.abstract>.
- Lyndsey A. Emery, Anusri Tripathi, Chialin King, Maureen Kavanah, Jane Mendez, Michael D. Stone, Antonio de las Morenas, Paola Sebastiani, i Carol L. Rosenberg. Early dysregulation of cell adhesion and extracellular matrix pathways in breast cancer progression. *The American Journal of Pathology*, 175:1292–1302, 2009. doi: 10.2353/ajpath.2009.090115. URL <http://dx.doi.org/10.2353/ajpath.2009.090115>.
- P. Erdős i A Rényi. On random graphs i. *Publ. Maths. Debrecen*, 6:290–297, 1959.
- P. Erdős i A Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.
- Ernesto Estrada. Characterization of 3d molecular structure. *Chemical Physics Letters*, 319(5–6):713 – 718, 2000. ISSN 0009-2614. doi: [http://dx.doi.org/10.1016/S0009-2614\(00\)00158-5](http://dx.doi.org/10.1016/S0009-2614(00)00158-5). URL <http://www.sciencedirect.com/science/article/pii/S0009261400001585>.
- Ernesto Estrada i Juan A. Rodríguez-Velázquez. Subgraph centrality in complex networks. *Phys. Rev. E*, 71:056103, May 2005. doi: 10.1103/PhysRevE.71.056103. URL <http://link.aps.org/doi/10.1103/PhysRevE.71.056103>.
- Cheng Fan, Daniel S. Oh, Lodewyk Wessels, Britta Weigelt, Dimitry S.A. Nuyten, Andrew B. Nobel, Laura J. van't Veer, i Charles M. Perou. Concordance among gene-expression-based predictors for breast cancer. *New England Journal of Medicine*, 355(6):560–569, 2006. doi: 10.1056/NEJMoa052933.
- Jacques Ferlay, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, i Freddie Bray. Cancer incidence and mortality worldwide: Sources, methods and major patterns in globo-

- can 2012. *International Journal of Cancer*, 136(5):E359–E386, 2015. ISSN 1097-0215. doi: 10.1002/ijc.29210. URL <http://dx.doi.org/10.1002/ijc.29210>.
- Linton C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1):35–41, March 1977. URL <http://links.jstor.org/sici?sici=0038-0431%28197703%2940%3A1%3C35%3AASOMOC%3E2.O.CO%3B2-H>.
- David A. Fruman i Christian Rommel. Pi3k and cancer: lessons, challenges and opportunities. *Nat Rev Drug Discov*, 13(2):140–156, 2014. ISSN 1474-1776. doi: 10.1038/nrd4204. URL <http://dx.doi.org/10.1038/nrd4204>.
- O Galamb, S Spisak, F Sipos, K Toth, N Solymosi, B Wichmann, T Krenacs, G Valcz, Z Tulassay, i B Molnar. Reversal of gene expression changes in the colorectal normal-adenoma pathway by ns398 selective cox2 inhibitor. *Br J Cancer*, 102(4):765–773, 2010. doi: 10.1038/sj.bjc.6605515. URL <http://dx.doi.org/10.1038/sj.bjc.6605515>.
- Mathew J. Garnett, Elena J. Edelman, Sonja J. Heidorn, Chris D. Greenman, Anahita Dastur, King Wai Lau, Patricia Greninger, I. Richard Thompson, Xi Luo, Jorge Soares, Qingsong Liu, Francesco Iorio, Didier Surdez, Li Chen, Randy J. Milano, Graham R. Bignell, Ah T. Tam, Helen Davies, Jesse A. Stevenson, Syd Barthorpe, Stephen R. Lutz, Fiona Kogera, Karl Lawrence, Anne McLaren-Douglas, Xenia Mitropoulos, Tatiana Mironenko, Helen Thi, Laura Richardson, Wenjun Zhou, Frances Jewitt, Tinghu Zhang, Patrick O’Brien, Jessica L. Boisvert, Stacey Price, Wooyoung Hur, Wanjuan Yang, Xianming Deng, Adam Butler, Hwan Geun Choi, Jae Won Chang, Jose Baselga, Ivan Stamenkovic, Jeffrey A. Engelman, Sreenath V. Sharma, Olivier Delattre, Julio Saez-Rodriguez, Nathanael S. Gray, Jeffrey Settleman, P. Andrew Futreal, Daniel A. Haber, Michael R. Stratton, Sridhar Ramaswamy, Ultan McDermott, i Cyril H. Benes. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483(7391):570–575, 03 2012. ISSN 0028-0836. doi: 10.1038/nature11005. URL <http://dx.doi.org/10.1038/nature11005>.
- Kwang-Il Goh, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal, i Albert-László Barabási. The human disease network. *Proceedings of the National Academy of Sciences*, 104(21):8685–8690, 2007. doi: 10.1073/pnas.0701361104.
- Aric A. Hagberg, Daniel A. Schult, i Pieter J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA, August 2008.
- Jing-Dong J. Han, Denis Dupuy, Nicolas Bertin, Michael E. Cusick, i Marc Vidal. Effect of sampling on topology predictions of protein-protein interaction networks. *Nat Biotech*, 23(7):839–844, July 2005. ISSN 1087-0156. doi: 10.1038/nbt1116. URL <http://dx.doi.org/10.1038/nbt1116>.
- Douglas Hanahan i Robert A Weinberg. The hallmarks of cancer. *Cell*, 100(1):57 – 70, 2000. doi: [http://dx.doi.org/10.1016/S0092-8674\(00\)81683-9](http://dx.doi.org/10.1016/S0092-8674(00)81683-9).
- Douglas Hanahan i Robert A. Weinberg. Hallmarks of cancer: The next generation.

- Cell*, 144(5):646 – 674, 2011. ISSN 0092-8674. doi: <http://dx.doi.org/10.1016/j.cell.2011.02.013>. URL <http://www.sciencedirect.com/science/article/pii/S0092867411001279>.
- Renu A. Heller, Mark Schena, Andrew Chai, Dari Shalon, Tod Bedilion, James Gilmore, David E. Woolley, i Ronald W. Davis. Discovery and analysis of inflammatory disease-related genes using cdna microarrays. *Proceedings of the National Academy of Sciences*, 94(6):2150–2155, 1997. URL <http://www.pnas.org/content/94/6/2150.abstract>.
- Pilar Hernández, Jaime Huerta-Cepas, David Montaner, Fátima AlShahrour, Joan Valls, Laia Gómez, Gabriel Capellá, Joaquín Dopazo, i Miguel Angel Pujana. Evidence for systems-level molecular mechanisms of tumorigenesis. *BMC Genomics*, 8:185, 2007. doi: 10.1186/1471-2164-8-185.
- Desmond J. Higham, Gabriela Kalna, i Milla Kibble. Spectral clustering and its use in bioinformatics. *Journal of Computational and Applied Mathematics*, 204(1):25 – 37, 2007. ISSN 0377-0427. doi: <http://dx.doi.org/10.1016/j.cam.2006.04.026>. URL <http://www.sciencedirect.com/science/article/pii/S0377042706002366>. Special issue dedicated to Professor Shinnosuke Oharu on the occasion of his 65th birthday.
- Caitriona Holohan, Sandra Van Schaeybroeck, Daniel B. Longley, i Patrick G. Johnston. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer*, 13(10):714–726, 2013. ISSN 1474-175X. doi: 10.1038/nrc3599. URL <http://dx.doi.org/10.1038/nrc3599>.
- S. Itzkovitz, R. Milo, N. Kashtan, G. Ziv, i U. Alon. Subgraphs in random networks. *Phys. Rev. E*, 68:026127, Aug 2003. doi: 10.1103/PhysRevE.68.026127. URL <http://link.aps.org/doi/10.1103/PhysRevE.68.026127>.
- Filip Janku, Jennifer J. Wheler, Shannon N. Westin, Stacy L. Moulder, Aung Naing, Apostolia M. Tsimberidou, Siqing Fu, Gerald S. Falchook, David S. Hong, Ignacio Garrido-Laguna, Rajyalakshmi Luthra, J. Jack Lee, Karen H. Lu, i Razole Kurzrock. Pi3k/akt/mtor inhibitors in patients with breast and gynecologic malignancies harboring pik3ca mutations. *Journal of Clinical Oncology*, 30(8):777–782, 2012. doi: 10.1200/JCO.2011.36.1196. URL <http://jco.ascopubs.org/content/30/8/777.abstract>.
- H. Jeong, S. P. Mason, A. L. Barabasi, i Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, May 2001. ISSN 0028-0836. doi: 10.1038/35075138. URL <http://dx.doi.org/10.1038/35075138>.
- Pall F. Jonsson i Paul A. Bates. Global topological features of cancer proteins in the human interactome. *Bioinformatics*, 22(18):2291–2297, 2006. doi: 10.1093/bioinformatics/btl390. URL <http://bioinformatics.oxfordjournals.org/content/22/18/2291.abstract>.
- Ashish Juvekar, Laura N. Burga, Hai Hu, Elaine P. Lunsford, Yasir H. Ibrahim, Judith Balmañà, Anbazhagan Rajendran, Antonella Papa, Katherine Spencer,

- Costas A. Lyssiotis, Caterina Nardella, Pier Paolo Pandolfi, José Baselga, Ralph Scully, John M. Asara, Lewis C. Cantley, i Gerburg M. Wulf. Combining a pi3k inhibitor with a parp inhibitor provides an effective therapy for brca1-related breast cancer. *Cancer Discovery*, 2(11):1048–1063, 2012. doi: 10.1158/2159-8290.CD-11-0336. URL <http://cancerdiscovery.aacrjournals.org/content/2/11/1048.abstract>.
- Minoru Kanehisa, Yoko Sato, Masayuki Kawashima, Miho Furumichi, i Mao Tanabe. Kegg as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 2015. doi: 10.1093/nar/gkv1070. URL <http://nar.oxfordjournals.org/content/early/2015/10/15/nar.gkv1070.abstract>.
- T. S. Keshava Prasad, Renu Goel, Kumaran Kandasamy, Shivakumar Keerthikumar, Sameer Kumar, Suresh Mathivanan, Deepthi Telikicherla, Rajesh Raju, Beema Shafreen, Abhilash Venugopal, Lavanya Balakrishnan, Arivusudar Marimuthu, Sutopa Banerjee, Devi S. Somanathan, Aimy Sebastian, Sandhya Rani, Somak Ray, C. J. Harrys Kishore, Sashi Kanth, Mukhtar Ahmed, Manoj K. Kashyap, Riaz Mohmood, Y. L. Ramachandra, V. Krishna, B. Abdul Rahiman, Sujatha Mohan, Prathibha Ranganathan, Subhashri Ramabadran, Raghothama Chaerkady, i Akhilesh Pandey. Human protein reference database—2009 update. *Nucleic Acids Research*, 37(suppl 1):D767–D772, 2009. doi: 10.1093/nar/gkn892.
- Min-Sik Kim, Sneha M. Pinto, Derese Getnet, Raja S. Nirujogi, Srikanth S. Manda, Raghothama Chaerkady, Anil K. Madugundu, Dhanashree S. Kelkar, Ruth Isserlin, Shobhit Jain, Joji K. Thomas, Babylakshmi Muthusamy, Pamela Leal-Rojas, Praveen Kumar, Nandini A. Sahasrabuddhe, Lavanya Balakrishnan, Jayshree Advani, Bijesh George, Santosh Renuse, Lakshmi D. Selvan, Arun H. Patil, Vishalakshi Nanjappa, Aneesha Radhakrishnan, Samarjeet Prasad, Tejaswini Subbannayya, Rajesh Raju, Manish Kumar, Sreelakshmi K. Sreenivasamurthy, Arivusudar Marimuthu, Gajanan J. Sathe, Sandip Chavan, Keshava K. Datta, Yashwanth Subbannayya, Apeksha Sahu, Soujanya D. Yelamanchi, Savita Jayaram, Pavithra Rajagopalan, Jyoti Sharma, Krishna R. Murthy, Nazia Syed, Renu Goel, Aafaque A. Khan, Sartaj Ahmad, Gourav Dey, Keshav Mudgal, Aditi Chatterjee, Tai-Chung Huang, Jun Zhong, Xinyan Wu, Patrick G. Shaw, Donald Freed, Muhammad S. Zahari, Kanchan K. Mukherjee, Subramanian Shankar, Anita Mahadevan, Henry Lam, Christopher J. Mitchell, Susarla K. Shankar, Parthasarathy Satishchandra, John T. Schroeder, Ravi Sirdeshmukh, Anirban Maitra, Steven D. Leach, Charles G. Drake, Marc K. Halushka, T. S. Keshava Prasad, Ralph H. Hruban, Candace L. Kerr, Gary D. Bader, Christine A. Iacobuzio-Donahue, Harsha Gowda, i Akhilesh Pandey. A draft map of the human proteome. *Nature*, 509(7502):575–581, May 2014. ISSN 0028-0836. doi: 10.1038/nature13302. URL <http://dx.doi.org/10.1038/nature13302>.
- Hiroaki Kitano. Systems biology: A brief overview. *Science*, 295(5560):1662–1664, 2002. doi: 10.1126/science.1069492.
- Kakajan Komurov, Jen-Te Tseng, Melissa Muller, Elena G Seviour, Tyler J Moss,

- Lifeng Yang, Deepak Nagrath, i Prahlad T Ram. The glucose-deprivation network counteracts lapatinib-induced toxicity in resistant erbb2-positive breast cancer cells. *Molecular Systems Biology*, 8(1), 2012. ISSN 1744-4292. doi: 10.1038/msb.2012.25.
- Pamela K. Kreeger i Douglas A. Lauffenburger. Cancer systems biology: a network modeling perspective. *Carcinogenesis*, 31(1):2–8, 2010. doi: 10.1093/carcin/bgp261. URL <http://carcin.oxfordjournals.org/content/31/1/2.abstract>.
- Petric Kuballa, Whitney M. Nolte, Adam B. Castoreno, i Ramnik J. Xavier. Autophagy and the immune system. *Annual Review of Immunology*, 30(1):611–646, 2012. doi: 10.1146/annurev-immunol-020711-074948. URL <http://dx.doi.org/10.1146/annurev-immunol-020711-074948>. PMID: 22449030.
- K Lai, M C Killingsworth, i C S Lee. Gene of the month: Pik3ca. *Journal of Clinical Pathology*, 68(4):253–257, 2015. doi: 10.1136/jclinpath-2015-202885. URL <http://jcp.bmj.com/content/68/4/253.abstract>.
- F Laloo i D G Evans. Familial breast cancer. *Clinical Genetics*, 82(2):105–114, 2012. ISSN 1399-0004. doi: 10.1111/j.1399-0004.2012.01859.x. URL <http://dx.doi.org/10.1111/j.1399-0004.2012.01859.x>.
- Justin Lamb, Sridhar Ramaswamy, Heide L. Ford, Bernardo Contreras, Robert V. Martinez, Frances S. Kittrell, Cynthia A. Zahnow, Nick Patterson, Todd R. Golub, i Mark E. Ewen. A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. *Cell*, 114(3):323–334, August 2003. ISSN 0092-8674. URL <http://view.ncbi.nlm.nih.gov/pubmed/12914697>.
- Pierre Legrain, Ruedi Aebersold, Alexander Archakov, Amos Bairoch, Kumar Bala, Laura Beretta, John Bergeron, Christoph H. Borchers, Garry L. Corthals, Catherine E. Costello, Eric W. Deutsch, Bruno Domon, William Hancock, Fuchu He, Denis Hochstrasser, György Marko-Varga, Ghasem Hosseini Salekdeh, Salvatore Sechi, Michael Snyder, Sudhir Srivastava, Mathias Uhlén, Cathy H. Wu, Tadashi Yamamoto, Young-Ki Paik, i Gilbert S. Omenn. The human proteome project: Current state and future direction. *Molecular & Cellular Proteomics*, 10(7), 2011. doi: 10.1074/mcp.M111.009993. URL <http://www.mcponline.org/content/10/7/M111.009993.abstract>.
- Joseph Lehar, Andrew S. Krueger, William Avery, Adrian M. Heilbut, Lisa M. Johansen, E. Roydon Price, Richard J. Rickles, Glenn F. Short, Jane E. Staunton, Xiaowei Jin, Margaret S. Lee, Grant R. Zimmermann, i Alexis A. Borisy. Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nature Biotechnology*, 27(7):659–666, July 2009. ISSN 1087-0156. doi: 10.1038/nbt.1549. URL <http://dx.doi.org/10.1038/nbt.1549>.
- D. J. Lockhart, H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, i E. L. Brown. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Na-*

- ture biotechnology*, 14(13):1675–1680, December 1996. ISSN 1087-0156. doi: 10.1038/nbt1296-1675. URL <http://dx.doi.org/10.1038/nbt1296-1675>.
- Xiao-Jun Ma, Sonika Dahiya, Elizabeth Richardson, Mark Erlander, i Dennis C. Sgroi. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Research*, 11(1):R7+, February 2009. ISSN 1465-5411. doi: 10.1186/bcr2222. URL <http://dx.doi.org/10.1186/bcr2222>.
- Donna Maglott, Jim Ostell, Kim D. Pruitt, i Tatiana Tatusova. Entrez gene: gene-centered information at ncbi. *Nucleic Acids Research*, 33(suppl 1):D54–D58, 2005. doi: 10.1093/nar/gki031.
- Tobias Maier, Marc Güell, i Luis Serrano. Correlation of mrna and protein in complex biological samples. *{FEBS} Letters*, 583(24):3966 – 3973, 2009. ISSN 0014-5793. doi: <http://dx.doi.org/10.1016/j.febslet.2009.10.036>. URL <http://www.sciencedirect.com/science/article/pii/S0014579309008126>. Systems Biology - Nobel Symposium 146.
- Mathew J. Marton, Joseph L. DeRisi, Holly A. Bennet, Vishwanath R. Iyer, Michael R. Meyer, Christofer J. Roberts, Roland Stoughton, Julja Burchard, David Slade, Hongyue Dai, Douglas E. Basset, Leland H. Hartwell, Patrick O. Brown, i Stephen H Friend. Drug target validation and identification of secondary drug target effects using dna microarrays. *Nat. Med.*, 4(11):1293–1301, 11 1998. URL <http://dx.doi.org/10.1038/3282>.
- Suleiman Massarweh, C. Kent Osborne, Chad J. Creighton, Lanfang Qin, Anna Tsimelzon, Shixia Huang, Heidi Weiss, Mothaffar Rimawi, i Rachel Schiff. Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function. *Cancer Research*, 68(3):826–833, 2008. doi: 10.1158/0008-5472.CAN-07-2707. URL <http://cancerres.aacrjournals.org/content/68/3/826.abstract>.
- Stanley Milgram. The small-world problem. *Psychology Today*, 1(1):61–67, 5 1967.
- R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, i U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002. doi: 10.1126/science.298.5594.824. URL <http://www.sciencemag.org/content/298/5594/824.abstract>.
- Kevin Moreau, Maurizio Renna, i David C. Rubinsztein. Connections between {SNAREs} and autophagy. *Trends in Biochemical Sciences*, 38(2):57 – 63, 2013. ISSN 0968-0004. doi: <http://dx.doi.org/10.1016/j.tibs.2012.11.004>. URL <http://www.sciencedirect.com/science/article/pii/S0968000412001879>.
- Richard M. Neve, Koei Chin, Jane Fridlyand, Jennifer Yeh, Frederick L. Baehner, Tea Fevr, Laura Clark, Nora Bayani, Jean-Philippe Coppe, Frances Tong, Terry Speed, Paul T. Spellman, Sandy DeVries, Anna Lapuk, Nick J. Wang, Wen-Lin Kuo, Jackie L. Stilwell, Daniel Pinkel, Donna G. Albertson, Frederic M. Waldman, Frank McCormick, Robert B. Dickson, Michael D. Johnson, Marc Lippman, Stephen Ethier, Adi Gazdar, i Joe W. Gray. A collection of breast cancer

- cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*, 10(6):515–527, July 2006. ISSN 1535-6108. doi: 10.1016/j.ccr.2006.10.008. URL <http://dx.doi.org/10.1016/j.ccr.2006.10.008>.
- M. E. J. Newman. Assortative Mixing in Networks. *Physical Review Letters*, 89(20):208701+, October 2002. ISSN 0031-9007. doi: 10.1103/physrevlett.89.208701. URL <http://dx.doi.org/10.1103/physrevlett.89.208701>.
- M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003. doi: 10.1137/S003614450342480.
- Sandra Orchard, Mais Ammari, Bruno Aranda, Lionel Breuza, Leonardo Briganti, Fiona Broackes-Carter, Nancy H Campbell, Gayatri Chavali, Carol Chen, Noemi del Toro, Margaret Duesbury, Marine Dumousseau, Eugenia Galeota, Ursula Hinz, Marta Iannuccelli, Sruthi Jagannathan, Rafael Jimenez, Jyoti Khadake, Astrid Lagreid, Luana Licata, Ruth C Lovering, Birgit Meldal, Anna N Melidoni, Mila Milagros, Daniele Peluso, Livia Perfetto, Pablo Porras, Arathi Raghunath, Sylvie Ricard-Blum, Bernd Roechert, Andre Stutz, Michael Tognolli, Kim van Roey, Gianni Cesareni, i Henning Hermjakob. The mintact project—intact as a common curation platform for 11 molecular interaction databases. *Nucleic acids research*, 42(Database issue):D358—63, January 2014. ISSN 0305-1048. doi: 10.1093/nar/gkt1115. URL <http://europepmc.org/articles/PMC3965093>.
- The Human Proteome Organization. A gene-centric human proteome project: Hupo—the human proteome organization. *Molecular & Cellular Proteomics*, 9(2):427–429, 2010. doi: 10.1074/mcp.H900001-MCP200. URL <http://www.mcponline.org/content/9/2/427.short>.
- C. Kent Osborne. Tamoxifen in the treatment of breast cancer. *New England Journal of Medicine*, 339(22):1609–1618, 1998. doi: 10.1056/NEJM199811263392207. URL <http://dx.doi.org/10.1056/NEJM199811263392207>. PMID: 9828250.
- Soonmyung Paik, Steven Shak, Gong Tang, Chungyeul Kim, Joffre Baker, Maureen Cronin, Frederick L. Baehner, Michael G. Walker, Drew Watson, Taesung Park, William Hiller, Edwin R. Fisher, D. Lawrence Wickerham, John Bryant, i Norman Wolmark. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine*, 351(27):2817–2826, 2004. doi: 10.1056/NEJMoa041588. URL <http://dx.doi.org/10.1056/NEJMoa041588>.
- Young-Ki Paik, Seul-Ki Jeong, Gilbert S. Omenn, Mathias Uhlen, Samir Hanash, Sang Y. Cho, Hyoung-Joo Lee, Keun Na, Eun-Young Choi, Fangfei Yan, Fan Zhang, Yue Zhang, Michael Snyder, Yong Cheng, Rui Chen, Gyorgy Marko-Varga, Eric W. Deutsch, Hoguen Kim, Ja-Young Kwon, Ruedi Aebersold, Amos Bairoch, Allen D. Taylor, Kwang Y. Kim, Eun-Young Lee, Denis Hochstrasser, Pierre Legrain, i William S. Hancock. The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat Biotech*,



- 30(3):221–223, March 2012. ISSN 1087-0156. doi: 10.1038/nbt.2152. URL <http://dx.doi.org/10.1038/nbt.2152>.
- C. M. Perou, T. Sørli, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lønning, A. L. Børresen-Dale, P. O. Brown, i D. Botstein. Molecular portraits of human breast tumours. *Nature*, 406(6797): 747–752, August 2000. ISSN 0028-0836. doi: 10.1038/35021093.
- Charles M. Perou, Stefanie S. Jeffrey, Matt van de Rijn, Christian A. Rees, Michael B. Eisen, Douglas T. Ross, Alexander Pergamenschikov, Cheryl F. Williams, Shirley X. Zhu, Jeffrey C. F. Lee, Deval Lashkari, Dari Shalon, Patrick O. Brown, i David Botstein. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proceedings of the National Academy of Sciences*, 96(16):9212–9217, August 1999. ISSN 1091-6490. doi: 10.1073/pnas.96.16.9212. URL <http://dx.doi.org/10.1073/pnas.96.16.9212>.
- Davide Rambaldi, Federico M. Giorgi, Fabrizio Capuani, Andrea Ciliberto, i Francesca D. Ciccarelli. Low duplicability and network fragility of cancer genes. *Trends in Genetics*, 24(9):427–430, 2015. doi: 10.1016/j.tig.2008.06.003. URL <http://dx.doi.org/10.1016/j.tig.2008.06.003>.
- Erzsébet Ravasz i Albert-László Barabási. Hierarchical organization in complex networks. *Phys. Rev. E*, 67:026112, Feb 2003. doi: 10.1103/PhysRevE.67.026112. URL <http://link.aps.org/doi/10.1103/PhysRevE.67.026112>.
- A. L. Richardson, Z. C. Wang, A. De Nicolo, X. Lu, M. Brown, A. Miron, X. Liao, J. D. Iglehart, D. M. Livingston, i S. Ganesan. X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell*, 9(2):121–132, February 2006. ISSN 1535-6108. doi: 10.1016/j.ccr.2006.01.013. URL <http://dx.doi.org/10.1016/j.ccr.2006.01.013>.
- Alexander W. Rives i Timothy Galitski. Modular organization of cellular networks. *Proceedings of the National Academy of Sciences*, 100(3):1128–1133, 2003. doi: 10.1073/pnas.0237338100. URL <http://www.pnas.org/content/100/3/1128.abstract>.
- Thomas Rolland, Murat Taşan, Benoit Charlotiaux, Samuel J. Pevzner, Quan Zhong, Nidhi Sahni, Song Yi, Irma Lemmens, Celia Fontanillo, Roberto Mosca, Atanas Kamburov, Susan D. Ghiassian, Xinping Yang, Lila Ghamsari, Dawit Balcha, Bridget E. Begg, Pascal Braun, Marc Brehme, Martin P. Broly, Anne-Ruxandra Carvunis, Dan Convery-Zupan, Roser Corominas, Jasmin Coulombe-Huntington, Elizabeth Dann, Matija Dreze, Amélie Dricot, Changyu Fan, Eric Franzosa, Fana Gebreab, Bryan J. Gutierrez, Madeleine F. Hardy, Mike Jin, Shuli Kang, Ruth Kiros, Guan N. Lin, Katja Luck, Andrew MacWilliams, Jörg Menche, Ryan R. Murray, Alexandre Palagi, Matthew M. Poulin, Xavier Rambout, John Rasla, Patrick Reichert, Viviana Romero, Elien Ruysinck, Julie M. Sahalie, Annemarie Scholz, Akash A. Shah, Amitabh Sharma, Yun Shen, Kerstin

- Spirohn, Stanley Tam, Alexander O. Tejada, Shelly A. Trigg, Jean-Claude Twizere, Kerwin Vega, Jennifer Walsh, Michael E. Cusick, Yu Xia, Albert-László Barabási, Lilia M. Iakoucheva, Patrick Aloy, Javier De Las Rivas, Jan Tavernier, Michael A. Calderwood, David E. Hill, Tong Hao, Frederick P. Roth, i Marc Vidal. A Proteome-Scale Map of the Human Interactome Network. *Cell*, 159(5): 1212–1226, November 2014. ISSN 00928674. doi: 10.1016/j.cell.2014.10.050. URL <http://dx.doi.org/10.1016/j.cell.2014.10.050>.
- P. Ross-Macdonald, P. S. Coelho, T. Roemer, S. Agarwal, A. Kumar, R. Jansen, K. H. Cheung, A. Sheehan, D. Symoniatis, L. Umansky, M. Heidtman, F. K. Nelson, H. Iwasaki, K. Hager, M. Gerstein, P. Miller, G. S. Roeder, i M. Snyder. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature*, 402(6760):413–418, November 1999. ISSN 0028-0836. doi: 10.1038/46558. URL <http://dx.doi.org/10.1038/46558>.
- Jean-François Rual, Kavitha Venkatesan, Tong Hao, Tomoko Hirozane-Kishikawa, Amélie Dricot, Ning Li, Gabriel F. Berriz, Francis D. Gibbons, Matija Dreze, Nono Ayivi-Guedehoussou, Niels Klitgord, Christophe Simon, Mike Boxem, Stuart Milstein, Jennifer Rosenberg, Debra S. Goldberg, Lan V. Zhang, Sharyl L. Wong, Giovanni Franklin, Siming Li, Joanna S. Albala, Janghoo Lim, Carlene Fraughton, Estelle Llamosas, Sebiha Cevik, Camille Bex, Philippe Lamesch, Robert S. Sikorski, Jean Vandenhoute, Huda Y. Zoghbi, Alex Smolyar, Stephanie Bosak, Reynaldo Sequerra, Lynn Doucette-Stamm, Michael E. Cusick, David E. Hill, Frederick P. Roth, i Marc Vidal. Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, 437(7062):1173–1178, September 2005a. ISSN 0028-0836. doi: 10.1038/nature04209. URL <http://dx.doi.org/10.1038/nature04209>.
- Jean-François Rual, Kavitha Venkatesan, Tong Hao, Tomoko Hirozane-Kishikawa, Amélie Dricot, Ning Li, Gabriel F. Berriz, Francis D. Gibbons, Matija Dreze, Nono Ayivi-Guedehoussou, Niels Klitgord, Christophe Simon, Mike Boxem, Stuart Milstein, Jennifer Rosenberg, Debra S. Goldberg, Lan V. Zhang, Sharyl L. Wong, Giovanni Franklin, Siming Li, Joanna S. Albala, Janghoo Lim, Carlene Fraughton, Estelle Llamosas, Sebiha Cevik, Camille Bex, Philippe Lamesch, Robert S. Sikorski, Jean Vandenhoute, Huda Y. Zoghbi, Alex Smolyar, Stephanie Bosak, Reynaldo Sequerra, Lynn Doucette-Stamm, Michael E. Cusick, David E. Hill, Frederick P. Roth, i Marc Vidal. Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, 437(7062):1173–1178, September 2005b. ISSN 0028-0836. doi: 10.1038/nature04209. URL <http://dx.doi.org/10.1038/nature04209>.
- Anguraj Sadanandam, Costas A. Lyssiotis, Krisztian Homicsko, Eric A. Collisson, William J. Gibb, Stephan Wullschleger, Liliane C. Ostos, William A. Lannon, Carsten Grotzinger, Maguy Del Rio, Benoît Lhermitte, Adam B. Olshen, Bertram Wiedenmann, Lewis C. Cantley, Joe W. Gray, i Douglas Hanahan. A colorectal cancer classification system that associates cellular phenotype and responses to

- therapy. *Nat Med*, 19(5):619–625, May 2013. ISSN 1546-170X. doi: 10.1038/nm.3175. URL <http://dx.doi.org/10.1038/nm.3175>.
- F. Sanger i H. Tuppy. The amino-acid sequence in the phenylalanyl chain of insulin. 2. the investigation of peptides from enzymic hydrolysates. *Biochem. J.*, 49(4): 481–490, 1951.
- Christina S. Schuetz, Michael Bonin, Susan E. Clare, Kay Nieselt, Karl Sotlar, Michael Walter, Tanja Fehm, Erich Solomayer, Olaf Riess, Diethelm Wallwiener, Raffael Kurek, i Hans J. Neubauer. Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. *Cancer Research*, 66(10):5278–5286, 2006. doi: 10.1158/0008-5472.CAN-05-4610. URL <http://cancerres.aacrjournals.org/content/66/10/5278.abstract>.
- Violeta Serra, Pieter J.A. Eichhorn, Celina García-García, Yasir H. Ibrahim, Ludmila Prudkin, Gertrudis Sánchez, Olga Rodríguez, Pilar Antón, Josep-Lluís Parra, Sara Marlow, Maurizio Scaltriti, José Pérez-García, Aleix Prat, Joaquín Arribas, William C. Hahn, So Young Kim, i José Baselga. Rsk3/4 mediate resistance to pi3k pathway inhibitors in breast cancer. *The Journal of Clinical Investigation*, 123(6):2551–2563, 6 2013. doi: 10.1172/JCI66343. URL <http://www.jci.org/articles/view/66343>.
- Jordi Serra-Musach, Helena Aguilar, Francesco Iorio, Francesc Comellas, Antoni Berenguer, Joan Brunet, Julio Saez-Rodriguez, i Miguel Angel Pujana. Cancer develops, progresses and responds to therapies through restricted perturbation of the protein-protein interaction network. *Integr. Biol.*, 4:1038–1048, 2012. doi: 10.1039/C2IB20052J. URL <http://dx.doi.org/10.1039/C2IB20052J>.
- R. Shen, A. M. Chinnaiyan, i D. Ghosh. Pathway analysis reveals functional convergence of gene expression profiles in breast cancer. *BMC medical genomics*, 1, 2008. ISSN 1755-8794. doi: 10.1186/1755-8794-1-28. URL <http://dx.doi.org/10.1186/1755-8794-1-28>.
- Yigong Shi. Serine/threonine phosphatases: Mechanism through structure. *Cell*, 139(3):468 – 484, 2009. ISSN 0092-8674. doi: <http://dx.doi.org/10.1016/j.cell.2009.10.006>. URL <http://www.sciencedirect.com/science/article/pii/S0092867409012549>.
- Brandon George Smaglo, Hongkun Wang, Kenneth Steadman, Joseph Murray, Michael Pishvaian, Aiwu Ruth He, Jimmy J. Hwang, Deepa Suresh Subramaniam, John F. Deeken, i Louis M. Weiner. A phase i study of the bcr-abl tyrosine kinase inhibitor nilotinib and cetuximab in patients with solid tumors that can be treated with cetuximab. *J Clin Oncol*, 31(TPS2624), 2013.
- Ashley G. Smart, Luis A. N. Amaral, i Julio M. Ottino. Cascading failure and robustness in metabolic networks. *Proceedings of the National Academy of Sciences*, 105(36):13223–13228, 2008. doi: 10.1073/pnas.0803571105. URL <http://www.pnas.org/content/105/36/13223.abstract>.

- Chris Stark, Bobby-Joe Breitzkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitzkreutz, i Mike Tyers. Biogrid: a general repository for interaction datasets. *Nucleic Acids Research*, 34(suppl 1):D535–D539, 2006. doi: 10.1093/nar/gkj109.
- Ulrich Stelzl, Uwe Worm, Maciej Lalowski, Christian Haenig, Felix H. Brembeck, Heike Goehler, Martin Stroedicke, Martina Zenkner, Anke Schoenherr, Susanne Koeppen, Jan Timm, Sascha Mintzlaff, Claudia Abraham, Nicole Bock, Silvia Kietzmann, Astrid Goedde, Engin Toksöz, Anja Droege, Sylvia Krobisch, Bernhard Korn, Walter Birchmeier, Hans Lehrach, i Erich E. Wanker. A human protein-protein interaction network: A resource for annotating the proteome. *Cell*, 122(6):957 – 968, 2005. ISSN 0092-8674. doi: <http://dx.doi.org/10.1016/j.cell.2005.08.029>. URL <http://www.sciencedirect.com/science/article/pii/S0092867405008664>.
- Josef Stoer, Roland Bulirsch, Richard H. Bartels, Walter Gautschi, i Christoph Witzgall. *Introduction to numerical analysis*. Texts in applied mathematics. Springer, New York, 2002. ISBN 0-387-95452-X. URL <http://opac.inria.fr/record=b1098819>.
- Michael R. Stratton, Peter J. Campbell, i P. Andrew Futreal. The cancer genome. *Nature*, 458(7239):719–724, April 2009. ISSN 0028-0836. doi: 10.1038/nature07943. URL <http://dx.doi.org/10.1038/nature07943>.
- Michael P. H. Stumpf, Carsten Wiuf, i Robert M. May. Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(12):4221–4224, 2005. doi: 10.1073/pnas.0501179102. URL <http://www.pnas.org/content/102/12/4221.abstract>.
- Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, i Jill P. Mesirov. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005. doi: 10.1073/pnas.0506580102. URL <http://www.pnas.org/content/102/43/15545.abstract>.
- Adnan S. Syed, Matteo D’Antonio, i Francesca D. Ciccarelli. Network of cancer genes: a web resource to analyze duplicability, orthology and network properties of cancer genes. *Nucleic Acids Research*, 38(suppl 1):D670–D675, 2010. doi: 10.1093/nar/gkp957. URL [http://nar.oxfordjournals.org/content/38/suppl\\_1/D670.abstract](http://nar.oxfordjournals.org/content/38/suppl_1/D670.abstract).
- Ian W. Taylor, Rune Linding, David Warde-Farley, Yongmei Liu, Catia Pesquita, Daniel Faria, Shelley Bull, Tony Pawson, Quaid Morris, i Jeffrey L. Wrana. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnology*, 27(2):199–204, February 2009. ISSN 1087-0156. doi: 10.1038/nbt.1522. URL <http://dx.doi.org/10.1038/nbt.1522>.

- the International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, February 2001. ISSN 0028-0836. doi: 10.1038/35057062. URL <http://dx.doi.org/10.1038/35057062>.
- Qiang Tian, Serguei B. Stepaniants, Mao Mao, Lee Weng, Megan C. Feetham, Michelle J. Doyle, Eugene C. Yi, Hongyue Dai, Vesteynn Thorsson, Jimmy Eng, David Goodlett, Joel P. Berger, Bert Gunter, Peter S. Linseley, Roland B. Stoughton, Ruedi Aebersold, Steven J. Collins, William A. Hanlon, i Leroy E. Hood. Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Molecular & Cellular Proteomics*, 3(10):960–969, 2004. doi: 10.1074/mcp.M400055-MCP200. URL <http://www.mcponline.org/content/3/10/960.abstract>.
- Virginia Goss Tusher, Robert Tibshirani, i Gilbert Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, 98(9):5116–5121, 2001. doi: 10.1073/pnas.091062498. URL <http://www.pnas.org/content/98/9/5116.abstract>.
- Marc J. van de Vijver, Yudong D. He, Laura J. van 't Veer, Hongyue Dai, Augustinus A.M. Hart, Dorien W. Voskuil, George J. Schreiber, Johannes L. Peterse, Chris Roberts, Matthew J. Marton, Mark Parrish, Douwe Atsma, Anke Witteveen, Annuska Glas, Leonie Delahaye, Tony van der Velde, Harry Bartelink, Sjoerd Rodenhuis, Emiel T. Rutgers, Stephen H. Friend, i René Bernards. A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine*, 347(25):1999–2009, 2002. doi: 10.1056/NEJMoa021967. PMID: 12490681.
- Wessel N. van Wieringen i Aad W. van der Vaart. Statistical analysis of the cancer cell's molecular entropy using high-throughput data. *Bioinformatics*, 27(4):556–563, 2011. doi: 10.1093/bioinformatics/btq704. URL <http://bioinformatics.oxfordjournals.org/content/27/4/556.abstract>.
- Victor E. Velculescu, Lin Zhang, Bert Vogelstein, i Kenneth W. Kinzler. Serial analysis of gene expression. *Science*, 270(5235):484–487, 1995. doi: 10.1126/science.270.5235.484. URL <http://www.sciencemag.org/content/270/5235/484.abstract>.
- Daniel V. Veres, Dávid M. Gyurkó, Benedek Thaler, Kristóf Z. Szalay, Dávid Fazekas, Tamás Korcsmáros, i Peter Csermely. CompPi: a cellular compartment-specific database for protein–protein interaction network analysis. *Nucleic Acids Research*, 2014. doi: 10.1093/nar/gku1007. URL <http://nar.oxfordjournals.org/content/early/2014/10/27/nar.gku1007.abstract>.
- Marc Vidal. A unifying view of 21st century systems biology. *{FEBS} Letters*, 583(24):3891–3894, 2009. ISSN 0014-5793. doi: <http://dx.doi.org/10.1016/j.febslet.2009.11.024>.
- Shinichiro Wachi, Ken Yoneda, i Reen Wu. Interactome-transcriptome analysis

- reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics*, 21(23):4205–4208, 2005. doi: 10.1093/bioinformatics/bti688. URL <http://bioinformatics.oxfordjournals.org/content/21/23/4205.abstract>.
- Jian-Wei Wang i Li-Li Rong. Cascade-based attack vulnerability on the {US} power grid. *Safety Science*, 47(10):1332 – 1336, 2009. ISSN 0925-7535. doi: <http://dx.doi.org/10.1016/j.ssci.2009.02.002>. URL <http://www.sciencedirect.com/science/article/pii/S0925753509000174>.
- Duncan J. Watts i Steven H. Strogatz. Collective dynamics of /‘small-world/’ networks. *Nature*, 393(6684):440–442, June 1998. ISSN 0028-0836. doi: 10.1038/30918. URL <http://dx.doi.org/10.1038/30918>.
- Britta Weigelt, Zhiyuan Hu, Xiaping He, Chad Livasy, Lisa A. Carey, Matthew G. Ewend, Annuska M. Glas, Charles M. Perou, i Laura J. van’t Veer. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Research*, 65(20):9155–9158, 2005a. doi: 10.1158/0008-5472.CAN-05-2553. URL <http://cancerres.aacrjournals.org/content/65/20/9155.abstract>.
- Britta Weigelt, Johannes L. Peterse, i Laura J. van’t Veer. Breast cancer metastasis: markers and models. *Nat Rev Cancer*, 5(8):591–602, 2005b. doi: 10.1038/nrc1670. URL <http://dx.doi.org/10.1038/nrc1670>.
- Henrica M. J. Werner, Gordon B. Mills, i Prahlad T. Ram. Cancer Systems Biology: a peek into the future of patient care? *Nat Rev Clin Oncol*, 11(3):167–176, March 2014. ISSN 1759-4774. doi: 10.1038/nrclinonc.2014.6. URL <http://dx.doi.org/10.1038/nrclinonc.2014.6>.
- I. Xenarios, D. W. Rice, L. Salwinski, M. K. Baron, E. M. Marcotte, i D. Eisenberg. DIP: the database of interacting proteins. *Nucleic acids research*, 28(1):289–291, January 2000. ISSN 0305-1048. doi: 10.1093/nar/28.1.289.
- Soon-Hyung Yook, Zoltán N. Oltvai, i Albert-László Barabási. Functional and topological characterization of protein interaction networks. *PROTEOMICS*, 4(4):928–942, 2004. ISSN 1615-9861. doi: 10.1002/pmic.200300636. URL <http://dx.doi.org/10.1002/pmic.200300636>.
- Hui Zhang, Jeanne Holden-Wiltse, Jiong Wang, i Hua Liang. A strategy to model nonmonotonic dose-response curve and estimate  $i_{c;sub;50;}/sub;_i$ . *PLoS ONE*, 8(8):e69301, 08 2013. doi: 10.1371/journal.pone.0069301. URL <http://dx.doi.org/10.1371/journal.pone.0069301>.

**Part VIII**

**Annexes**

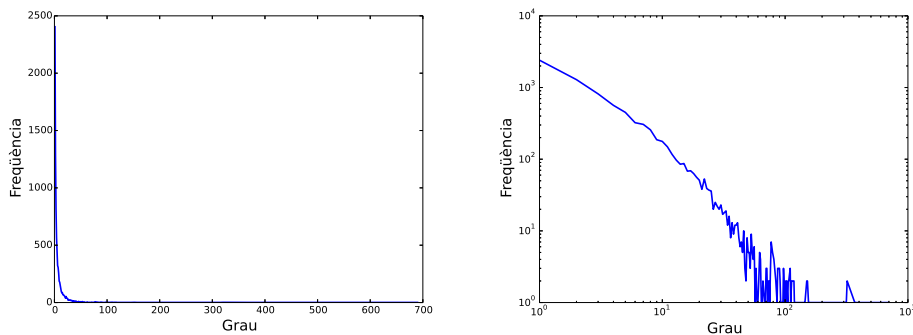




# A

## Descripció de la xarxa de l'interactoma IntAct

Les dades de *IntAct Molecular Interaction Database* (IntAct) s'han filtrat descartant els gens que no estan anotats per Entrez (Maglott et al., 2005), prenent la component principal de la xarxa obtinguda de les dades filtrades i eliminant els llaços. Finalment, s'obté la xarxa amb 8 292 vèrtexs i 33 794 arestes. El diàmetre de la xarxa és 12.

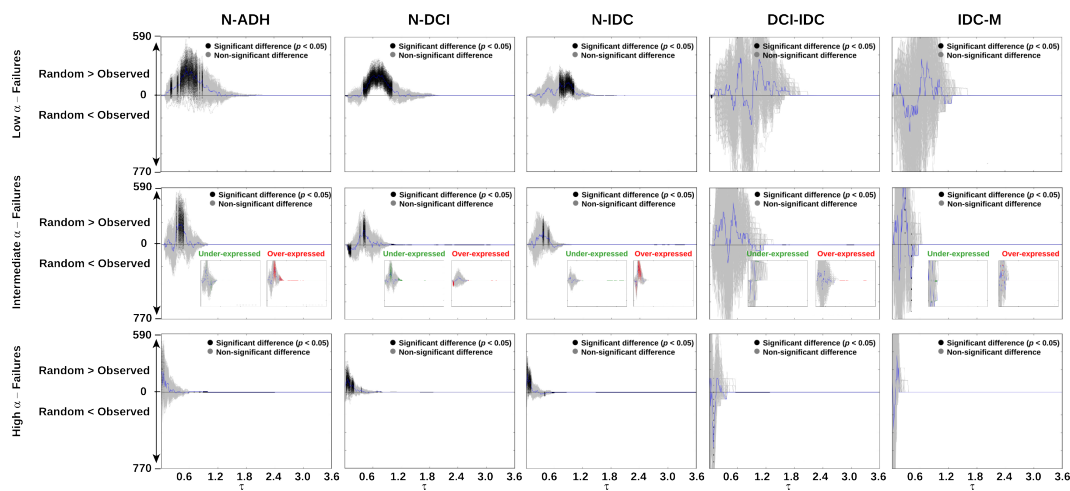


**Figura A.0.1:** La xarxa construïda amb les dades IntAct. Els gràfics mostren la distribució del grau, en escala logarítmica el gràfic de l'esquerra. L'ajustament per una funció potencial dona un exponent de -1.55.



# B

## Anàlisi topològica de la xarxa de l'interactoma IntAct



**Figura B.0.1:** Anàlisi dels patrons topològics de la xarxa IntAct respecte dels conjunts associats al desenvolupament i progressió del càncer. Els gràfics mostren la diferència del nombre de fallades entre les observades i les provocades per atzar. De dalt a baix,  $\alpha$  petita, mitjana i gran. Els patrons són molt similars als obtinguts en la xarxa HPRD



# C

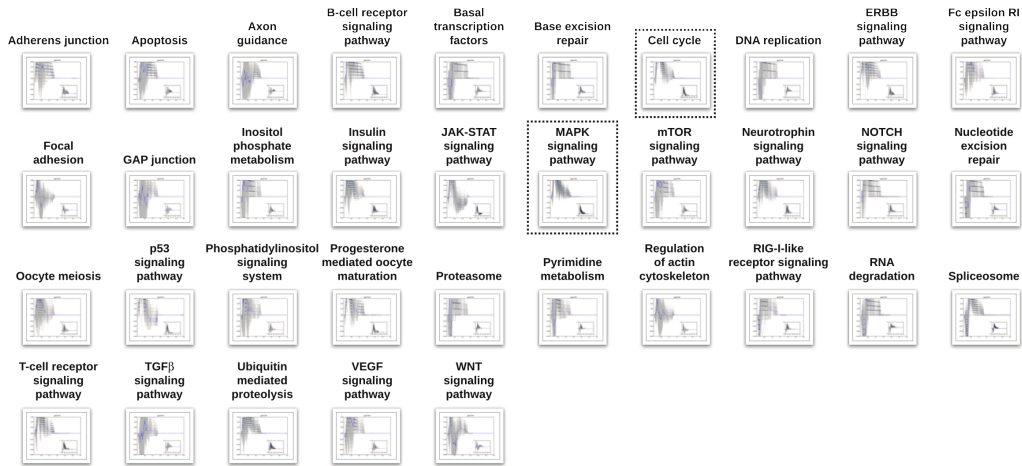
## Cascades de fallades relatives a les vies KEGG

### C.1 Vies KEGG que revelen patrons diferents als del càncer

**Figura C.1.1:** A la figura, pàgina següent, es mostren tots els gràfics del nombre de fallades, respecte les produïdes per atzar, per a totes les vies de senyalització KEGG que mostren patrons diferents als del càncer (de no robustesa). S'han marcat els casos que la diferència entre el nombre de fallades a l'atzar i observades màxima és significativa.



## C.2 Vies KEGG amb patrons similars als del càncer



**Figura C.2.1:** Es mostres tots els gràfics del nombre de fallades, respecte les produïdes per atzar, per a totes les vies de senyalització KEGG que generen patrons similars als del càncer, és a dir, de robustesa. S'han marcat els casos que la diferència entre el nombre de fallades aleatòries i observades màxima és significativa.





# D

## Classificació de fàrmacs

**Taula D.0.1:** Classificació dels fàrmacs en funció del nombre de dianes.

fàrmac	tipus	fàrmac	tipus	fàrmac	tipus
681640	Unique target	CMK	Unique target	Nutlin-3a	Unique target
17-AAG	Multitarget	Cyclopamine	Unique target	NVP-BEZ235	Unique target
A-443654	Unique target	Cytarabine	Multitarget	NVP-TAE684	Unique target
A-769662	Unique target	Dasatinib	Unique target	Obatoclox Mesylate	Multi target
A-770041	Unique target	DMOG	Unique target	OSI-906	Unique target
ABT-263	Multitarget	Docetaxel	Multitarget	PAC-1	Unique target
ABT-888	Unique target	Doxorubicin	Multitarget	Paclitaxel	Multitarget
AICAR	Unique target	Elesclomol	Multitarget	Parthenolide	Multitarget
AKT inhibitor VIII	Unique target	Embelin	Multitarget	Pazopanib	Unique target
AMG-706	Unique target	Epothilone B	Multitarget	PD-0325901	Unique target
AP-24534	Unique target	Erlotinib	Unique target	PD-0332991	Unique target
AS601245	Unique target	Etoposide	Multitarget	PD-173074	Unique target
ATRA	Unique target	FH535	Unique target	PF-02341066	Unique target
AUY922	Multitarget	FTI-277	Unique target	PF-562271	Unique target
Axitinib	Unique target	GDC-0449	Unique target	PHA-665752	Unique target
AZ628	Unique target	GDC0941	Unique target	PLX4720	Unique target
AZD-0530	Unique target	Gefitinib	Unique target	Pyrimethamine	Multitarget
AZD-2281	Unique target	Gemcitabine	Multitarget	QS11	Unique target
AZD6244	Unique target	GNF-2	Unique target	Rapamycin	Unique target
AZD6482	Unique target	GSK-650394	Unique target	RDEA119	Unique target
AZD7762	Unique target	GSK269962A	Unique target	RO-3306	Unique target
AZD8055	Unique target	GW 441756	Unique target	Roscovitine	Unique target
BAY 61-3606	Unique target	GW843682X	Unique target	S-Trityl-L-cysteine	Unique target
Bexarotene	Unique target	Imatinib	Unique target	Salubrinal	Unique target
BI-2536	Unique target	IPA-3	Unique target	SB 216763	Unique target
BI-D1870	Unique target	JNK Inhibitor VIII	Unique target	SB590885	Unique target
BIBW2992	Unique target	JNK-9L	Unique target	SL 0101-1	Unique target
Bicalutamide	Unique target	JW-7-52-1	Unique target	Shikonin	-
BIRB 0796	Unique target	KIN001-135	Unique target	Sorafenib	Unique target
Bleomycin	Multitarget	KU-55933	Unique target	Sunitinib	Unique target
BMS-509744	Unique target	Lapatinib	Unique target	Temsirolimus	Unique target
BMS-536924	Unique target	Lenalidomide	Multitarget	Thapsigargin	Unique target
BMS-754807	Unique target	LFM-A13	Unique target	Tipifarnib	Unique target
Bortezomib	Multitarget	Metformin	Unique target	Vinblastine	Multitarget
Bosutinib	Unique target	Methotrexate	Multitarget	Vinorelbine	Multitarget
Bryostatín 1	Unique target	MG-132	Multitarget	Vorinostat	Multitarget
BX-795	Unique target	Midostaurin	Multitarget	VX-680	Unique target
Camptothecin	Multitarget	Mitomycin C	Multitarget	VX-702	Unique target
CEP-701	Unique target	MK-2206	Unique target	WH-4-023	Unique target
CGP-082996	Unique target	MS-275	Multitarget	WZ-1-84	Unique target
CGP-60474	Unique target	Nilotinib	Unique target	XMD8-85	Unique target
CHIR-99021	Unique target	NSC-87877	Unique target	Z-LLNc-CHO	Unique target
CI-1040	Unique target	NU-7441	Unique target	ZM-447439	Unique target
Cisplatin	Multitarget				



# E

## Diferències entre característiques dels fàrmacs

**Taula E.0.1:** Diferències significatives entre les distribucions dels coeficients de correlació de Pearson, test de Wilcoxon amb nivell de significança 5%, entre les diferents característiques dels fàrmacs.

<b>Fàrmac classe I</b>	
Clínic	Experimental
Experimental	Quimioteràpia
Experimental	En desenvolupament clínic

<b>Fàrmac classe II</b>	
Citotòxic	amb diana

<b>Família de les dianes</b>	
Quimioteràpia	Receptor tirosina-cinasa
Quimioteràpia	CTKs
Quimioteràpia	STKs
STKs	CTKs

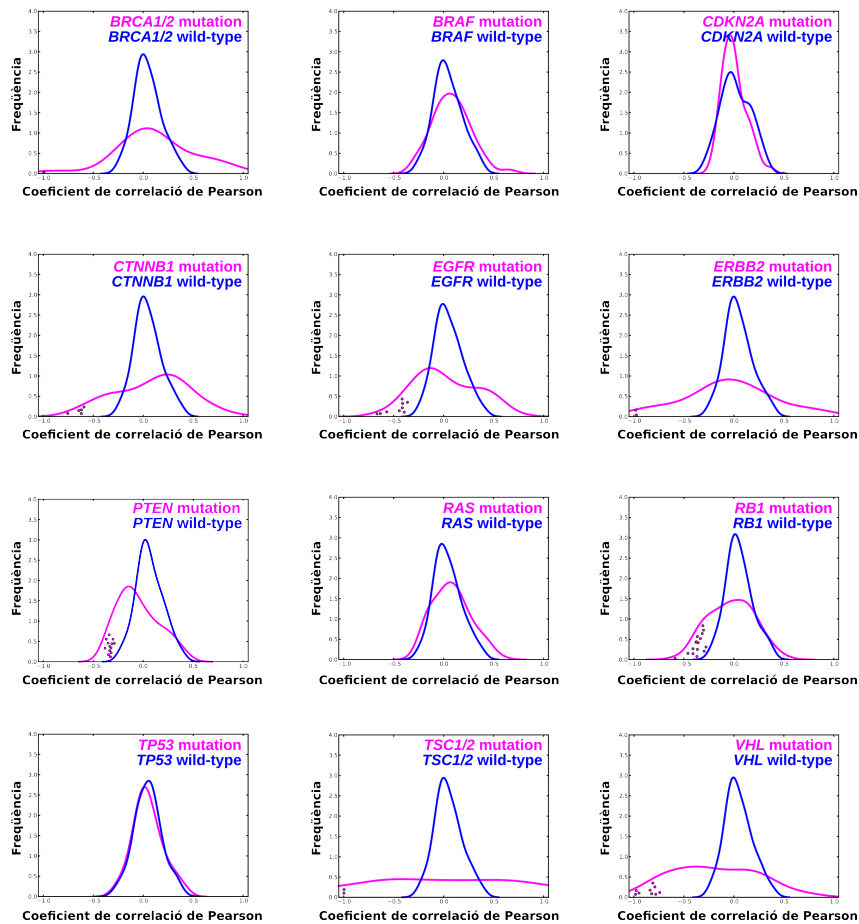
  

<b>Vies de senyalització</b>	
Reparació de l'ADN	Citoesquelet
Reparació de l'ADN	Senyalització ERK
Reparació de l'ADN	Apoptosi
Senyalització ERK	Apoptosi
Mitosi	Reparació de l'ADN
PI3K/mTOR	Reparació de l'ADN
Replicació	Citoesquelet
Replicació	Senyalització ERK
Replicació	Mitosi
Replicació	PI3K/mTOR
<i>Stress pathway</i>	Citoesquelet
<i>Stress pathway</i>	Senyalització ERK
<i>Stress pathway</i>	Mitosi
<i>Stress pathway</i>	PI3K/mTOR
Transcripció	Citoesquelet
Transcripció	Senyalització ERK
Transcripció	Mitosi
Transcripció	PI3K/mTOR



# F

## Diferències explicades per les mutacions dels gens



**Figura F.0.1:** Diferències entre els fàrmacs respecte de tots els gens conductors del càncer considerats.



# G

Fàrmacs associats a l'estatus de mutació dels  
gens

**Taula G.0.1:** Fàrmacs amb correlacions negatives entre les  $NA$  i les  $IC_{50}$  ( $p < 0.1$ )

Gene/Compound	PCC	<i>p</i> value	Gene/Compound	PCC	<i>p</i> value
<b>CTNNB1</b>			<b>BRCA1/2</b>		
SB590885	-0.76	3.02E-02	CHIR-99021	-0.99	7.30E-02
JNK Inhibitor VIII	-0.64	6.08E-02			
PD-173074	-0.62	7.53E-02	<b>PTEN</b>		
AMG-706	-0.62	7.62E-02	CEP-701	-0.37	3.12E-03
BIRB 0796	-0.59	9.30E-02	AICAR	-0.34	5.15E-03
			ATRA	-0.35	8.04E-03
<b>EGFR</b>			BI-D1870	-0.33	8.99E-03
Rapamycin	-0.67	3.53E-02	ABT-888	-0.33	9.50E-03
Nilotinib	-0.57	1.10E-02	AZD8055	-0.32	9.78E-03
KIN001-135	-0.63	4.92E-02	AZD-2281	-0.32	1.11E-02
GDC-0449	-0.44	5.69E-02	GDC-0449	-0.31	1.29E-02
VX-702	-0.42	7.68E-02	JNK-9L	-0.38	1.38E-02
ATRA	-0.39	9.47E-02	VX-702	-0.29	2.19E-02
PD-173074	-0.39	9.78E-02	GSK-650394	-0.34	2.55E-02
AZD6244	-0.41	9.85E-02	PF-562271	-0.35	2.62E-02
			MK-2206	-0.29	2.64E-02
<b>ERBB2</b>			681640	-0.28	2.95E-02
FH535	-0.70	3.54E-02	Bexarotene	-0.30	5.31E-02
ZM-447439	-0.52	5.75E-02	Tipifarnib	-0.27	7.51E-02
			AKT inhibitor VIII	-0.28	7.81E-02
<b>PIK3CA/R1</b>			BMS-754807	-0.27	8.02E-02
SL 0101-1	-0.35	1.32E-02			
AZD-2281	-0.34	1.49E-02	<b>RB1</b>		
AKT inhibitor VIII	-0.33	3.36E-02	Thapsigargin	-0.59	1.59E-03
MK-2206	-0.30	3.88E-02	AZD-2281	-0.41	5.26E-03
Metformin	-0.27	3.93E-02	SL 0101-1	-0.38	1.13E-02
AZD8055	-0.28	4.73E-02	Nutlin-3a	-0.36	1.58E-02
IPA-3	-0.28	6.73E-02	SB590885	-0.37	1.71E-02
Rapamycin	-0.45	8.23E-02	Bexarotene	-0.46	1.81E-02
			GSK-650394	-0.41	3.80E-02
			ATRA	-0.31	4.00E-02
			VX-702	-0.30	4.25E-02
			SB 216763	-0.30	4.71E-02
			Tipifarnib	-0.38	5.90E-02
			AS601245	-0.37	6.28E-02
			GDC-0449	-0.28	6.35E-02
			FH535	-0.36	6.69E-02
			ABT-888	-0.27	7.51E-02
			PF-562271	-0.34	9.40E-02
			<b>TSC1/2</b>		
			Temsirolimus	-1.00	3.88E-02
			AZD7762	-1.00	5.18E-02
			<b>VHL</b>		
			CEP-701	-0.95	3.74E-03
			Bosutinib	-0.84	3.70E-02
			Nilotinib	-0.82	4.42E-02
			ZM-447439	-0.79	6.12E-02
			Temsirolimus	-0.78	6.63E-02
			AZD7762	-0.74	9.15E-02
			BIBW2992	-0.74	9.45E-02
			Axitinib	-0.74	9.56E-02
			Dasatinib	-0.81	9.59E-02



**Part IX**

**Currículum**



## BIOGRAPHICAL SKETCH

<b>NAME:</b> Jordi Serra	<b>POSITION TITLE:</b> PhD Student
<b>ADDRESS:</b> Breast Cancer and Systems Biology Lab Catalan Institute of Oncology, IDIBELL L'Hospitalet del Lobregat 08908, Barcelona, Spain Ph.: +34-9326074500 (lab) / +34-610797159 (cell) E-mail: <a href="mailto:jmusach@gmail.com">jmusach@gmail.com</a> <a href="http://cancersystemsbiology.cat/index.php/group/students/jordi-serra">http://cancersystemsbiology.cat/index.php/group/students/jordi-serra</a>	Breast Cancer and Systems Biology Lab Catalan Institute of Oncology, IDIBELL L'Hospitalet de Llobregat, Barcelona, Spain <a href="http://cancersystemsbiology.cat">http://cancersystemsbiology.cat</a>

## EDUCATION / TRAINING

INSTITUTION AND LOCATION	DEGREE (if applicable)	YEAR(s)	FIELD OF STUDY
University of Barcelona	B.Sc.	2007	Mathematics
Polytechnic University of Catalonia	M.Sc.	2010	Modeling in biosciences
Catalan Institute of Oncology, L'Hospitalet, Barcelona, Spain	Ph.D.	2015	Cancer Systems Biology

## A. Positions and Honors

- 2008 - present PhD student, Breast Cancer Systems biology. Supported by RTICC Breast Cancer.
- 2012 Research stay at the European Bioinformatics Institute (Saez-Rodriguez research group) for three months awarded by Ferran Sunyer i Balaguer foundation.

## B. Peer-reviewed Publications (in chronological order)

- M. Gustafsson, D. R. Gawel, L. Alfredsson, S. Baranzini, J. Björkander, R. Blomgran, S. Hellberg, D. Eklund, J. Ernerudh, I. Kockum, A. Konstantinell, R. Lahesmaa, A. Lentini, H. R. I. Liljenström, L. Mattson, A. Matussek, J. Mellergård, M. Mendez, T. Olsson, M. A. Pujana, O. Rasool, **J. Serra-Musach**, M. Stenmarker, S. Tripathi, M. Vitala, H. Wang, H. Zhang, C. E. Nestor, and M. Benson. A validated gene regulatory network and gwas identifies early regulators of t cell-associated diseases. *Science Translational Medicine*, 7(313):313ra178–313ra178, 2015.
- G. Ruiz de Garibay, C. Herranz, A. Llorente, J. Boni, **J. Serra-Musach**, F. Mateo, H. Aguilar, L. Gómez-Baldó, A. Petit, A. Vidal, F. Climent, J. Hernández-Losa, Á. Cordero, E. González-Suárez, J. V. Sánchez-Mut, M. Esteller, R. Llatjós, M. Varela, J. I. López, N. García, A. I. Extremera, A. Gumà, R. Ortega, M. J. Plà, A. Fernández, S. Pernas, C. Faló, I. Morilla, M. Campos, M. Gil, A. Román, M. Molina-Molina, P. Ussetti, R. Laporta, C. Valenzuela, J. Ancochea, A. Xaubet, Á. Casanova, and M. A. Pujana. Lymphangioliomyomatosis biomarkers linked to lung metastatic potential and cell stemness. *PLoS ONE*, 10(7):e0132546, 07 2015.
- B. Barneda-Zahonero, O. Collazo, A. Azagra, I. Fernandez-Duran, **J. Serra-Musach**, A. B. M. M. K. Islam, N. Vega-Garcia, R. Malatesta, M. Camos, A. Gomez, L. Roman-Gonzalez, A. Vidal, N. Lopez-Bigas, A. Villanueva, M. Esteller, and M. Parra. The transcriptional repressor hdac7 promotes apoptosis and c-myc downregulation in particular types of leukemia and lymphoma. *Cell Death Dis*, 6:e1635, 2015.
- H. Aguilar, A. Urruticoechea, P. Halonen, K. Kiyotani, T. Mushiroda, X. Barril, **J. Serra-Musach**, A. Islam, L. Caizzi, L. Di Croce, E. Nevedomskaya, W. Zwart, J. Bostner, E. Karlsson, G. Perez Tenorio, T. Fornander, D. Sgroi, R. Garcia-Mata, M. Jansen, N. Garcia, N. Bonifaci, F. Climent, M. Soler, A. Rodriguez-Vida, M. Gil, J. Brunet, G. Martrat, L. Gomez-Baldo, A. Extremera, A. Figueras, J. Balart, R. Clarke, K. Burnstein, K. Carlson, J. Katzenellenbogen, M. Vizoso, M. Esteller, A. Villanueva, A. Rodriguez-Pena, X. Bustelo, Y. Nakamura, H. Zembutsu, O. Stal, R. Beijersbergen, and M. Pujana. VAV3 mediates resistance to breast cancer endocrine therapy. *Breast Cancer Research*, 16(3):R53, 2014.
- M. Gustafsson, M. Edstrom, D. Gawel, C. Nestor, H. Wang, H. Zhang, F. Barrenas, J. Tojo, I. Kockum, T. Olsson, **J. Serra-Musach**, N. Bonifaci, M. Pujana, J. Ernerudh, and M. Benson. Integrated genomic and prospective

- clinical studies show the importance of modular pleiotropy for disease susceptibility, diagnosis and treatment. *Genome Medicine*, 6(2):17, 2014.
6. N. Bonifaci, E. Colas, **J. Serra-Musach**, N. Karbalai, J. Brunet, A. Gómez, M. Esteller, E. Fernández-Taboada, A. Berenguer, J. Reventós, B. Müller-Myhsok, L. Amundadottir, E. J. Duell, and M. À. Pujana. Integrating gene expression and epidemiological data for the discovery of genetic interactions associated with cancer risk. *Carcinogenesis*, 35(3):578–585, 2014.
  7. T. Cokelaer, D. Pultz, L. M. Harder, **J. Serra-Musach**, and J. Saez-Rodriguez. Bioservices: a common python package to access biological web services programmatically. *Bioinformatics*, 2013.
  8. **J. Serra-Musach**, H. Aguilar, F. Iorio, F. Comellas, A. Berenguer, J. Brunet, J. Saez-Rodriguez, and M. A. Pujana. Cancer develops, progresses and responds to therapies through restricted perturbation of the protein-protein interaction network. *Integr. Biol.*, 4:1038–1048, 2012.
  9. H. Heyn, N. Li, H. J. Ferreira, S. Moran, D. G. Pisano, A. Gomez, J. Diez, J. V. Sanchez-Mut, F. Setien, F. J. Carmona, A. A. Puca, S. Sayols, M. A. Pujana, **J. Serra-Musach**, I. Iglesias-Platas, F. Formiga, A. F. Fernandez, M. F. Fraga, S. C. Heath, A. Valencia, I. G. Gut, J. Wang, and M. Esteller. Distinct dna methylomes of newborns and centenarians. *Proceedings of the National Academy of Sciences*, 109(26):10522–10527, 2012.
  10. R. Sanz-Pamplona, A. Berenguer, X. Sole, D. Cordero, M. Crous-Bou, **J. Serra-Musach**, E. Guinó, M. Pujana, and V. Moreno. Tools for protein-protein interaction network analysis in cancer research. *Clinical and Translational Oncology*, 14:3–14, 2012.
  11. C. A. Maxwell, J. Benítez, L. Gómez-Baldó, A. Osorio, N. Bonifaci, R. Fernández-Ramires, S. V. Costes, E. Guinó, H. Chen, G. J. R. Evans, P. Mohan, I. Català, A. Petit, H. Aguilar, A. Villanueva, A. Aytes, **J. Serra-Musach**, G. Rennert, F. Lejbkowitz, P. Peterlongo, S. Manoukian, B. Peissel, C. B. Ripamonti, B. Bonanni, A. Viel, A. Allavena, L. Bernard, P. Radice, E. Friedman, B. Kaufman, Y. Laitman, M. Dubrovsky, R. Milgrom, A. Jakubowska, C. Cybulski, B. Gorski, K. Jaworska, K. Durda, G. Sukiennicki, J. Lubiński, Y. Y. Shugart, S. M. Domchek, R. Letrero, B. L. Weber, F. B. L. Hogervorst, M. A. Rookus, J. M. Collee, P. Devilee, M. J. Ligtenberg, R. B. van der Luijt, C. M. Aalfs, Q. Waisfisz, J. Wijnen, C. E. P. van Roozendaal, D. F. Easton, S. Peock, M. Cook, C. Oliver, D. Frost, P. Harrington, D. G. Evans, F. Lalloo, R. Eeles, L. Izatt, C. Chu, D. Eccles, F. Douglas, C. Brewer, H. Nevanlinna, T. Heikkinen, F. J. Couch, N. M. Lindor, X. Wang, A. K. Godwin, M. A. Caligo, G. Lombardi, N. Loman, P. Karlsson, H. Ehrencrona, A. von Wachenfeldt, R. Bjork Barkardottir, U. Hamann, M. U. Rashid, A. Lasa, T. Caldés, R. Andrés, M. Schmitt, V. Assmann, K. Stevens, K. Offit, J. Curado, H. Tilgner, R. Guigó, G. Aiza, J. Brunet, J. Castellsagué, G. Martrat, A. Urruticoechea, I. Blanco, L. Tihomirova, D. E. Goldgar, S. Buys, E. M. John, A. Miron, M. Southey, M. B. Daly, R. K. Schmutzler, B. Wappenschmidt, A. Meindl, N. Arnold, H. Deissler, R. Varon-Mateeva, C. Sutter, D. Niederacher, E. Imyamtov, O. M. Sinilnikova, D. Stoppa-Lyonne, S. Mazoyer, C. Verny-Pierre, L. Castera, A. de Pauw, Y.-J. Bignon, N. Uhrhammer, J.-P. Peyrat, P. Vennin, S. Fert Ferrer, M.-A. Collonge-Rame, I. Mortemousque, A. B. Spurdle, J. Beesley, X. Chen, S. Healey, M. H. Barcellos-Hoff, M. Vidal, S. B. Gruber, C. Lázaro, G. Capellá, L. McGuffog, K. L. Nathanson, A. C. Antoniou, G. Chenevix-Trench, M. C. Fleisch, V. Moreno, and M. A. Pujana. Interplay between *brc1* and *rhamm* regulates epithelial apicobasal polarization and may influence risk of breast cancer. *PLoS Biol*, 9(11):e1001199, 11 2011.
  12. J. Sandoval, H. Heyn, S. Moran, **J. Serra-Musach**, M. Pujana, M. Bibikova, and M. Esteller. Validation of a dna methylation microarray for 450,000 cpg sites in the human genome. *Epigenetics*, 6:692 – 702, 2011.
  13. N. Bonifaci, B. Górski, B. Masojć, D. Wokolorczyk, A. Jakubowska, T. Dębniak, A. Berenguer, **J. Serra Musach**, J. Brunet, J. Dopazo, S. A. Narod, J. Lubiński, C. Lázaro, C. Cybulski, and M. A. Pujana. Exploring the link between germline and somatic genetic alterations in breast carcinogenesis. *PLoS ONE*, 5(11):e14078, 11 2010.
  14. H. Aguilar, X. Sole, N. Bonifaci, **J. Serra-Musach**, A. Islam, N. Lopez-Bigas, M. Mendez-Pertuz, R. L. Beijersbergen, C. Lázaro, A. Urruticoechea, and P. M. A. Biological reprogramming in acquired resistance to endocrine therapy of breast cancer. *Oncogene*, 29:6071–6083, 2010.