

II. PREMISSA

Fins ara hem vist la gran varietat d'estudis dirigits a intentar esbrinar quins són els processos i canvis que pateix una cèl·lula normal perquè es transformi en una cèl·lula tumoral, i com aquests coneixements ens poden ajudar a pronosticar el curs evolutiu de la malaltia.

El problema cabdal del càncer és com simplificar un tema tan genèric per poder abordar-lo de la millor manera. Les línies cel·lulars i els models animals són una ajuda molt important, però mai deixen de ser models que sovint no són aplicables a les cèl·lules humanes *in vivo*. Sí que ens ajuden a l'estudi d'una via concreta, però després cal situar aquell procés dins d'un entramat de vies paral·leles, que s'entrecreuen, o que compensen aquell efecte. És per tant molt més complicat del que podem dibuixar sobre un paper amb fletxes amunt i avall.

Trobats en aquest punt, la nostra manera d'intentar clarificar el panorama, dins del càncer colorectal almenys, va ser abordar-lo des de dos punts de vista diferents, però complementaris. Per una banda, volíem veure quines eren les alteracions a nivell de DNA de forma general, és a dir, a nivell dels cromosomes, no pas de les mutacions concretes. Per això ens vam aprofitar de la tècnica de la CGH, per estudiar així les alteracions a nivell cromosòmic. Per altra banda, ens vam centrar en l'estudi dels patrons d'expressió gènica. En aquest cas vam utilitzar la tècnica de les microarrays per veure les alteracions a nivell transcripcional.

El nostre material d'estudi eren biòpsies humanes de càncer colorectal de pacients sense antecedents familiars. La quantitat de mostra era limitada, i per això ens vam haver de restringir a un estalvi màxim del material a utilitzar. Les dues tècniques esmentades són molt adients perquè amb una quantitat relativament petita de mostra obtenim molta informació de l'estat del DNA i l'RNA respectivament. Són aproximacions globals, per respondre a qüestions complexes. Per tant, la nostra manera d'enfocar el problema serà des d'un punt de vista general, amb la intenció de poder donar explicació al comportament del conjunt de tumors a partir d'una gran quantitat d'informació que cal destriar molt bé.

L'empresa semblava molt ambiciosa, per això ens vam centrar en un grup de 50 mostres de càncer colorectal esporàdics, dels quals es disposava de l'historial clínic i patològic, així com de la presència o no de mutacions dels gens TP53 i KRAS.

III. OBJECTIUS

Objectiu general

Obtenir una classificació molecular dels tumors colorectals humans amb una possible aplicació pronòstica.

Objectius específics

Caracterització genòmica

1. Analitzar les alteracions a nivell cromosòmic d'una sèrie de tumors colorectals, i correlacionar-les amb les característiques clinicopatològiques.
2. Classificar els tumors en funció de les alteracions cromosòmiques per tal d'identificar subgrups amb un comportament diferent.

Caracterització transcripcional

3. Aplicació de la RAP-PCR com a mètode d'amplificació de l'RNA per hibridar les mostres sobre microarrays de cDNA.
4. Classificar els tumors en funció de les diferències d'expressió gèniques per tal d'identificar subgrups amb un comportament diferent.

Marcadors específics de pronòstic

5. Buscar nous marcadors moleculars de pronòstic del càncer colorectal.

IV. MATERIALS I MÈTODES

4.1. Pacients i tipus de mostra

Tots els teixits estudiats procedeixen de pacients diagnosticats de càncer colorectal a l'hospital de Bellvitge de l'Hospitalet del Llobregat entre els anys 1996 i el 1998. Els criteris d'inclusió foren els següents:

- adenocarcinomes primaris intervinguts de forma electiva (cirurgia no urgent) i recessió complerta del tumor.
- obtenció de mostres fresques de mucosa tumoral i la seva corresponent normal aparellada, en un interval inferior a les dues hores després de l'extirpació del tumor.
- pacients sense mort postoperatòria i amb un seguiment mínim de 4 anys.

Per tant, de tots ells es va obtenir tant mucosa normal com tumoral. Els teixits es recollien en fresc al quiròfan i es congelaven a $-80\text{ }^{\circ}\text{C}$ inclosos en OCT (Tissue-Tek[®] O.C.T.[™] Compound, Sakura Finetek Europe B. V.). Al Departament d'Anatomia Patològica del mateix hospital es va fer l'estudi patològic dels tumors. I el seguiment dels pacients es va dur a terme a l'hospital Duran i Reynals, també de l'Hospitalet.

De tota la sèrie del banc de tumors es van seleccionar 50 casos amb les següents característiques:

	característica	número de pacients	%
estadi de Dukes	B1	5	10
	B2	25	50
	C1	2	4
	C2	18	36
sexe	home	26	52
	dona	24	48
TP53	normal	33	66
	mutat	17	34
KRAS	normal	17	34
	mutat	33	66
localització	recte	20	40
	còlon esquerre	15	30
	còlon dret	15	30
teràpia	quimioteràpia	11	22
	radioteràpia	1	2
	quimio + radio	9	18
	sense tractar	29	58
recidiva	sí	21	42
	no	29	58
supervivència	mort	15	30
	viu	35	70

Taula 9. Resum dels 50 casos de càncer colorectal seleccionats.

L'edat mitjana dels 50 pacients és de 65 anys ($\pm 13,25$). Es van seleccionar només casos de Dukes B i C, i 29 d'ells no van seguir tractament quimioteràpic ni radioteràpic. A continuació es detallen les característiques clinicopatològiques de tots els casos:

Taula 10. Detall de les característiques clinicopatològiques dels 50 casos escollits.

nº cas	edat ⁽¹⁾	sexe	localització	estadi de Dukes ⁽²⁾	quimioteràpia	radioteràpia	sense recidiva després de 44 mesos	MSI ⁽³⁾
9	67	home	dret	B2	no	no	sí	sí
11	54	home	esquerre	B2	sí	sí	sí	no
12	55	home	esquerre	B2	sí	no	sí	no
14	64	dona	esquerre	B1	no	no	sí	no
16	70	dona	esquerre	C2	no	no	sí	no
17	31	dona	esquerre	C2	sí	sí	sí	no
19	49	dona	esquerre	B1	no	no	no	no
21	68	dona	esquerre	B1	no	no	sí	no
23	65	home	esquerre	B2	no	no	sí	no
25	67	home	esquerre	B1	no	no	sí	no
46	68	home	dret	B2	no	no	sí	no
53	61	dona	esquerre	C2	sí	sí	sí	no
63	74	dona	esquerre	C2	no	no	sí	no
67	56	home	esquerre	C2	sí	sí	sí	no
69	52	home	dret	B2	sí	sí	sí	sí
72	73	home	dret	C2	no	no	no	no
74	86	home	esquerre	C2	no	no	sí	no
75	76	home	esquerre	B2	sí	no	no	no
78	62	home	esquerre	B2	sí	no	sí	no
81	67	dona	dret	B2	no	no	sí	no
90	88	dona	esquerre	C2	no	no	no	no
99	65	home	dret	B2	no	no	no	sí
102	78	home	dret	B2	no	no	no	no
103	80	home	esquerre	B2	no	no	no	no
104	62	dona	dret	C2	sí	no	no	no
108	61	home	esquerre	C2	sí	no	no	no
109	43	dona	dret	B2	sí	no	sí	no
113	73	dona	dret	B2	sí	no	no	no
114	76	dona	dret	B2	no	no	sí	no
118	51	dona	esquerre	C2	sí	sí	no	no
122	62	home	esquerre	C1	no	no	sí	no
124	74	dona	esquerre	C2	sí	sí	sí	no
127	30	dona	esquerre	C2	sí	no	sí	no
135	63	dona	esquerre	B2	sí	sí	sí	no
138	80	dona	esquerre	B2	no	no	sí	no
143	37	dona	dret	C1	sí	no	sí	no
144	72	dona	esquerre	B2	no	no	no	no
145	73	dona	dret	B2	no	no	sí	no
146	65	dona	esquerre	B1	no	no	sí	no
147	76	dona	esquerre	C2	no	no	no	no
151	79	home	dret	B2	no	no	no	no
154	54	home	esquerre	C2	sí	sí	no	no
165	76	home	dret	C2	no	no	no	no
171	85	home	esquerre	C2	no	no	sí	no
175	46	dona	esquerre	B2	sí	sí	no	no
186	56	home	esquerre	B2	no	no	no	no
187	63	home	esquerre	B2	no	no	no	no
223	75	home	esquerre	B2	no	no	no	no
271	67	home	esquerre	B2	sí	no	no	no
282	78	home	esquerre	C2	no	sí	no	no

(1) edat en anys

(2) segons la classificació d'Astler-Coller

(3) inestabilitat de microsatèl·lits

4.2. Extracció de DNA

L'extracció de DNA es va fer seguint el protocol convencional d'extracció amb fenol/cloroform (Sambrook et al., 1989) a partir del teixit congelat. El DNA obtingut es valorava en un gel d'agarosa al 0,8% per comprovar la seva qualitat i concentració. Es diluïa a una concentració final d'uns 100 ng/ μ l, que era adient per després fer el marcatge per la CGH. Els mateixos casos es van diluir a 20 ng/ μ l per fer la PCR quantitativa.

4.3. Microdissecció

Prèvia a l'extracció de l'RNA s'enriqueix el teixit tumoral mitjançant microdissecció. Es feia amb l'objectiu de seleccionar la zona del tumor on hi ha un tant per cent més elevat de cèl·lules tumorals, evitant així les zones necròtiques o amb un elevat component inflamatori. Resumint es tracta de marcar l'àrea d'interès de la secció del bloc d'OCT i descartar la resta. Per això vam seguir la següent estratègia:

- es feia un tall de 5 μ m i es tenyia amb hematoxilina i eosina per fer-ne la valoració patològica.
- el patòleg seleccionava l'àrea d'interès del tall i macroscòpicament se solapava damunt del bloc per marcar la mateixa àrea i es feien 10 talls de 50 μ m cada un.
- es tornava a fer un control de la morfologia de la secció mitjançant un tall de 5 μ m i una nova tinció d'hematoxilina i eosina que valorava el patòleg.
- es tornava a seleccionar l'àrea d'interès i es feien 10 talls més de 50 μ m.
- prèviament a l'extracció de l'RNA es feien 2 rentats amb 1 ml de PBS 1 \times (de l'anglès, *phosphate-buffered saline*) (KH₂PO₄ 1.47mM, Na₂HPO₄ \times 7H₂O 4.29 mM, NaCl 137 mM, KCl 2.68 mM) i centrifugant un minut a 1000 rpm i a 4°C, amb l'objectiu d'eliminar les restes d'OCT.

4.4. Extracció de l'RNA total

L'extracció de l'RNA total es va fer seguint el protocol convencional d'extracció amb tiocianat de guanidina i fenol/cloroform (Chomczynski i Sacchi, 1987). El material de partida eren els 20 talls de biòpsies de còlon de 50 μ m de gruix. La disgregació mecànica es feia mitjançant una xeringa d'1 ml i 25 GA 5/8 (BD Plastipak™).

A continuació es purifica l'RNA amb el kit *RNeasy® Mini Kit* (Qiagen Inc.). Vam seguir el protocol de *l'RNA Cleanup*, amb l'objectiu d'eliminar els fragments d'RNA degradats, així com possibles restes dels reactius d'extracció. El protocol es basa en una columna que conté una membrana de silicagel, on s'hi uneix l'RNA específicament. Després es renten els contaminants i es dilueix l'RNA amb aigua.

Finalment es resolien els RNAs en un gel d'agarosa al 0,8% per valorar la seva qualitat. Si era correcta es procedia a la quantificació a l'espectrofotòmetre i es diluïen a 5 ng/ μ l en aigua.

4.5. Hibridació genòmica comparada (CGH, de l'anglès *Comparative genomic hybridization*)

a) Introducció a la tècnica

La CGH és una hibridació *in situ* damunt de cromosomes metafàsics de quantitats iguals de DNA total de dues mostres marcades de diferent manera (Kallioniemi et al., 1994; Kallioniemi et al., 1993). Es tracta de marcar el DNA d'una mescla de cèl·lules tumorals amb un fluorocrom, i el DNA d'una mostra de referència amb un altre fluorocrom, i s'hibriden sobre cromosomes metafàsics normals. Així podem visualitzar i mapar sobre els cromosomes la quantitat relativa del DNA del tumor respecte la referència. L'objectiu d'aquesta tècnica és identificar regions cromosòmiques amb un increment del número de còpies de DNA (guanys o amplifacions), i d'altres amb una reducció del número de còpies (pèrdues). D'aquesta manera es poden reconèixer regions dels cromosomes on hi hagi possibles oncogens (regions amb guanys), o d'altres on es pot pensar la presència putativa de gens supressors de tumors (regions amb pèrdues) (Figura 12).

La mesura de la senyal de fluorescència resultant de la hibridació es fa mitjançant un sistema d'anàlisi de la imatge digital. Amb l'ajuda d'un *software* obtenim les intensitats netes dels dos fluorocroms, després de restar-li la senyal de fons, així com la raó de les intensitats netes al llarg de tots els cromosomes, des de l'1p fins l'Xq. La resolució de la tècnica és de 10 Mb.

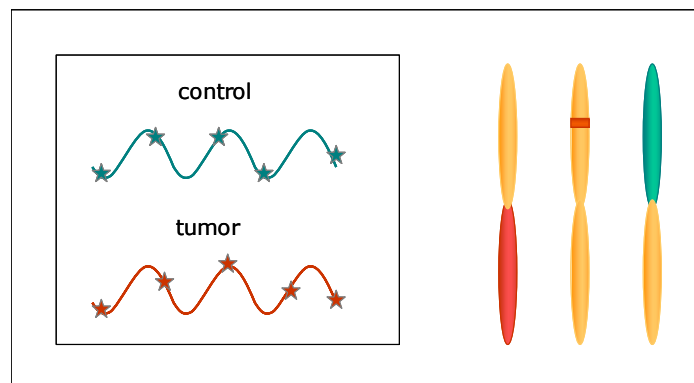


Figura 12. Esquema de la CGH (hibridació genòmica comparada). La tècnica es basa en la hibridació de quantitats iguals de DNA normal (en aquest cas de color verd) i tumoral (en aquest cas de color vermell). Així, quan hi ha la mateixa quantitat de la mostra d'estudi i del control, veurem els cromosomes tenyits de color groguenc, quan predomini el DNA tumoral, es veurà de color vermell (i parlarem de guany cromosòmic), i quan hi hagi menys quantitat de la mostra tumoral en relació a la normal es veurà de color verd (i parlarem de pèrdua cromosòmica).

La CGH es basa en una tècnica anterior com és el FISH (de l'anglès, *Fluorescence in situ hybridization*), però amb un abordatge totalment invers. Així, en un FISH estem marcant una sonda de DNA específica i coneguda i l'hibridem sobre cromosomes de cèl·lules tumorals. En canvi amb la CGH, hibridem tot el DNA d'una mostra, format per seqüències anònimes i desconegudes sobre cromosomes metafàsics normals. D'aquesta manera tenim informació de l'estat global del genoma de la mostra en estudi, obtenint un perfil de totes les regions cromosòmiques guanyades i perdudes. Amb l'actual coneixement de la seqüència del genoma humà i de la seva localització subcromosòmica, aquesta informació ens pot ser molt útil per tal d'identificar quins gens hi ha en les regions alterades. Com que la resolució de la tècnica és de 10 Mb, serà necessari buscar el gen o gens candidats emprant altres tècniques com poden ser el FISH, o la PCR quantitativa. És un abordatge una mica arriscat, perquè en un fragment de 10 Mb hi poden haver molts gens. Però d'aquesta manera, si es confirma el guany, amplificació, o pèrdua d'un gen determinat es pot postular la relació entre aquest i el procés tumoral.

Aplicacions

- caracterització de llocs recurrents (o llocs calents, de l'anglès *hot spots*) de reorganitzacions genòmiques desequilibrades.
- definició de nous gens implicats en el procés tumoral (Sakakura et al., 2001).
- anàlisi de la progressió i l'evolució clonal dels tumors (Heselmeyer et al., 1996).
- classificació dels tumors (Chibon et al., 2003).
- estudi i caracterització de models per a l'estudi del càncer com són les línies cel·lulars existents de diferents tipus de tumors (Ribas et al., 2003).
- diagnòstic de determinats tipus de tumors. És així en leucèmies i limfomes, perquè gràcies a la CGH es poden detectar canvis quan l'activitat mitòtica d'aquestes cèl·lules és baixa i no es podrien estudiar mitjançant la citogenètica convencional. Cal assegurar-se que més del 50-70% de les cèl·lules de la mostra són d'origen tumoral (Ried et al., 1999).
- pronòstic dels tumors: els resultats de CGH s'han correlacionat amb la supervivència dels pacients, arribant a la conclusió que un major nombre d'alteracions genètiques resulta amb una major agressivitat del tumor (Visakorpi et al., 1995).

Avantatges

- només calen 800 ng de DNA.

- la dificultat en l'estudi dels tumors sòlids mitjançant la citogenètica clàssica degut a que és molt difícil o impossible d'obtenir metafases, es pot solucionar gràcies a la CGH. Aquests tumors presenten moltes reorganitzacions desequilibrades, per tant aquesta tècnica ens permet estudiar el seu perfil genètic. Podem parlar de l'obtenció d'un *fingerprint* (o empremta genètica) dels cromosomes per cada mostra.
- la possibilitat d'estudiar un gran nombre de mostres, i fer un *screening* (o cribratge) dels canvis en el número de còpies de seqüències de DNA de sèries de tumors grans, degut a la facilitat de la tècnica.
- proporciona el mapatge directe d'aquelles regions guanyades o perdudes.

Inconvenients

- degut a l'heterogeneïtat del tumor pot ser que només unes quantes cèl·lules presentin canvis en el número de còpies. Si aquests guanys o pèrdues no estan presents en més del 50% de les cèl·lules no seran detectables per la tècnica de la CGH (Ried et al., 1999).
- el mateix problema tindrem si hi ha contaminació de teixit normal, cosa que es pot evitar mitjançant l'enriquiment de cèl·lules tumorals per microdissecció.
- tampoc podem detectar les translocacions equilibrades, inversions, ni reorganitzacions cromosòmiques petites, ja que només estem estudiant el contingut total de DNA d'una barreja de cèl·lules tumorals en referència a un control normal.
- no ens informa de la ploïdia de les cèl·lules en estudi, ja que com s'ha dit estem mirant el contingut total de DNA en relació al control, i en la preparació de les mostres s'igualen les quantitats de DNA del cas d'estudi respecte la referència per tal que siguin comparables. En aquest cas la millor solució és utilitzar altres tècniques alternatives amb aquesta finalitat, com són la citogenètica convencional o la citometria de flux.
- les regions centromèriques, i les regions heterocromàtiques dels cromosomes 1, 9, 16, 19, i 22 no es poden valorar degut a l'elevat polimorfisme en el número de còpies entre diferents individus.

b) Protocol

Per dur a terme una hibridació de CGH necessitem marcar les dues mostres control i tumor (sondes) per una banda, i preparar els portes amb les extensions de cromosomes metafàsics, per l'altra. A la Figura 13 estan esquematitzats els diferents passos a seguir.

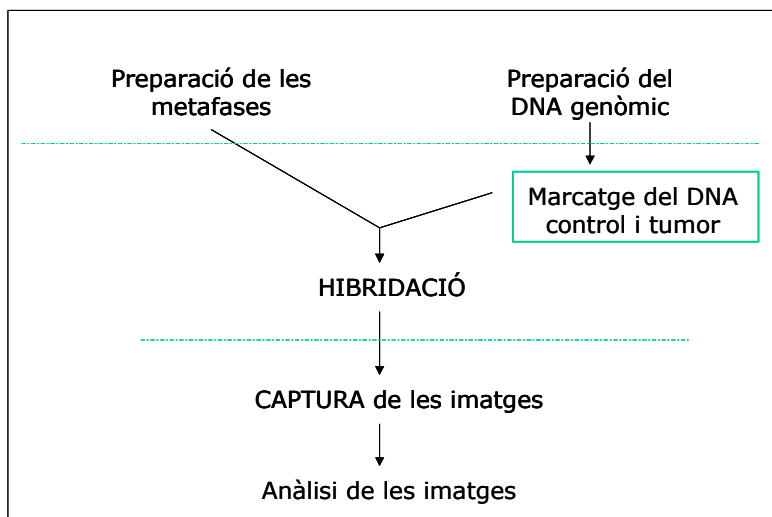


Figura 13. Esquematzació gràfica dels passos a seguir per fer una CGH.

Marcatge del DNA

Utilitzem el *Nick Translation Kit* de la casa Vysis, Inc. Per cada cas es marquen 800 ng de DNA genòmic del tumor amb el fluorocrom vermell, i DNA genòmic de sang perifèrica amb el fluorocrom verd, que fem servir de control.

En un volum final de 50 μ l es fa la següent mescla per a la reacció de marcatge:

Taula 11. Marcatge de les sondes de la CGH.

mescla:	H ₂ O lliure de nucleases	20,5 – X μ l
	DNA (800 ng)	X μ l
	dTTP 0.1 mM	5 μ l
	dNTP a 0.1 mM cada un (dATP, dCTP, dGTP)	10 μ l
	Buffer 10 \times ¹	5 μ l
	<i>SpectrumGreen</i> dUTP o bé <i>Red</i> dUTP ²	2,5 μ l
	enzim <i>nick translation</i> ¹	7 μ l
condicions de la reacció: 1 hora i 15 minuts a 16 °C		
10 minuts a 70 °C		

¹ DNA polimerasa I, DNasa I en 50% de glicerol, 50 mM Tris-HCl pH 7,2, 10 mM MgSO₄, 0,1 mM DTT, 0,5 mg/ml, que se subministra amb el seu tampó concentrat 10 vegades (500 mM Tris-HCl, pH 7,2, 100 mM MgSO₄, 1mM DTT) de la casa Vysis, Inc. ² Fluorocroms de la casa Vysis, Inc.

En aquest punt es pot aturar el protocol i guardar els tubs a la nevera fins l'endemà o bé continuar amb els següents passos.

Precipitació de les sondes

- Ajuntem els 50 μ l de la mostra procedent del tumor i els 50 μ l de la mostra control
- Afegim 10 μ l (10 μ g) de DNA Cot-1 humà (1mg/ml) (Invitrogen, Life Technologies)
- Afegim 10 μ l d'acetat sòdic 3 M (0,1 volums)
- Afegim 300 μ l d'etanol al 100% fred (2,5 volums)
- Ho deixem precipitant fins al dia següent a la nevera a 4 °C
- L'endemà ho centrifuguem 12 minuts a 12000 rpm a temperatura ambient
- Decanem el sobrenedant i deixem assecar el *pellet* a 37 °C
- Ho resuspenem amb 12 μ l de solució A (50% formamida desionitzada i 50% MASTER MIX: 20% dextrà sulfat en 2 \times SSC)

Preparació de les metafases

- Descongelar les extensions metafàsiques normals de limfòcits de sang perifèrica (Vysis, Inc.), passant-les del congelador a la nevera, i després a temperatura ambient.
- Cada portaobjectes conté dues extensions, a dret i esquerra, per això fem una ratlla separadora al mig amb un llapis de diamant per distingir les dues àrees.
- Incubar el porta 5 minuts en una solució desnaturalitzant a 70 °C dins d'una cubeta *Coplin*. SOLUCIÓ DESNATURALITZANT:
 - 49 ml de formamida
 - 7 ml de 20 \times SSC pH
 - 14 ml aigua mQ
- Deshidratar els portes passant-los 2 minuts en les següents solucions i en el següent ordre (també s'utilitzen *Coplin*):
 - 1er) etanol al 70% fred
 - 2on) etanol al 90% fred
 - 3er) etanol al 100% fred
- Deixar assecar a temperatura ambient davant d'un ventilador.

Hibridació

- Desnaturalitzar la sonda 5 minuts a 70 °C.
- Posar tot el volum (els 12 μ l) a cada una de les meitats del portaobjectes.
- Posar un cobreobjectes de 22 \times 22 mm al damunt de cada extensió de cromosomes i segellar tot el seu perímetre amb pega per evitar l'evaporació de la sonda.

- Incubar dins d'una cambra fosca i humida a 37 °C durant 36 o 72 hores (de 2 a 3 nits).

Rentats

- Retirar els cobreobjectes i la pega amb cura.
- Rentar durant dos minuts a 70 °C amb una solució de SSC 0,4× i Tween-20 al 0,3%.
- Rentar durant 2 minuts a temperatura ambient amb una solució de SSC 2× i Tween-20 al 0,1%. (Aquests dos rentats es fan dins d'un *Coplin* i es van agitant els portaobjectes amb l'ajuda d'unes pinces).
- Deixar assecar davant d'un ventilador.
- Un cop sec posar una gota de DAPI II/Antifade (4.6-diamidino-2-fenilindol dihidroclorid, a 0,1 ug/ml, de Vysis, Inc.) a cada una de les àrees on hi ha les extensions de cromosomes i posar un cobreobjectes damunt. Aquesta última tinció tenyeix les bandes dels cromosomes i ens servirà per tal de poder-los identificar.

Un cop acabat ja es pot mirar al microscopi de fluorescència amb els filtres corresponents per veure els tres fluorocroms: *Sepctrum Red*, *Spectrum Green* i DAPI.

Captura i anàlisi de les imatges

MICROSCOPI: Olympus Optical BX51, Hamburg, Germany.

PROGRAMARI: Cytovision Ultraworkstation (Applied Imaging, Sunderland, UK).

Es busquen aquelles metafases on tots els cromosomes estiguin ben estesos i amb el mínim nombre possible de solapaments, i que alhora presentin una bona tinció amb els tres colors (verd, vermell i DAPI). Es capturen les imatges dels tres canals i després s'analitzen mitjançant el programa Cytovision. Resumint cal seguir els següents passos:

- fer el cariotip de cada metafase amb la imatge marcada amb DAPI.
- corregir els eixos dels cromosomes.
- analitzar com a mínim 10 metafases per cada cas i obtenir la mitjana resultant dels canvis entre els dos canals en forma de perfil de les alteracions damunt de cada cromosoma (fent la raó del canal vermell/verd).
- avaluació dels perfils de cada cas.

c) Anàlisi de les dades

S'introdueixen en una fulla de càlcul d'Excel els resultats de guanys i pèrdues en valors numèrics de la següent manera:

+1	guany
-1	pèrdua
0	no canvi

Cada braç de cromosoma és introduït a la taula independentment, a excepció dels cromosomes 13, 14, 15, i 19 on només considerem el braç llarg. Així els camps d'estudi són els següents:

ch01p ch01q ch02p ch02q ch03p ch03q ch04p ch04q ch05p ch05q
ch06p ch06q ch07p ch07q ch08p ch08q ch09p ch09q ch10p ch10q
ch11p ch11q ch12p ch12q ch13 ch14 ch15 ch16p ch16q ch17p
ch17q ch18p ch18q ch19 ch20p ch20q chxp chxq

Les amplificacions són considerades a part, perquè els mecanismes subjacents al guany d'un braç de cromosoma són diferents al d'una amplificació. Així estudiarem aquests dos successos de forma independent.

Totes les anàlisis estadístiques es van fer amb el paquet estadístic SPSS per a Windows, versió 9.0 (SPSS, Inc.).

Els clusters es van fer amb el programa estadístic *R* (<http://www.r-project.org/>). Per a més detalls veure l'apartat 4.9. de l'anàlisi estadístic.

4.6. PCR quantitativa (QPCR)

Amb l'objectiu d'identificar el gen candidat d'una de les zones amplificades recurrentment a la sèrie de 50 casos es va utilitzar la PCR quantitativa. Aquesta tècnica ens permet monitoritzar a temps real els cicles d'amplificació d'una seqüència mitjançant la utilització de fluorocroms a la reacció, de manera que es quantifica el número de còpies inicial de DNA.

Hi ha diferents mètodes que es basen en la utilització de fluorocroms inespecífics de la doble cadena de DNA com el SYBR® Green, o bé en la utilització de sondes específiques pel gen que estem estudiant. Nosaltres vam escollir aquesta segona aproximació. I concretament vam utilitzar el sistema anomenat Taqman (Applied Biosystems), que s'aprofita de l'activitat 5'-exonucleasa de la DNA polimerasa. Breument, es tracta de dissenyar uns oligonucleòtids que amplifiquen un

fragment del gen d'interès, alhora que s'utilitza una sonda que hibrida amb una part de la zona amplificada. Aquesta sonda està associada a la part 5' a una molècula fluorescent que emet en una determinada longitud d'ona. I a la part 3' hi ha una altra molècula (*quencher*, en anglès *apagar*) que inhibeix l'emissió de la fluorescència. Així, mentre la sonda està intacta no s'emet fluorescència (Figura 14). A cada cicle de la PCR, la sonda es pot unir a la cadena sencilla quan se separa el dsDNA, i en el pas de la síntesi de la cadena nova, s'hidrolitza la sonda per l'activitat exonucleasa i s'allibera el fluorocrom donant un pic de fluorescència. Així a cada cicle, com que s'incrementa la quantitat de cadena motlle nova de forma exponencial, també s'incrementa la quantitat de fluorescència emesa en el pas de l'allargament o síntesi. Aquesta senyal és directament proporcional al nombre de molècules present al final de cada cicle d'amplificació.

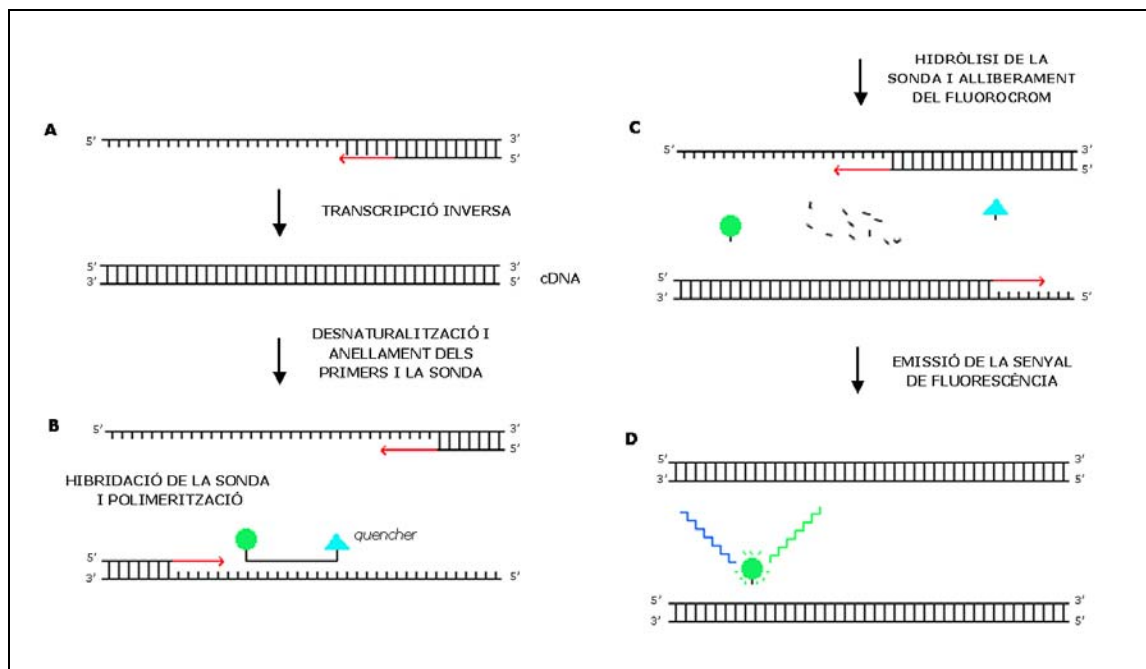


Figura 14. Esquema del mètode de PCR quantitativa *Taqman* (vegeu el text).

La manera de quantificar el resultat és a partir de la definició del valor **Ct** (de l'anglès, *threshold cycle*): cicle a partir del qual la senyal de fluorescència comença a ser exponencial. Així, com més quantitat de DNA motlle hi hagi a l'inici de la reacció, menys cicles tardarà a arribar al seu Ct. I aquest valor es pot traduir a un resultat quantitatiu si construïm una recta patró amb una mostra de concentració coneguda.

Per poder calcular les vegades que un gen està guanyat, necessitem un control intern d'un gen no amplificat. Nosaltres vam estudiar el gen de la ciclina D1 (CCND1) a nivell del DNA, i com a control de no amplificació el gen de la $\beta 2$ -

microglobulina. Alhora, calculem els resultats en funció d'un cas control que no presenti amplificació. Per tant, la diferència del número de cicle en el qual comença l'amplificació exponencial (Ct) entre dues mostres es pot expressar en valors relatius del DNA inicial. Els càlculs que fem són els següents:

$$\mathbf{Ct\ CICLINA\ D1 - Ct\ B2M = \delta Ct}$$

$$\mathbf{\delta\delta Ct = \delta Ct\ problema - \delta Ct\ control}$$

$$\mathbf{contingut\ de\ DNA = 2^{\delta\delta Ct}}$$

Per tant, quan no hi ha guany ni pèrdua el $\delta\delta Ct$ és zero, i el contingut de DNA normal serà $2^0=1$. Alhora de determinar els llimars de canvi vam utilitzar 10 casos que no presentaven cap canvi al cromosoma 11 (on es localitza el gen CCND1), ni al 15 (on es localitza el gen de la B2M). D'aquesta manera vam establir que a partir d'un contingut de DNA d'1,6 podíem parlar de guany, i a partir de 0,4 de pèrdua (vegeu l'apartat de Resultats).

Així i tot, prèviament, per poder quantificar el DNA inicial necessitem assegurar-nos que la cinètica de la reacció d'amplificació segueix una resposta lineal; com a mínim dins un rang conegut de concentracions, amb el qual després haurem de treballar. Per això abans d'estudiar les mostres vam construir una recta patró per diferents concentracions conegudes amb els dos gens, el d'estudi i el de control (Figura 15). La pendent de la recta en ambdós gens difereix només en 0,1, i per tant es poden comparar els resultats obtinguts entre aquests dos gens per fer els càlculs pertinents.

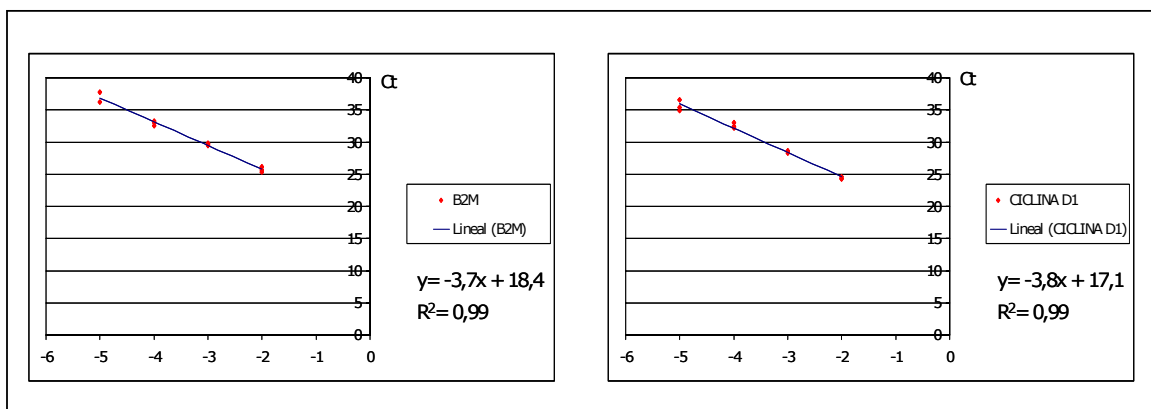


Figura 15. Recta patró dels gens $\beta 2$ -microglobulina i ciclina D1. Es representa la concentració en l'eix de les abscisses i el valor de Ct en l'eix de les ordenades. Es van utilitzar 4 dilucions per triplicat: 20 ng/ μ l (-1), 2ng/ μ l (-2), 0,2 ng/ μ l (-3), 0,02 2ng/ μ l (-4), i 0,002 ng/ μ l (-5).


Vam utilitzar el termociclador d'ABI Prism 7700 (Applied Biosystems). Les PCRs es feien en un volum final de 25 μ l i per triplicat. Les condicions de les dues PCRs eren les següents:

Taula 12. Condicions de la PCR quantitativa pel gen ciclina D1.

oligonucleòtids: *forward* 5'- CCATGAACTACCTGGACCGCT -3'
reverse 5'- CAGCTGCAGGCGGCTCT -3'

sonda: 5'- CCTGTCGCTGGAGCCCGTGAAA -3'

mescla:	H ₂ O	8 μ l
	<i>Master Mix</i> 2 \times ¹	12,5 μ l
	<i>forward</i> a 10 μ M ²	0,75 μ l
	<i>reverse</i> a 10 μ M ²	0,75 μ l
	sonda ³	1 μ l
	DNA (40 ng)	2 μ l

condicions de la PCR: 10 min a 95 °C
 15 sec a 95 °C
 1 min a 60 °C  40 cicles


¹TaqMan® Universal PCR Master Mix, que conté l'enzim AmpliTaq Gold® DNA polimerasa, AmpErase®UNG, dNTPs amb dUTP. ²Oligonucleòtid de Invitrogen, Life Technologies. ³Sonda TaqMan® marcada amb el fluorocrom FAM de la casa Applied Biosystems.

Taula 13. Condicions de la PCR quantitativa pel gen β 2-microglobulina.

oligonucleòtids: *forward* 5'- GGAATTGATTTGGGAGAGCATC -3'
reverse 5'- CAGGTCCTGGCTCTACAATTTACTAA -3'

sonda: 5'- AGTGTGACTGGGCAGATCATCCACCTTC -3'

mescla:	H ₂ O	8 μ l
	<i>Master Mix</i> 2 \times ¹	12,5 μ l
	<i>forward</i> a 10 μ M ²	0,75 μ l
	<i>reverse</i> a 10 μ M ²	0,75 μ l
	sonda a 5 pmol/ μ l ³	1 μ l
	DNA (40 ng)	2 μ l

condicions de la PCR: 10 min a 95 °C
 15 sec a 95 °C
 1 min a 60 °C  40 cicles

¹TaqMan® Universal PCR Master Mix, que conté l'enzim AmpliTaq Gold® DNA polimerasa, AmpErase®UNG, dNTPs amb dUTP (Applied biosystems). ²Oligonucleòtid de Invitrogen, Life Technologies. ³Sonda TaqMan® marcada amb el fluorocrom VIC de la casa Applied Biosystems.

4.7. RNA Arbitrarily Primed PCR (RAP-PCR)

a) Introducció a la tècnica

La RAP-PCR és una tècnica de *fingerprint* que permet detectar canvis d'expressió gènica que va ser introduïda per Welsh l'any 1992 (Welsh et al., 1992). Es basa en el mètode descrit anteriorment també per Welsh i McClelland l'any 1990 de l'AP-PCR (*Arbitrarily Primed PCR*), usat per amplificar seqüències de DNA (Welsh i McClelland, 1990). La diferència en aquest cas és que partim de l'RNA total d'una mostra.

La RAP-PCR té molts avantatges enfront d'altres mètodes de detecció d'expressió diferencial de gens, ja que és una tècnica fàcil, calen pocs ng d'RNA, i ens dona informació dels patrons generals d'expressió gènica entre diferents tipus de mostres. D'aquesta manera permet l'estudi de molts casos simultàniament d'una manera senzilla i ràpida.

Breument es basa en la utilització d'un oligonucleòtid arbitrari per a la síntesi de la primera cadena (el cDNA) a partir de l'RNA total o missatge. En aquest pas de la transcripció inversa estem fent una selecció d'aquelles seqüències d'RNA que tinguin una millor homologia amb l'oligonucleòtid escollit. A continuació se sintetitza la segona cadena mitjançant l'amplificació del cDNA amb el mateix oligonucleòtid. En aquest segon pas se segueix la mateixa estratègia que l'AP-PCR, amb cinc cicles d'amplificació inicials de baixa astringència, durant els quals l'oligonucleòtid es pot unir a múltiples llocs del cDNA amb més o menys homologia. En aquests primers cicles se sintetitzen diferents productes de PCR que incorporen la mateixa seqüència de l'oligonucleòtid a l'altre extrem. A continuació s'augmenta la temperatura durant 35 cicles més d'alta astringència i s'acaben d'amplificar tots els fragments. El producte resultant es pot resoldre en un gel d'acrilamida obtenint un patró de bandes discretes (*fingerprint*). Tots els productes amplificats tindran als dos extrems la mateixa seqüència (complementàries) de l'oligonucleòtid utilitzat. I cada oligonucleòtid té el seu patró característic d'unes 50 a 100 bandes, i s'amplifiquen fragments de 50 a 2000 parells de bases aproximadament (Figura 16).

Avantatges

L'aspecte més interessant és que la RAP-PCR és semiquantitativa, és a dir, ens informa de la quantitat relativa d'una seqüència. Això s'aconsegueix gràcies a la competició que s'estableix entre tots els possibles llocs d'unió de l'oligonucleòtid, permetent que una diferència d'intensitat en el gel sigui el reflex de l'abundància relativa de les seqüències de cDNA inicials. D'aquesta manera podem comparar dues mostres i veure quines bandes estan sobreexpressades o infraexpressades

en un teixit tumoral respecte la seva parella de mucosa normal. Per tant, d'una manera fàcil podem identificar una gran quantitat de marcadors tumorals.

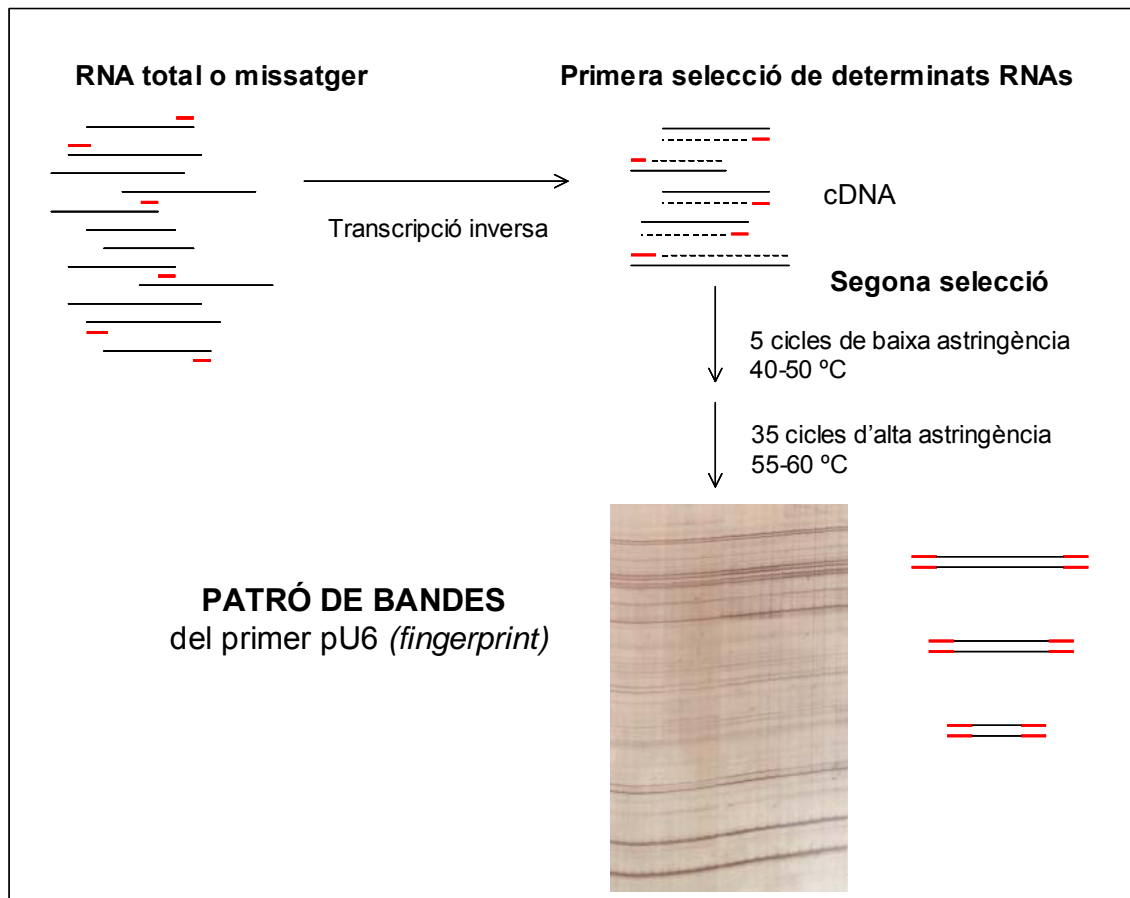


Figura 16. Esquema del procés d'amplificació de l'RNA mitjançant *RNA Arbitrarily Primed PCR* (RAP-PCR).

Limitacions

L'únic inconvenient que hi ha és que estem treballant amb seqüències anònimes (bandes), i per tant, per tal de verificar quin és el gen que està formant una banda discreta al gel, s'haurà de clonar i seqüenciar. Actualment, això no és gaire problema gràcies a la senzillesa d'aquestes tècniques. Amb l'objectiu de confirmar que la seqüència clonada és la correcta, es pot fer un *Southern Blot* del *fingerprint* utilitzant com a sonda el clon marcat radioactivament.

El nostre abordatge

En aquesta tesi l'estratègia ha estat una mica diferent. El pas de la RAP-PCR s'ha fet amb l'objectiu d'amplificar l'RNA de les mostres d'una manera semi-quantitativa per després poder marcar i hibridar aquest producte sobre microarrays de cDNA. El nostre interès era obtenir patrons de bandes prou rics en

número de seqüències per tenir una màxima representativitat del transcriptoma de la cèl·lula. Tot i així, hem de tenir en compte que estem fent una selecció d'una part de tots els gens que s'estan transcrivint a la cèl·lula en un moment determinat. Amb l'avantatge que alhora estem enriquint la mostra amb RNAs que es trobin a baixes concentracions, però que tinguin una bona homologia amb l'oligonucleòtid escollit. Així podrem detectar canvis d'expressió subtils de gens de baixa expressió, que d'altra manera no podríem estudiar (Taula 14).

Taula 14. Ús de mostres de complexitat reduïda enfront cDNA total en estudis d'expressió gènica per arrays.

Avantatges	Limitacions
<ul style="list-style-type: none">• Sense biaix entre mRNAs rars i comuns• Solapament entre oligonucleòtids del 20% al 40%• Es pot estudiar fins un 40%-60% dels gens del transcriptoma• Poc requeriment de material: ~1µg RNA total permet fer més de 20 experiments• Elevada sensibilitat i especificitat	<ul style="list-style-type: none">• Sense resposta lineal• Per tenir la representació de més del 90% del transcriptoma es requereixen entre 5 i 6 RAP-PCRs per mostra

Disseny de l'experiment

De cada cas es feien dues reaccions de transcripció inversa (RT, de l'anglès *Reverse transcription*) en el pas de la síntesi de la primera cadena de cDNA. A continuació de cada un dels dos tubs es feien tres rèpliques en el pas de la síntesi de la segona cadena i amplificació (Figura 17).

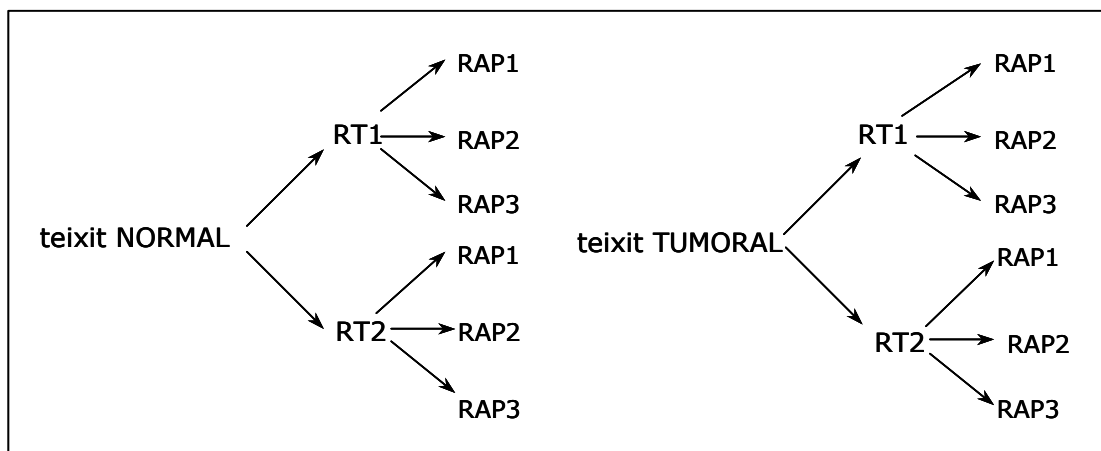


Figura 17. Esquema del disseny experimental dut a terme amb les 50 mostres de l'estudi. RAP: RAP-PCR.

A continuació es purificaven els productes de la RAP-PCR i es resolien en gels d'acrilamida al 6% per comprovar que el patró de bandes era correcte. Es barrejaven tots els productes de les rèpliques del teixit normal per una banda i el tumoral per l'altra, i es procedia al marcatge per a fer la hibridació sobre microarrays. Més endavant en aquest capítol s'expliquen els protocols detalladament.

b) Protocol

Transcripció inversa

La reacció de síntesi del cDNA es va fer a partir de les dilucions d'RNA a 5 ng/μl i en un volum final de 20 μl. Les condicions de la reacció i l'oligonucleòtid utilitzat foren les següents:

Taula 15. Condicions de la transcripció inversa amb l'oligonucleòtid pU6.

oligonucleòtid pU6: 5'- GCTTCTGACTTATTCTTGCTCTGAG -3'	
mescla:	
H ₂ O	0,4 μl
Buffer 5× ¹	4 μl
dNTP 5 mM de cada un ²	2 μl
DTT ¹	2 μl
pU6 a 100 μM ³	0,1 μl
Inhibidor RNases ⁴	0,5 μl
M-MLV <i>Reverse Transcriptase</i> ¹	1 μl
RNA a 5 ng/μl	10 μl
condicions de la PCR: 1 hora a 37 °C	
5 min a 95 °C	

¹Transcriptasa inversa de Moloney Murine Leukemia Virus (M-MLV RT) a 200 U/μl, que se subministra amb el seu tampó concentrat 5 vegades (250 mM Tris-HCl pH 8,3, 375 mM KCl, 15 mM MgCl₂) i 0,1 M DTT (Invitrogen, Life Technologies). ² dNTP: desoxiribonucleòtids trifosfat (Amersham Biosciences Corp). ³Oligonucleòtid de Invitrogen, Life Technologies ⁴RNA Guard™ (inhibidor porcí d'RNases) a 30000 unitats/ml (Amersham Biosciences Corp).

Amplificació mitjançant RAP-PCR



La reacció d'amplificació es va fer a partir del cDNA en un volum final de 50 μl. Les condicions de la reacció i l'oligonucleòtid utilitzat foren les que apareixen a la Taula 16.

Purificació dels productes de la RAP-PCR

Vam utilitzar el kit *CONCERT™ Rapid PCR Purification System* (GibcoBRL, Life Technologies) per tal de purificar el DNA i així eliminar totes les restes de la mescla de la PCR. Es van diluir en aigua per tal de procedir amb el marcatge

posterior. Abans però, calia comprovar que la PCR havia funcionat bé mitjançant l'electroforesi.

Taula 16. Condicions de la PCR amb l'oligonucleòtid pU6.

oligonucleòtid pU6: 5'- GCTTCTGACTTATTCTTGCTCTGAG -3'	
mescla:	
H ₂ O	33 µl
Buffer 10× ¹	5 µl
dNTP 1 mM de cada un ²	6 µl
MgCl ₂ a 100 mM	0,5 µl
pU6 a 100 µM ³	1 µl
Taq DNA polimerasa ¹	0,5 µl
producte de la RT	4 µl
condicions de la PCR:	
1 min a 95 °C	
1 min a 94 °C	 5 cicles
45 sec a 40 °C	
1 min 15 sec a 72 °C	
1 min a 94 °C	 35 cicles
45 sec a 60 °C	
1 min 15 sec a 72 °C	
5 min a 72 °C	

¹DNA polimerasa *Thermus aquaticus* (Taq) que se subministra amb el seu tampó concentrat 10 vegades (100 mM Tris-HCl pH 8,0, 500 mM MgCl₂, 0,1% gelatina) de la casa Roche. ² dNTP: desoxiribonucleòtids trifosfat (Amersham Biosciences Corp). ³Oligonucleòtid de Invitrogen, Life Technologies.

Electroforesi

Per tal de veure que els productes de la RAP-PCR s'havien amplificat bé, i que els patrons eren els correctes, es resolien les mostres en gels de seqüenciació desnaturalitzants amb el 6% d'acrilamida i 8 M d'urea. La composició de la mescla per fer el gel era la següent:

- 51 ml de TBE 1,2× i urea a 9,4 M
- 9 ml de solució al 40% acrilamida:bisacrilamida (29:1)
- més 600 µl de persulfat amònic al 10% i 25 µl de TEMED (N, N, N', N',-tetra-metiletilen-diamina), que activen la polimerització del gel

Els gels feien 50 cm de llarg, 30 cm d'ample i 0,4 mm de gruix. Abans de muntar els vidres es tractaven de la següent manera:

- al vidre gran s'hi escampaven 800 µl de SIGMACOTE (Sigma Chemical Co.), una solució repel·lent de silicona que evita que s'hi enganxi el gel.
- el vidre petit es tractava amb una solució que contenia GORP (γ-metacriloxipropil-trimetoxisilan) de la casa Sigma Chemical Co. Aquest

producte fa que el gel s'enganxi al vidre i així després es pot tenyir sense perill que es trenqui degut a que aquests gels són molt prims. La composició de la solució del GORP era la següent:

75 µl d'àcid acètic al 10%
2,5 ml d'etanol absolut
11 µl de GORP

S'escampava tot el volum damunt del vidre i es deixava actuar durant 5 minuts. A continuació es netejava el vidre per treure l'excés i acabar-los d'assecar amb etanol al 70%, i per últim amb etanol absolut.

Finalment es munten els vidres amb els separadors de 4 mm de gruix i les pintes, i se subjecten amb pinces. Cal deixar que el gel polimeritzi un mínim de dues hores.

Les mostres es diluïen (1:4) amb el tampó desnaturalitzant de càrrega DLB (93% formamida desionitzada, 0.01 M EDTA, 0.01 M NaOH, 0.01% blau de bromofenol, 0.01% blau de xilen cianol). Es desnaturalitzaven a 95 °C durant 3 minuts, i a continuació es col·locaven en gel fins al moment de carregar-les per evitar que renaturalitzessin.

Un cop el gel havia polimeritzat es muntava a al cubeta d'electroforesi que s'omplia amb TBE 1× (90 mM Tris-Borat, 2 mM EDTA) filtrat, i es feien precórrer 30 minuts a 80W per equilibrar el tampó de la cubeta i el gel. A continuació es carregaven 3 µl de cada mostra i es corrien unes 3 hores i mitja a 55W, fins que l'últim blau (el xilen cianol) es perdia per l'extrem del gel.

Un cop finalitzada l'electroforesi se separaven els vidres amb cura, sabent que el gel queda enganxat al vidre petit, i es procedia a la tinció de DNA mitjançant el nitrat de plata.

Tinció de plata

Basat en el protocol descrit per Bassam (Bassam et al., 1991). Tots els passos es fan dins d'una cubeta on s'hi van afegint 1 litre de cada un dels reactius. Aquests es poden arribar a reciclar fins a 5 vegades.

- Fixació: 10 minuts en etanol al 10%.
- Oxidació: 3 minuts en una solució d'àcid nítric a l'1%.
- Rentats: 2 rentats amb aigua destil·lada.
- Tinció: 20 minuts en una solució amb nitrat de plata (AgNO_3) a una concentració de 2 g/l.
- Rentats: 2 rentats amb aigua destil·lada.

- Revelat: es fa amb una solució de carbonat de sodi que es prepara al moment en un litre d'aigua més

29.6 g de NaCO ₃
450 µl de formaldehid al 37%

Primer es posen dins la cubeta uns 250 ml de la solució i s'espera a que s'enfosqueixi degut a un precipitat marró. Es llença aquest volum i s'acaba de posar la resta del líquid. Mentre anem barrejant es veu com van apareixent les bandes de mica en mica.

- Aturar la reacció: un cop les bandes són prou nítides, i abans que s'enfosqueixi massa el fons del gel, s'atura la reacció amb una solució d'àcid acètic al 10% durant uns 5 minuts.
- Rentat: es fa un rentat amb aigua per treure l'àcid acètic. I es deixa uns 10 minuts submergit en aigua per acabar d'eliminar-lo.
- Assecar i escannejar: es deixa assecar el gel a temperatura ambient i així es pot escannejar sense que es faci malbé.
- Conservació: es torna a posar el gel en una cubeta amb aigua un mínim d'una hora perquè s'estovi. Es rasca amb l'ajuda d'un regle, quedant així el gel tot cargolat, i dins d'una cubeta amb aigua s'estén altre cop amb l'ajuda dels dits. S'agafa (es pesca) damunt d'un paper de filtre intentant que quedi llis i s'asseca al buit a 85°C. D'aquesta manera es poden guardar permanentment.

4.8. Microarrays de cDNA

a) Introducció a la tècnica

La tecnologia dels arrays és una tècnica que permet l'estudi massiu de molts gens a nivell del seu RNA. Per microarray s'entén una matriu bidimensional de material genètic fixat en un suport sòlid. Breument i en general, es tracta de la immobilització de milers de seqüències sobre les quals s'hibrida una mostra complexa de cDNA marcat mitjançant diferents mètodes possibles.

M'agradaria destacar el perquè de l'anglicisme *array* (en català *graella* o *matriu*). Es tracta de la disposició dels gens diana seguint una graella ordenada de punts en dues dimensions (Figura 18). D'aquesta manera s'aconsegueix miniaturitzar el format clàssic dels *Dot Blots* i alhora es capgira el seu disseny, és a dir, es marca la mostra més complexa i s'immobilitzen els gens o seqüències d'interès. El principal avantatge d'utilitzar el mínim espai per posar el màxim de punts el trobem en el fet de poder obtenir informació de milers de gens alhora en un sol experiment o hibridació. La limitació en aquest cas serà la tècnica: robots més precisos i puntes més fines permeten fer punts més petits i junts, ocupant menys

espai. Tot això amb l'objectiu d'obtenir el màxim d'informació amb el mínim espai i despesa possibles.

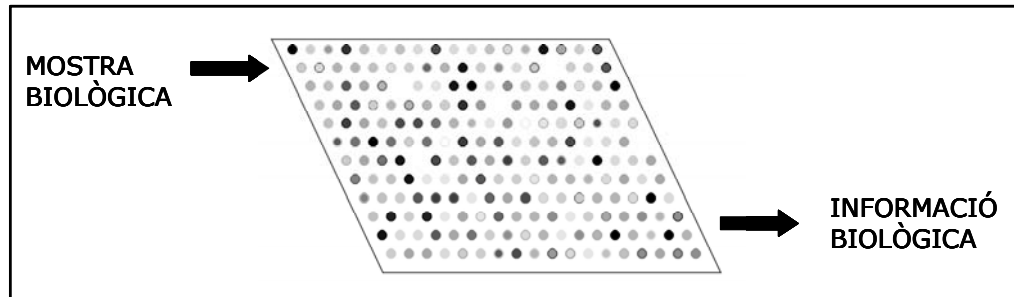


Figura 18. Concepte d'un experiment en format de microarray. Una mostra és marcada amb una molècula fluorescent o radioactiva i es fa reaccionar amb una sèrie de molècules ordenades en un suport sòlid. La localització i mesura de les molècules unides a l'array proporcionen informació de la identitat i la quantitat del producte gènic present a la mostra d'estudi. El tipus d'informació biològica obtinguda dependrà de les mostres fixades en el suport sòlid i de la mostra marcada en la fase líquida.

Classificació de les microarrays

El format dels microarrays des del seu inici ha suposat una plataforma molt flexible on l'avenç de la tècnica ha fet que n'apareguessin una gran varietat de tipus i aplicacions. És per això que es poden classificar en base a diversos paràmetres:

Taula 17. Possibles característiques de diferents formats de microarrays.

CRITERI	CARACTERÍSTIQUES	NOM
suport	<ul style="list-style-type: none"> no porós porós 	<ul style="list-style-type: none"> suport rígid (vidre o plàstic) membranes
material immobilitzat	<ul style="list-style-type: none"> oligonucleòtids cDNA 	<ul style="list-style-type: none"> xips de gens arrays de cDNA
densitat de punts	<ul style="list-style-type: none"> alta baixa 	<ul style="list-style-type: none"> microarrays macroarrays
impressió	<ul style="list-style-type: none"> generació dels punts 	<ul style="list-style-type: none"> "in situ" per dipòsit

Tipus de suport

Les hibridacions clàssiques desenvolupades als anys 70 utilitzaven membranes flexibles de nitrocel·lulosa o niló, radioactivitat, i autoradiografies. En canvi, els arrays, a part de les tècniques basades amb la radioactivitat també utilitzen suports sòlids com el vidre, i un marcatge i sistema de detecció basats en la fluorescència. En comparació amb el format macroscòpic dels experiments amb filtres, la miniaturització dels anomenats *bioxips* (per analogia als xips dels ordinadors) representa una revolució en els anàlisis biològics.

L'avantatge d'utilitzar superfícies sòlides rau en el fet que no són poroses, i d'aquesta manera permeten la deposició de petites quantitats de material en una posició molt precisa i definida. A diferència dels substrats com la nitrocel·lulosa, on no es pot evitar la difusió del material i per tant no són apropiats pels formats microarray.

Els substrats no porosos també són molt adients perquè permeten l'ús de volums molt petits de mostra. Les reaccions en volums petits redueixen el consum de reactius i el cost, alhora que incrementen la concentració d'àcids nucleics reactius de l'experiment unes 100.000 vegades en comparació als mètodes tradicionals. Passem de tenir concentracions de l'ordre de 0,4-4 pM a 0,1-1 µM. L'elevada concentració accelera la cinètica de la reacció, reduint el temps necessari per obtenir una senyal de fluorescència forta. A la Taula 18 es resumeixen les característiques dels dos formats, els filtres i els microarrays:

Taula 18. Comparació dels dos tipus de suport dels arrays: el filtre i el vidre.

CRITERI	FILTRE	VIDRE
superfície	porosa	no porosa
substrat	no uniforme	uniforme
estructura	flexible	rígida
format	semiparal·lel	paral·lel
format en miniatura possible	no	sí
compatible amb fluorescència	no	sí
concentració de la mostra	baixa	alta
cinètica de la hibridació	lenta	ràpida
volum del reactius	gran	petit

El seu format flexible permet alhora un gran ventall d'aplicacions, en funció principalment del tipus de seqüències que fixem damunt del suport sòlid:

Taula 19. Aplicacions de les arrays.

TIPUS DE SEQÜÈNCIA	APLICACIÓ
cDNA	- canvis d'expressió gènica
oligos	- canvis d'expressió gènica - mutacions puntuals - SNPs

Avantatges i inconvenients de les microarrays

Sempre que apareix una nova tècnica, s'acull amb un optimisme a vegades desmesurat. I aquest va ser el cas de les microarrays. La possibilitat d'estudiar

tots els gens d'una mostra en un sol experiment és molt atractiva. I en un inici es publicaven articles sense que els resultats s'haguessin confirmat mitjançant cap altra tècnica. Posteriorment, es va veure que molts dels gens escollits després del cribratge per microarrays eren falsos positius i calia extremar molt bé els paràmetres d'error de l'anàlisi estadístic. Per tant, si el que es vol és buscar gens cal confirmar els resultats mitjançant altres tècniques, com poden ser la PCR quantitativa, o el *Northern blot*.

Les fonts d'error poden ser moltes:

- hibridació creuada
- *splicings* alternatius
- contaminació dels clons
- errors de seqüenciació
- el fet que les condicions d'hibridació són iguals per a totes les seqüències de l'*array*, i per tant l'eficiència d'hibridació no serà igual per a tots els clons

Per això, passada la novetat, s'han fet molts esforços per tal d'estudiar la reproductibilitat de la tècnica. També hi ha hagut autors que han comparat els resultats obtinguts mitjançant suports diferents, com són els d'*Affymetrix* (basat en oligonucleòtids sintètics) i els de cDNA (Lee et al., 2003). En aquest cas, per exemple, busquen aquells gens que donin el mateix resultat en tots dos suports, i la concordança global es defineix amb el paràmetre anomenat "coeficient de la concordança de concordances".

Microarrays de cDNA

Els pioners en la utilització de la tècnica de les microarrays van ser el grup de Patrick O. Brown l'any 1995 (Schena et al., 1995). Des d'aquell moment la seva utilització va anar creixent de forma exponencial. En aquesta tesi s'han usat microarrays de cDNA, per això em centraré en aquest format.

Les microarrays de cDNA s'elaboren a partir de llibreries de gens immobilitzats en un portaobjectes de vidre. Cada punt *individual* de la graella conté el cDNA d'un gen únic o un EST (de l'anglès, *Expressed Sequenced Tag*). Aquestes seqüències són amplificades per PCR a partir d'una llibreria de clons. Són varies les companyies o consorcis que disposen d'un elevat nombre de clons que representen la gran majoria de gens que s'expressen en una cèl·lula. El gran repte d'aquest tipus d'arrays és el control de les contaminacions en tot el procés de clonatge, seqüenciació i amplificació.

Actualment, molts laboratoris que es van afanyar a fabricar microarrays de cDNA amb el màxim de gens possibles, estan reseqüenciant els clons dipositats als

arrays per confirmar les dades que obtinguin a partir d'aquests. És una feina lenta, però necessària per obtenir uns resultats prou fiables i robustos.

Per altra banda, hi ha l'avantatge d'utilitzar clons de gens coneguts, així com EST que encara no han estat estudiades. D'aquesta manera es poden fer troballes de nous gens implicats en el procés objecte del nostre estudi.

Sonda i diana

Abans de continuar, és necessari aclarir un concepte de nomenclatura. La confusió prové dels termes *sonda* i *diana* dins del context de la tècnica dels microarrays. En les hibridacions clàssiques, per exemple en un *Southern Blot*, anomenem sonda al "nostre" gen marcat i en la fase líquida o lliure, i la mostra fixada al suport sòlid és on hi ha la diana que és posada de manifest gràcies a la hibridació amb aquesta sonda. En el cas dels microarrays d'expressió, en canvi, la mostra marcada és la més complexa, el cDNA. És per això que a partir d'ara quan em refereixi a la **sonda** estaré parlant del cDNA; i en oposició a aquesta tots els gens immobilitzats en un suport sòlid seran les **dianes**. És una decisió arbitrària ja que en la literatura trobem l'ús de la paraula sonda (en anglès, *probe*) indistintament tant per referir-nos al cDNA com als gens que formen l'*array*.

El disseny experimental d'aquest tipus d'*array* es basa en la hibridació conjunta de dues mostres (control i estudi) marcades amb dos fluorocroms diferents. Així les diferències d'intensitat entre els dos fluorocroms representen diferències d'expressió del gen en qüestió. De forma general, se sintetitza el cDNA a partir de l'RNA i es marca amb el fluorocrom en una segona reacció (marcatge indirecte). També es pot marcar directament en el pas de l'RT (de l'anglès, *Reverse Transcription*). Existeixen diferents *kits* i fluorocroms al mercat.

Amplificació de l'RNA

El principal problema que cal tenir en compte és la quantitat d'RNA necessari per marcar i hibridar, que serà el pas més limitant per l'aplicació d'aquesta tècnica. I més tenint en compte que sovint les mostres d'estudi són biòpsies humanes petites i de les quals no es pot extreure una gran quantitat d'RNA. El marcatge típic per fer una hibridació requereix una quantitat de 2 a 4 µg d'RNA poli(A)⁺ o de 25 a 50 µg d'RNA total. És per això que també s'han elaborat estratègies per tal d'amplificar l'RNA mitjançant el promotor T7 de l'RNA polimerasa (Van Gelder et al., 1990). Es tracta de fer la síntesi del cDNA mitjançant un oligonucleòtid sintètic (dT) que conté el promotor T7. A continuació es degrada la cadena de l'RNA amb l'RNasa H, i se sintetitza la cadena complementària amb la DNA polimerasa d'*E. coli*, obtenint DNA de doble cadena (ds cDNA). L'RNA amplificat (aRNA) s'obté mitjançant la transcripció in vitro usant l'RNA polimerasa de T7. Els

productes amplificats fan entre 0,2 kb i 6 kb. Amb una sola ronda d'amplificació es pot arribar a produir un increment d'entre 200 i 700 vegades més de producte. S'han fet diferents estudis per tal de verificar la concordança entre les mostres amplificades i les no amplificades (Zhao et al., 2002). Arribant a la conclusió que el coeficient de correlació mitjà entre mostres amplificades i les que no és del 0,82. Paral·lelament, menys d'un 4% dels gens que presenten un canvi de dues vegades o més són diferents comparant els dos protocols.

En aquesta tesi, es va escollir un altre mètode per tal d'amplificar la quantitat d'RNA de partida, la RAP-PCR (de l'anglès, *RNA Arbitrarily Primed PCR*). Aquesta tècnica ens permet allargar la quantitat de mostra així com disminuir la seva complexitat. La informació que ens dóna no és la mateixa, però sí complementària.

Aquesta nova aproximació s'ha posat a punt i s'ha validat en el nostre laboratori, tal i com es descriu a l'article de l'annex II (Grau et al., 2005).

b) Esquema del mètode seguit

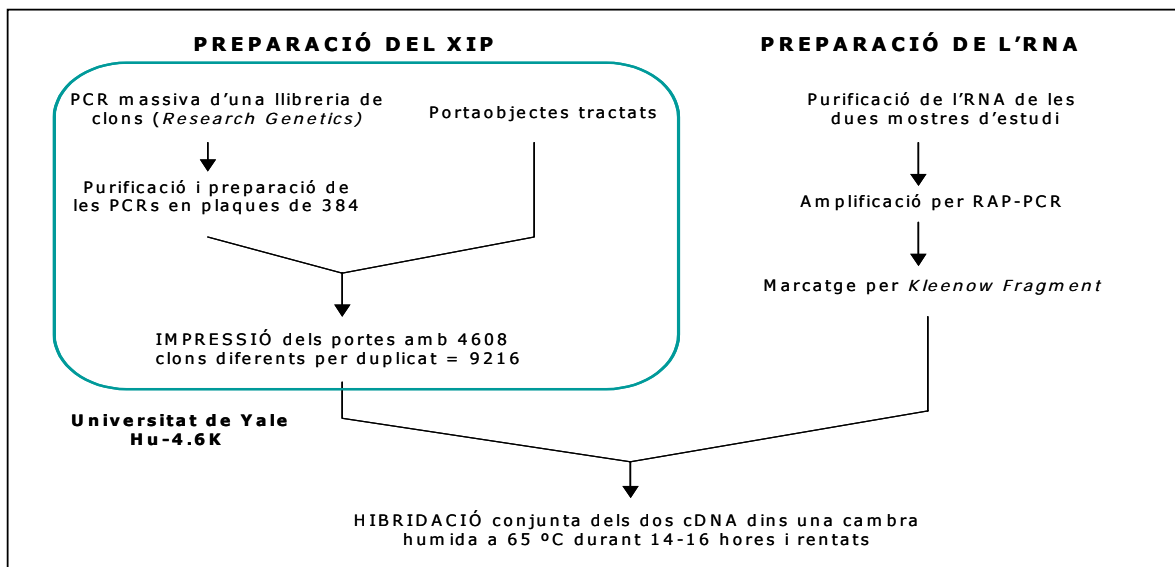


Figura 19. Esquema general del material necessari per dur a terme una hibridació sobre microarray.

Disseny de l'experiment

Es van hibridar els 50 casos de la següent manera:

- 1) el teixit normal marcat amb Cy3 i el teixit tumoral aparellat marcat amb Cy5.
- 2) una mescla de teixits normals marcat amb Cy3 (referència) i el tumor marcat amb Cy5.

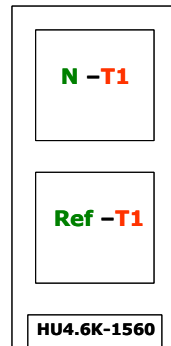


Figura 20a. Esquema de les hibridacions.

D'aquesta manera cada tumor s'hibridava dues vegades, amb la seva parella normal i amb una mostra de referència, que sempre era la mateixa per tots els casos. Així, amb aquesta doble estratègia, es tenia el tumor hibridat per duplicat, i alhora ens permetia estudiar dues vessants del grup de 50 casos:

- 1) les diferències d'expressió entre el teixit normal i el tumoral. Que ens permet veure quins són els gens importants per a la progressió del tumor.
- 2) les diferències d'expressió entre els 50 tumors. Així podem estudiar els perfils d'expressió del grup i fer noves classificacions.

c) Protocol

Marcatge de la sonda

El producte de la RAP-PCR purificada es marcava mitjançant el *kit* BioPrime[®] DNA Labeling System (Invitrogen, Life Technologies). S'usa un protocol modificat descrit per Pollack l'any 1999. Breument es tracta de marcar les seqüències mitjançant l'enzim *Klenow Fragment* que incorpora els nucleòtids marcats amb fluorescència. Encara que l'objectiu no és el d'amplificar, hi ha una síntesi neta de DNA d'entre 10 i 40 vegades de la mostra original. A la Taula 20 es detallen els reactius necessaris per al marcatge.

Al final teníem un volum de 55 µl de sonda marcada, juntament amb totes les restes de nucleòtids i fluorocroms no incorporats. El següent pas necessari era la purificació per tal d'eliminar totes aquestes restes.

Purificació dels productes marcats

Vam utilitzar el kit *CONCERT[™] Rapid PCR Purification System* (GibcoBRL, Life Technologies) per purificar la sonda i així eliminar totes les restes de la mescla del marcatge. Els productes resultants de les sondes es van diluir en 50 µl d'aigua per tal de procedir a la hibridació sobre les microarrays de cDNA.

Taula 20. Condicions del marcatge per *Klenow Fragment*.

oligonucleòtid: octàmers arbitraris	
mescla: producte de la RAP purificada	21 µl
<i>Random Primer Solution</i> 2,5× ¹	20 µl
5 minuts a 95 °C i mantenir en gel mentre s'afegeix la resta de reactius	
dNTP (1,2 mM dATP, dGTP, dTTP, 0,6 mM dCTP) ²	5 µl
enzim <i>Klenow Fragment</i> ¹	1 µl
dCTP-Cy3 o dCTP-Cy5 ³	3 µl
condicions del marcatge: 2 hores a 37 °C	
al final s'afegeixen 5 µl de l' <i>Stop Buffer</i> ¹ per aturar la reacció	

¹ *Klenow Fragment* (fragment gran de la DNA polimerasa I) a 40 U/µl en 50 mM fosfat de potassi pH 7, 100 mM KCl, 1 mM DTT, 50% glicerol, que se subministra amb la mescla de nucleòtids *Random Primers Solution* 2.5× (125 mM Tris-HCl pH 6.8, 12.5 mM MgCl₂, 25 mM 2-mercaptoetanol, oligodesoxi-ribonucleòtids a 750 µg/ml (octàmers arbitraris)) i l'*Stop Buffer* (0.5 M Na₂EDTA pH 8) (Invitrogen, Life Technologies). ² dNTP: desoxiribonucleòtids trifosfat (Amersham Biosciences Corp). ³ Fluorolink™ Cy3-dCTP o Cy5-dCTP (5-Amino-propargil-2'-deoxicitidina 5'-trifosfat conjugat al fluorocrom Cy3 (Cy3-AP3-dCTP) o Cy5 (Cy5-AP3-dCTP)) (Amersham Biosciences Corp).

Hibridació

Les sondes es van hibridar damunt d'uns xips impresos a la Universitat de Yale (EE.UU.) anomenats Human-4.6K (HU4.6K). Presenten 4608 cDNAs, d'una llibreria de Research Genetics. Els punts estan distribuïts en 4 × 4 subquadrants (*subarray*) fent un total de 16, on a cada un dels quals hi ha 24 × 24 punts. Cada cDNA està representat per duplicat mitjançant dos punts adjacents. La distància de centre a centre de cada punt és de 175 µm i tenen un diàmetre de 125 µm. Hi ha dos *arrays* idèntics a cada portaobjectes separats suficientment per permetre la hibridació de cada àrea amb un cobreobjectes independent (Figura 20b).

La llista amb els noms de tots els cDNA es pot obtenir a la pàgina web: http://keck.med.yale.edu/dna_arrays.htm#Human.

Sempre hibridàvem un total de 1200 ng de DNA tant del teixit normal o referència com del teixit tumoral. Per això calia quantificar la sonda mitjançant la lectura a l'espectrofotòmetre. Miràvem l'absorvència a 260 nm (on té el pic d'absorvència el dsDNA), a 550 nm (on té el pic d'absorvència el Cy3) i a 650 nm (on té el pic d'absorvència el Cy5).

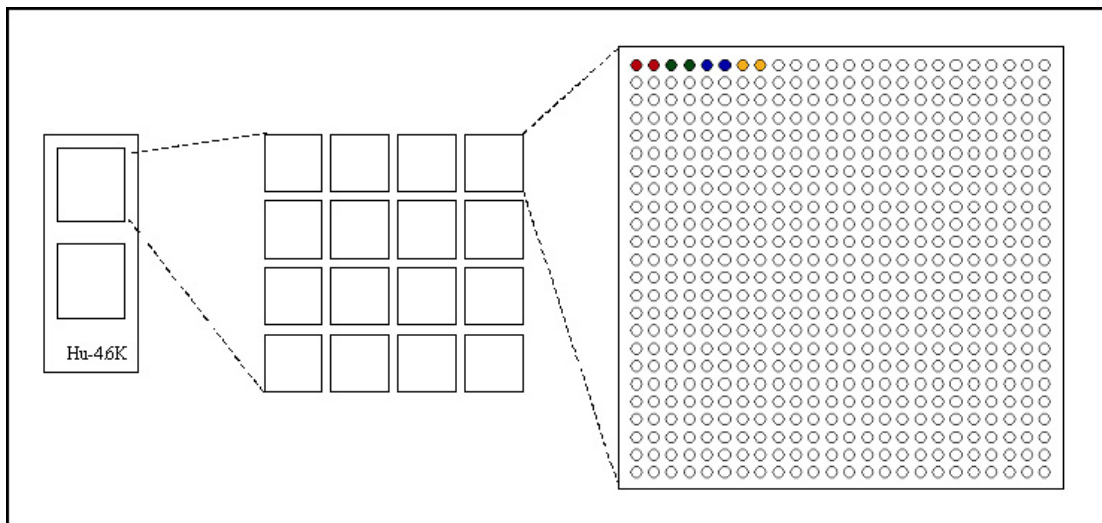


Figura 20b. Esquema del xip Human-4.6K (HU4.6K) fabricat a la Universitat de Yale (EE.UU.).

Amb aquests valors calculàvem:

- concentració del DNA

$$[\text{DNA}] = \text{Absorvència a } 260 \times 50 \times \text{factor de dilució (ng/}\mu\text{l)}$$

- marcatge incorporat

$$\left[\frac{\text{pmols Cy3}}{\mu\text{l}} \right] = \frac{\text{Abs. } 550}{0.15} \times \text{factor de dilució}$$

$$\left[\frac{\text{pmols Cy5}}{\mu\text{l}} \right] = \frac{\text{Abs. } 650}{0.25} \times \text{factor de dilució}$$

- eficiència del marcatge $\left[\frac{\text{pmols}}{\mu\text{g}} \right] = \frac{\text{pmols fluorocrom}/\mu\text{l}}{\text{concentració (ng}/\mu\text{l})} \times 1000$

Els valors de la quantitat del marcatge incorporat i de l'eficiència del marcatge ens servien de referència per veure que les sondes estaven ben marcades. Altrament, però, això ja es veia perquè les sondes quedaven d'un color rosa feble en el cas del Cy3 i blau cel, en el cas del Cy5. I a l'hora d'hibridar agafàvem

sempre la quantitat de 1200 ng de DNA total, perquè empíricament vam comprovar que ens donava una bona senyal de fluorescència.

La hibridació es feia seguint el protocol que van descriure Eisen M. B. i Brown P. O. l'any 1999 (Pollack et al., 1999). Així ajuntàvem el volum necessari per tenir els 1200 ng de la sonda de Cy3 amb la de Cy5 i els assecàvem a dins d'una centrífuga al buit. Aquest *pellet* sec el resuspeníem amb la següent mescla de tampó i solució de bloqueig:

Taula 21. Mescla de la sonda per hibridar

sonda marcada i seca: Cy3 + Cy5 (1200 ng de cada un)	
mescla:	H ₂ O 4,9 µl
	7× SSC, 0,06% SDS ¹ 7,3 µl
	solució de bloqueig (tRNA, polidA, Cot-1 DNA) ² 2,3 µl
condicions de la hibridació: de 14 a 16 hores a 65 °C dins d'una cambra individual de la casa CORNING	

¹ Aquesta solució a temperatura ambient precipita i és necessari escalfar-la a 65 °C just abans de ser utilitzada. ² Solució de bloqueig composta per 10 µg/µl de tRNA de llevat (Invitrogen, Life Technologies), 10 µg/µl polidA (oligonucleòtid 20mer de Invitrogen, Life Technologies), i 10 µg/µl Cot-1 DNA humà (Invitrogen, Life Technologies).

Així el *pellet* sec quedava resuspès amb la solució de bloqueig a una concentració final d'1,6 µg/µl de tRNA, polidA, Cot-1 DNA, i amb el tampó a 3× SSC i 0,3% SDS, en un volum final de 14,5 µl. La hibridació es muntava amb un cobreobjectes de 22×22 mm. S'utilitzaven cambres d'hibridació individuals de la casa *Corning* (Figura 21). Per mantenir la humitat a l'interior de la cambra hi ha dos pous on posàvem 10 µl de 2× SSC, es tancava quedant l'interior ben segellat, i se submergia al bany maria dins d'un forn d'hibridació model Mini Oven MKII (HYBAID). Es deixava de 14 a 16 hores a 65 °C.



Figura 21. Cambra d'hibridació de *Corning*.

Rentats

Dins de cubetes de vidre es posaven 200 ml de cada una de les solucions de rentat. Els portes se submergien en cada una de les solucions durant 2 minuts en agitació i a temperatura ambient. A cada pas es canviaven els *racks* per tal d'evitar arrossegar la solució d'una cubeta a l'altre. L'ordre era el següent:

- 2× SSC ($\text{Na}_3\text{C}_6\text{H}_5\text{O}_7 \times 5,5 \text{ H}_2\text{O}$ a 0,25 M, i NaCl 3M) i 0,1% SDS (dodecil sulfat sòdic): en el primer rentat s'ha d'anar amb cura a l'agitar fins que cauen els cobreobjectes.
- 1× SSC: agitant durant 2 minuts.
- 0,2× SSC: agitant durant 2 minuts.

A continuació i per assecar els portes se centrifugaven a 500 rpm durant 5 minuts a temperatura ambient. I si al final encara no estaven del tot secs es podien posar a l'estufa a 37 °C durant uns minuts.

Escanneig

Les imatges foren capturades amb l'escàner de la casa GSI Lumonics, *ScanArray 4000*. Els paràmetres inicials d'escanneig eren 80% de PMT (de l'anglès, *photomultiplier*) i 80% del làser (en anglès, *laser power*). Segons la qualitat de la hibridació s'ajustaven aquests dos paràmetres. Les imatges es guardaven en format *.TIF.

d) Anàlisi de les dades de microarrays

Les intensitats dels punts es van calcular amb el programa estadístic **R** (<http://www.r-project.org/>). Es basa en un llenguatge semblant a **S-Plus**, però amb la diferència de ser un *software* lliure, seguint el sistema operatiu del projecte GNU (<http://www.gnu.org/>) i les idees de la *Free Software Foundation* (<http://www.gnu.org/fsf/fsf.html>). Aquesta fundació té com a finalitat l'eliminació de les restriccions de còpia, redistribució, comprensió i modificació de programes d'ordinador. Per a assolir aquest objectiu promouen el desenvolupament i l'ús de programari lliure en totes les àrees de la computació. Per això es permet l'accés al codi font i fer canvis a qualsevol programa de forma lliure. D'aquesta manera es pot modificar *R* per adaptar-lo a les nostres necessitats.

Nosaltres, vam utilitzar el programa específic per a microarrays que s'anomena **Spot** (<http://spot.cmis.csiro.au/spot/index.php>). És un programari elaborat pel CSIRO australià (de l'anglès, *Commonwealth Scientific and Industrial Research Organization*). El programa *Spot* és una llibreria d'R, així un cop les imatges ja estan analitzades, les dades crues estan disponibles en llenguatge R per ser preprocessades. L'anàlisi és molt fàcil: mitjançant la introducció dels paràmetres corresponents al format del nostre xip, el programa *Spot* localitza tots els punts

de la graella de forma automàtica i calcula la intensitat tant del punt, com del fons. A continuació es fa l'anàlisi de les dades gràcies a una aplicació feta a mida amb el llenguatge d'R. L'anàlisi passa obligatòriament, per un filtratge dels punts dolents (els que no es poden aprofitar), el càlcul de les intensitats netes, i la normalització. Els detalls s'expliquen a continuació.

Càlcul de les intensitats netes

Les imatges escanejades tenen una resolució de 16 bits, per tant els valors de les intensitats obtingudes van de 2^0 (0 unitats) a 2^{16} (65.000 unitats). De la intensitat de cada punt es resta la intensitat del fons, i així obtenim les intensitats netes amb les que treballarem per la resta de l'anàlisi.

De cada hibridació tenim un resum dels resultats de tots els punts exemplificats amb les gràfiques de la Figura 22.

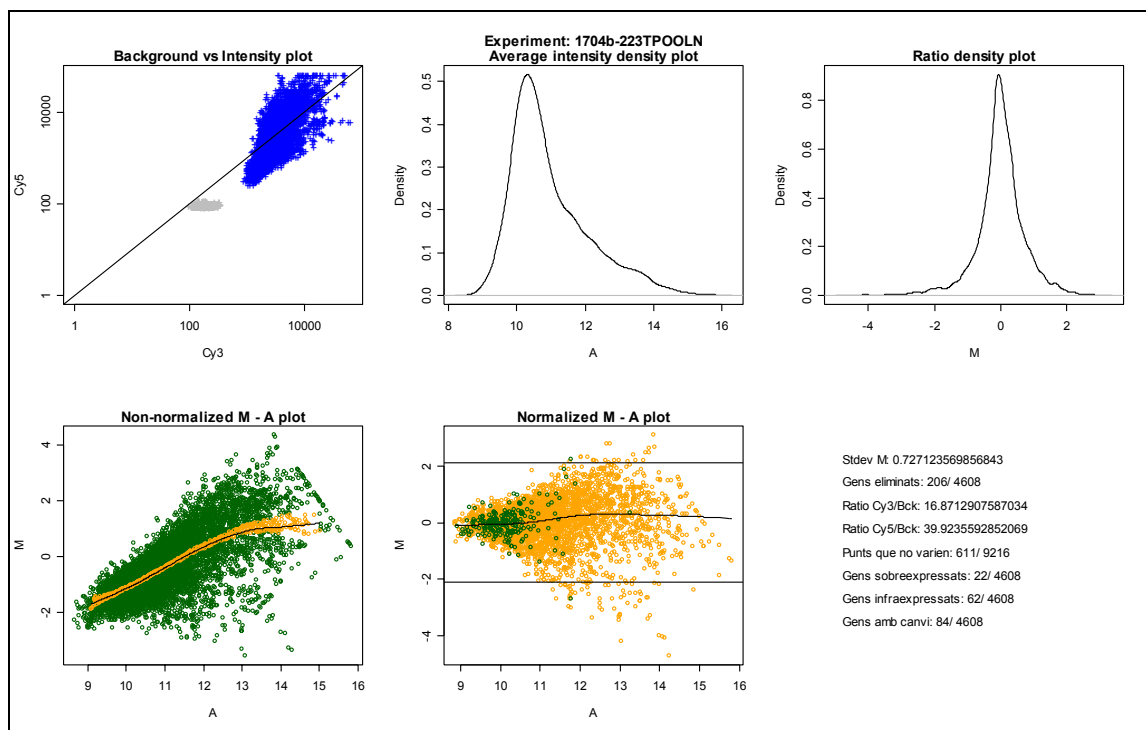


Figura 22. Exemple dels resultats de la hibridació del cas 223 i el POOLN.

Filtratge

És necessari fer un filtratge de les dades inicials per minimitzar els efectes distorsionadors d'elements com la pols, bombolles, rascades, etc. En el nostre cas vàrem decidir dissenyar un mètode automàtic de filtratge dels punts en funció d'alguns paràmetres de qualitat que ens ofereix el programa *Spot*: la circularitat, l'àrea i el perímetre. Així vam imposar un llindar que eliminaria aquells gens que tinguessin les dues rèpliques amb una àrea correcta, però que el coeficient de

variació entre les seves raons fos superior a 0,3. Per a aquells gens on només hi hagués un punt correcte (amb una àrea superior a 30) vam decidir quedar-nos amb aquest per no perdre tanta informació. D'aquesta manera, a cada hibridació es descarten entre un 5 i un 15% dels gens del xip.

Aquestes decisions van ser preses després de fer un estudi exhaustiu de diverses hibridacions i fent una inspecció visual dels punts. L'avantatge d'imposar un llindar determinista, és que l'error serà sempre el mateix, i no dependrà de la subjectivitat de l'investigador. A més, la feina de repassar visualment una imatge amb 9216 punts és poc fiable i pesada.

Normalització

La normalització de les dades és necessària per poder comparar diferents experiments entre ells, és dir, per poder relacionar les intensitats de dos canals d'hibridacions diferents.

En el nostre cas, degut al disseny experimental, tots els casos havien estat hibridats amb una mateixa mostra referència. Aquesta es va aconseguir fent una mescla del producte de la RAP-PCR de 12 teixits normals. Per tant, teníem les dades de 50 hibridacions d'aquesta referència. A l'hora de normalitzar vam escollir les 13 hibridacions més bones visualment, i vam calcular la mitjana de cada gen. Això ens donà que la hibridació anomenada de REFERÈNCIA seguia una determinada distribució (Figura 23).

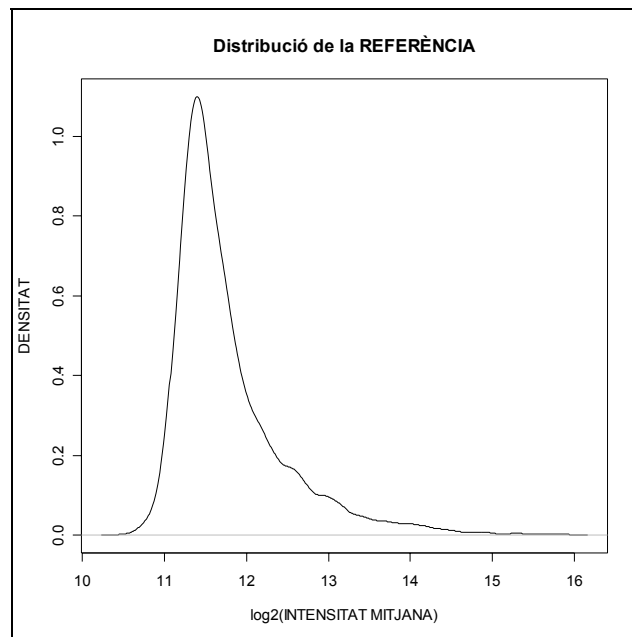


Figura 23. Gràfica de densitat de les intensitats mitjanes de la hibridació de 13 mostres REFERÈNCIA (mescla de 12 teixits normals). A l'eix de les abscisses es representen les intensitats expressades en logaritme en base 2.

A continuació es van normalitzar les intensitats per quantils respecte a aquesta hibridació referència. De manera que les intensitats dels 50 casos seguissin una distribució igual que la distribució de referència.

Càlculs

Un cop disposem de les dades del canal dels 50 teixits normals i dels 50 teixits tumorals normalitzades, vam procedir al càlcul de les raons. En el nostre cas, com que la REFERÈNCIA procedia de teixit normal, les raons entre el cas i la referència, equivalen a la divisió entre el teixit tumoral i un teixit normal. Encara que cal tenir en compte, que no estem parlant del teixit normal de la mostra aparellada. A la Figura 24 s'esquematitzen els càlculs realitzats.

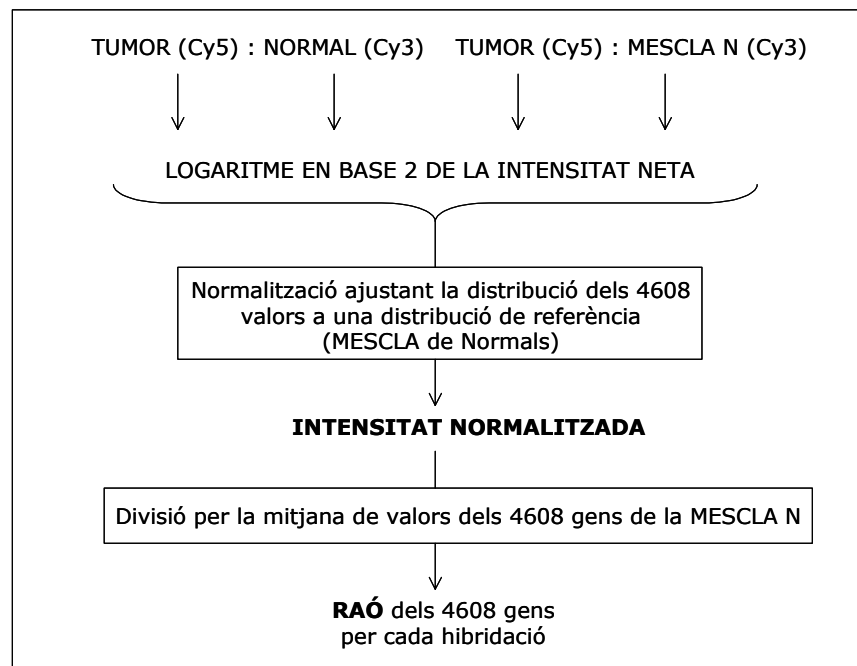


Figura 24. Esquema dels càlculs realitzats amb totes les hibridacions.

4.9. Anàlisi estadístic

Classificar

De forma general, classificar pot ser merament una forma convenient d'organitzar una gran quantitat de dades que ens serveix per poder extreure la informació de manera més eficient. El fet de descriure patrons de similituds i diferències entre els objectes en estudi mitjançant l'adjudicació a diferents classes o grups és una manera de resumir les dades.

Aquesta definició general té moltes aplicacions possibles, des dels estudis de mercat, fins la medicina i la biologia. En el cas de la medicina és molt útil per tal de distingir malalties que poden semblar la mateixa a nivell fenotípic, però que són originades per causes diferents. Així, en el nostre estudi, tot i que parlem de la mateixa malaltia, el càncer colorectal, sí que podem afirmar hi ha molts possibles camins per arribar-hi. Gràcies als estudis moleculars, tant a nivell de DNA com RNA, es pot diferenciar una malaltia a nivell genòmic o d'expressió.

Anàlisi de *clusters* no supervisats (anàlisi multivariable de classificació)

L'anàlisi multivariable de classificació, més freqüentment anomenada anàlisi *cluster*, és un mètode estadístic descriptiu que té per objectiu la recerca de grups similars de variables que es van agrupant en conglomerats (en anglès, *clusters*). Aquestes agrupacions desconegudes a priori, són suggerides per la pròpia essència de les dades. El resultat obtingut el formen les classes o clusters "interpretables" (és a dir, que els grups formats siguin homogenis dins el grup, i heterogenis entre ells).

L'interès d'aquest tipus de classificació rau en la possibilitat que ofereix d'analitzar i interpretar un gran nombre de dades mitjançant l'agrupament, generant així una informació més senzilla i pràctica.

Abans d'iniciar un anàlisi de clusters cal prendre tres decisions clau:

1. Seleccionar les variables
2. Escollir la unitat de mesura de les distàncies
3. Escollir el mètode de classificació de les dades

En el nostre cas les variables eren de dos tipus. En l'estudi de CGH, els **guanys i pèrdues** dels cromosomes. I en l'estudi de microarrays, les dades d'**expressió dels gens**.

Pel què fa a la mesura de les similituds, la majoria de mètodes utilitzen la **distància euclídea al quadrat**.

I per últim, el mètode de classificació és el que definirà com s'agrupen les mostres per tal de formar els clusters a analitzar. Dins dels mètodes de classificació no supervisats, es distingeixen dues categories:

- a) els **jeràrquics** permeten la construcció d'un arbre de classificació, denominat dendrograma, que serveix per visualitzar els resultats obtinguts.
- b) els **no jeràrquics**, tenen l'objectiu de realitzar una sola classificació dels individus en K grups. Això implica que l'investigador ha d'especificar a priori el número de grups que han de ser formats.

En el nostre cas, vam escollir un mètode de classificació jeràrquic anomenat Ward (Ward, 1963).

Aquest mètode calcula la mitjana de totes les variables de cada cluster. Després calcula la distància euclídea al quadrat entre cada mostra i la mitjana del seu grup. Per últim se sumen les distàncies de tots els casos. A cada pas de l'anàlisi, es proven totes les unions possibles de parelles de clusters, i s'escullen aquells que resulten amb el menor increment de la *pèrdua d'informació*. Aquest terme, de *pèrdua d'informació*, va ser definit per Ward com a l'error de la suma total de les distàncies al quadrat (ESS, de l'anglès *error sum-of-squares*).

Anàlisi de components principals (ACP)

L'ACP és un mètode estadístic utilitzat per identificar patrons a les dades, i expressar els resultats de manera que es ressalten tant les diferències com les similituds. Es basa en la construcció de les matrius de variància i covariàncies de totes les observacions (en aquest cas raons dels gens), per tal de descompondre les dades en unes noves variables. Així es defineixen N espais vectorials (el mateix nombre que variables) que presenten una variància màxima i per tant són totalment independents (ortogonals en l'espai) entre ells. D'aquesta manera el primer eix vectorial és el que explica la màxima variància, el segon eix vectorial explica la màxima variància restant, i així successivament. D'aquesta manera cada eix factorial queda definit per una combinació lineal de diversos gens, on cada un d'ells té un pes determinat (o contribució) a l'hora d'explicar la variància total d'aquell eix.

