

## **V. RESULTATS**



## 5. 1. Hibridació genòmica comparada

### a) Perfil

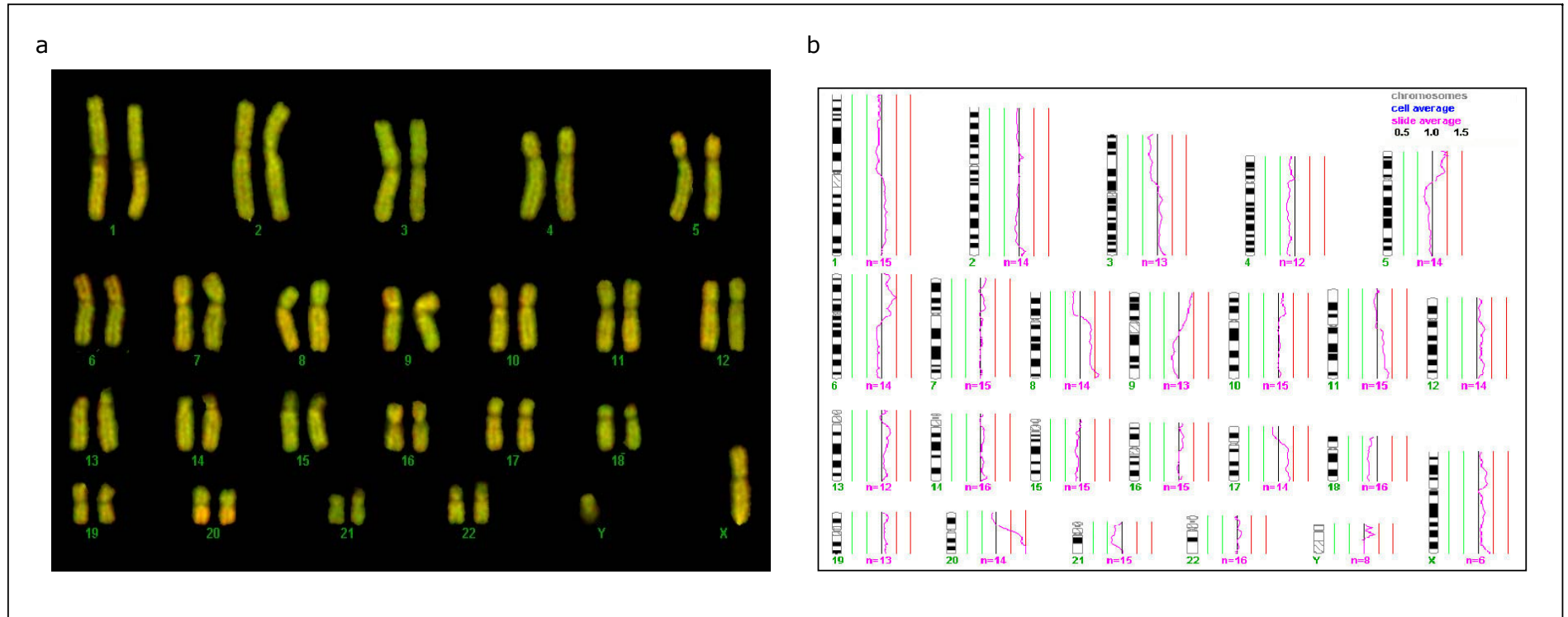
El resultat de la CGH s'obté de la raó del senyal de la fluorescència entre el cas en estudi (el tumor), i la mostra control de sang perifèrica. Per a cada cromosoma s'estudien un mínim de 8 metafases i el resultat es dibuixa esquemàticament en forma d'un perfil molecular (Figura 25). Sota de cada cromosoma es detallen el nombre de metafases que s'han estudiat per dibuixar aquell perfil. A la dreta de la línia negra es representen els guanys del tumor respecte al control, i a l'esquerra les pèrdues. A partir d'aquest perfil s'obtenen els guanys i pèrdues totals de cada cas, on es detallen totes les alteracions cromosòmiques. A tall d'exemple es detalla el resultat de la CGH del cas 109: -3p, -4, +5p, +6p, -8p, +8q, +9p, -9q, +10p, +11q, +12,-17p, +17q, -18, -20p, i +20q (Figura 25).

Un altre punt a considerar és la resolució de la tècnica. S'ha comprovat que el mínim canvi necessari per a ser observat és de 10 Mb (Kallioniemi et al., 1993), però això no és aplicable en tots els casos. Quan veiem un guany d'un tros de cromosoma ens pot estar indicant el guany de tota aquella regió major a 10 Mb, o bé que hi ha una amplificació d'un tros més petit, però com que hi ha moltes còpies el senyal de la fluorescència s'estén per una regió més gran. Així i tot, en la majoria de casos quan hi ha una amplificació es pot distingir perquè es representa com a un pic més acusat que supera els guanys en 1,5 vegades. El problema es troba en braços curts, com el cromosoma 20q, que sovint apareix guanyat per damunt de la línia d'1,5 perquè de seguida es pot saturar el senyal.

### b) Idiograma dels 50 casos

Una de les maneres de representar gràficament el resultat global de tots els casos de la CGH, és mitjançant un **idiograma**. En aquesta sèrie de 50 casos hem trobat alteracions molt recurrents com és el guany del braç 20q que afecta a un 74% dels casos, o la pèrdua del braç 18q que afecta a un 66% dels casos (Figura 26). Pel què fa al cromosoma 20 s'han dibuixat sempre com a guanys, tot i que hi ha molts casos que es pot considerar com a una amplificació, però com que és un braç curt és més difícil distingir entre el guany de poques còpies i l'amplificació.

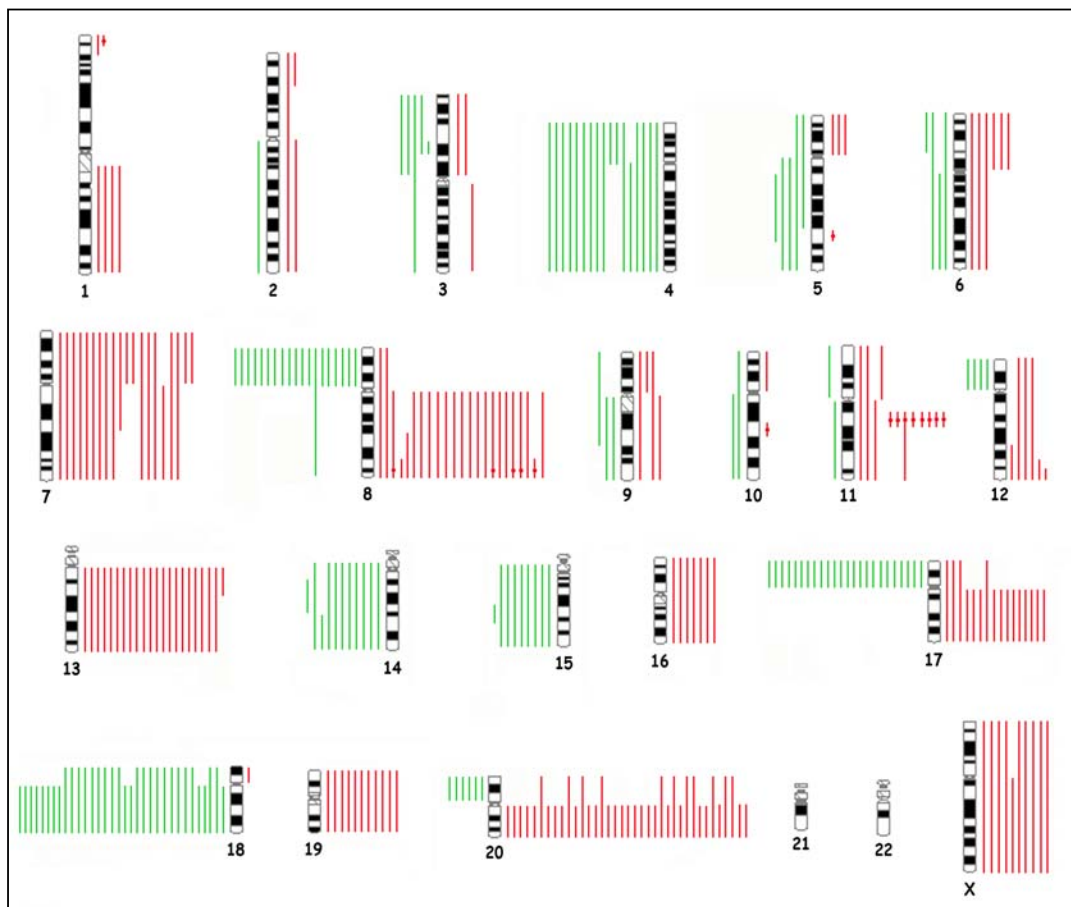
També hem observat que hi ha cromosomes que es guanyen i perden tots sencers o bé només un dels braços i en ambdós sentits, com són el 6 i el 9. En canvi d'altres sempre es veuen alterats en el mateix sentit, és a dir es perden sempre (cromosomes 4, 14 i 15), o es guanyen sempre (cromosomes 1, 7, 13, 16, 19 i X).



**Figura 25. Perfil i metafase del cas 109. (a)** CGH del cas 109. Resultat de la hibridació sobre cromosomes metafàsics. El tumor es marca amb el fluorocrom de color vermell i el control de verd. **(b)** Perfil mitjà de cada cromosoma del cas 109. El nombre de metafases analitzades s'indica a sota de cada cromosoma. La línia negra representa la raó de 1 (sense canvi), a la dreta d'aquesta es representen els guanys (>1,25) i a l'esquerra les pèrdues (<0,75). La raó del senyal de la fluorescència entre el tumor i el control es representa de color lila, i quan no hi ha canvi segueix la línia negra del mig, com és el cas del cromosoma 1.

Paral·lelament, també hem pogut identificar alteracions subcromosòmiques que afecten a una regió més petita que un braç. Però no hem identificat regions mínimes perquè tampoc disposàvem de molts casos. Es pot observar com la regió telomèrica de l'1p es veu guanyada en dos casos, o la pèrdua del cromosoma 5p que no sempre afecta a tot el braç. També es perden regions petites dels cromosomes 14 i 15.

Finalment, cal destacar que hi ha cromosomes que no es veuen mai alterats, com són el 21 i el 22. Però com que són els més petits sovint la CGH no és fàcil de valorar. Així com el cas del cromosoma Y que en aquest estudi no el vam tenir en compte, ja que és un cromosoma molt petit i amb molta heterocromatina. A més s'ha descrit la seva pèrdua en homes d'edat avançada, independentment de la presència de càncer (Mitelman et al., 1997).



**Figura 26. Idiograma** dels 50 casos de càncer colorectal de l'estudi. Es tracta del dibuix esquemàtic de cada un dels cromosomes, amb una línia a la dreta o a l'esquerra, en funció de si s'ha guanyat o perdut aquell tros de cromosoma respectivament. El codi de colors és arbitrari, i en el nostre cas, hem pintat de vermell els guanys (ja que el tumor es va marcar amb el fluorocrom *Spectrum Red*), i de verd les pèrdues (*Spectrum Green*). Les amplificacions, per altra banda, es representen mitjançant un cercle.

Les diferències en el número d'alteracions entre cada cromosoma ens poden donar idea de la importància d'aquell fragment en el procés tumoral. A continuació hi ha la relació dels resultats cromosoma per cromosoma en forma d'histograma (Figura 27). És una altra manera de representar els resultats que ja hem comentat abans amb l'idiograma. També ens ajuda a veure quin és el sentit majoritari de les alteracions en cada cromosoma. Així és fàcil apreciar com en la majoria de cromosomes hi ha una tendència, a guanyar-se o perdre's; i només en algun cas, com en el braç 12p hi ha quasi tants guanys com pèrdues.

### c) Detall de les alteracions cromosòmiques

A la Taula 22 podeu veure el total d'alteracions cromosòmiques presents a la sèrie dels 50 casos. Alhora es descriuen també totes les característiques clinicopatològiques i altres dades moleculars de les quals es disposa.

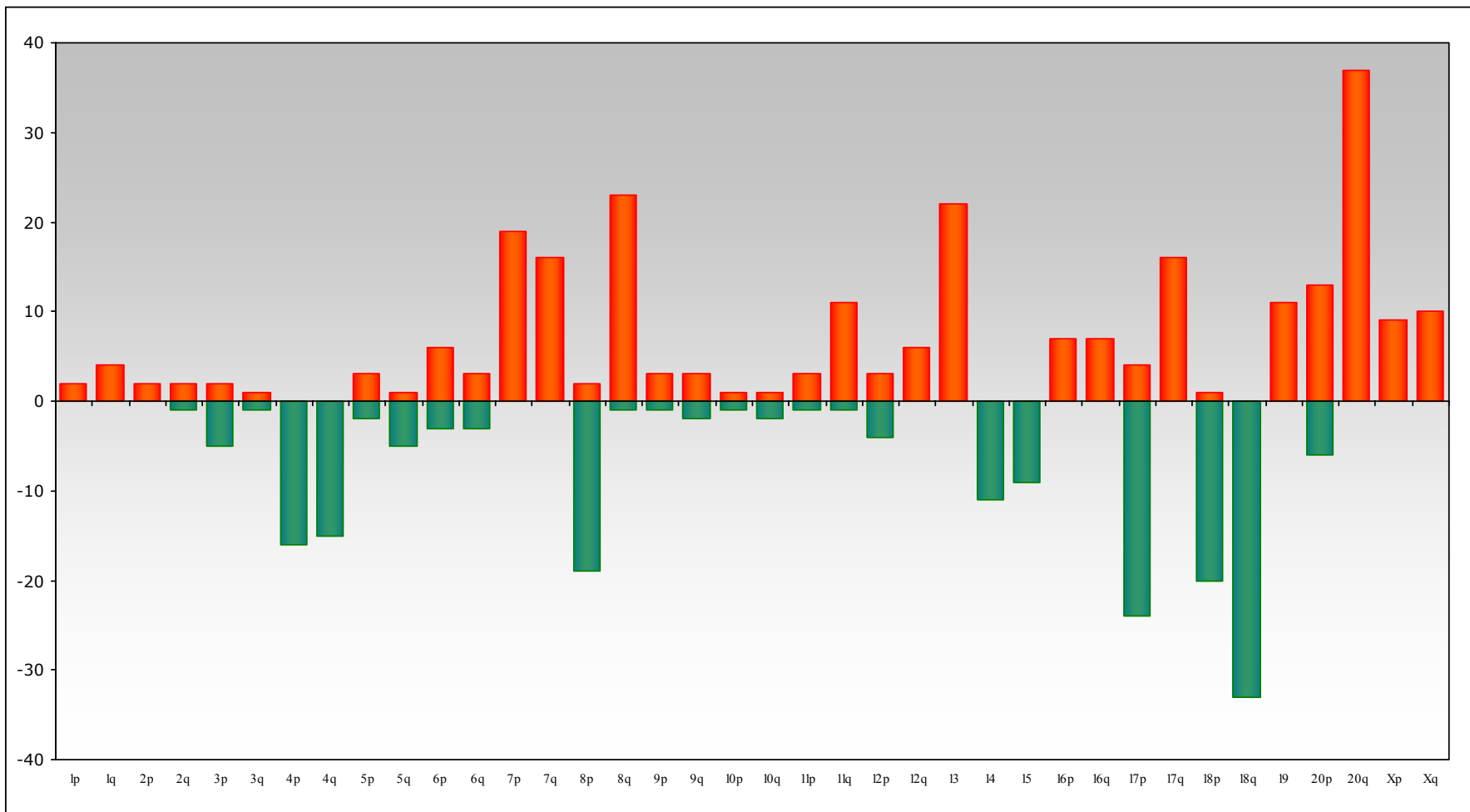
### **Alteracions cromosòmiques totals i recurrents**

Com ja s'ha explicat a l'apartat de *Materials i mètodes* quan parlem del número d'alteracions hem considerat com a base el braç cromosòmic, amb l'excepció dels cromosomes telocèntrics (13, 14 i 15) i el cromosoma 19 (on totes les alteracions detectades afecten el cromosoma complet) que es consideren com una unitat.

Comptat d'aquesta manera, els 50 casos presenten un total de 440 alteracions. El rang de variació va des de 0 fins a 20 alteracions. La mitjana és de 8,8 alteracions per cas amb una desviació estàndard de 6,3. Hi ha 9 casos sense cap alteració (18%) i 41 casos amb algun braç de cromosoma alterat (82%).

Si ho desglossem entre guanys i pèrdues, hi ha 186 braços de cromosoma que es perden i 256 que es guanyen. La mitjana de pèrdues per cas és de  $3,72 \pm 0,4$  i la de guanys és de  $5,1 \pm 0,6$ . Per tant hi ha més trossos de cromosoma que es guanyen que no pas es perden amb una raó de guanys dividit per pèrdues de 1,4.

Si ens fixem en les alteracions més recurrents, veiem que hi ha molts cromosomes alterats en més d'un 15% dels casos (Taula 23). Totes elles s'han descrit anteriorment en el càncer colorectal, i ens ajuden a definir el perfil d'aquest tipus de tumor. Les alteracions més freqüents són +20q (74%), -18q (66%), -17p (48%), +13 (44%) i -18p (40%).



**Figura 27. Histograma de freqüència dels guanys i pèrdues cromosòmiques.** Amb vermell i valors positius es veuen els guanys, i amb verd i negatiu les pèrdues.

**Taula 22.** Resum de les alteracions cromosòmiques dels 50 casos de la sèrie.

nº cas	guanys	pèrdues	edat <sup>(1)</sup>	sexe	localització	estadi de Dukes <sup>(2)</sup>	quimoteràpia	radioteràpia	sense recidiva després de 44 mesos	MSI <sup>(3)</sup>
9	X	-	67	home	dret	B2	no	no	sí	sí
11	7p, 17, 18p, 20	5pter-q23, 18q	54	home	esquerre	B2	sí	sí	sí	no
12	-	-	55	home	esquerre	B2	sí	no	sí	no
14	8q22-qter, 13, 17q, 19, amp20q, X	2q, 4, 17p, 18, 20p	64	dona	esquerre	B1	no	no	sí	no
16	20q	18	70	dona	esquerre	C2	no	no	sí	no
17	2q, 6, 13, 17q, 19, amp20q	4, 17p, 18	31	dona	esquerre	C2	sí	sí	sí	no
19	20	14, 18q	49	dona	esquerre	B1	no	no	no	no
21	-	-	68	dona	esquerre	B1	no	no	sí	no
23	6, 7, 8, amp11q13, 13, 17q, 20p, amp20q, X	4, 17p, 18	65	home	esquerre	B2	no	no	sí	no
25	6, 7, 8q, 13, 16q, 17, amp20q	4, 8p, 18q	67	home	esquerre	B1	no	no	sí	no
46	7, 8, 13, 19, 20q	3, 4, 5q, 14, 18q	68	home	dret	B2	no	no	sí	no
53	8q, 12, 16, 20q, X	6q, 8p, 17p, 18, 20p	61	dona	esquerre	C2	sí	sí	sí	no
63	8q, 13, amp20q	8p, 17p, 18	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	6p, 8q, amp11q13, 12q22-qter, 17, 20	4, 5, 18	73	home	dret	C2	no	no	no	no
74	1q, 7, 8q24-qter, 13, 20q, X	14, 17p, 18	86	home	esquerre	C2	no	no	sí	no
75	7, 13, 17q, 20, X	6p2-pter, 17p, 18q	76	home	esquerre	B2	sí	no	no	no
78	17q, 20q	6, 17p, 18q	62	home	esquerre	B2	sí	no	sí	no
81	19	3p14-pter, 6, 17p, 18q	67	dona	dret	B2	no	no	sí	no
90	-	-	88	dona	esquerre	C2	no	no	no	no
99	7	3p	65	home	dret	B2	no	no	no	sí
102	8q, 20p, amp20q	8p, 17p, 18	78	home	dret	B2	no	no	no	no
103	-	-	80	home	esquerre	B2	no	no	no	no
104	7q, 8q, amp8q24, 13	8p, 12p, 17p	62	dona	dret	C2	sí	no	no	no
108	7p, 20p, amp11q13, amp20q	4q, 8p, 18	61	home	esquerre	C2	sí	no	no	no
109	5p, 6p, 8q, 9p, 10p, 11q, 12, 17q, amp20q	3p, 4, 8p, 9q, 17p, 18, 20p	43	dona	dret	B2	sí	no	sí	no
113	1q, 2, 7, 11, 13, 19, 20q	4p, 8, 14, 15, 17p, 18q	73	dona	dret	B2	sí	no	no	no
114	1q, 13, 20q	14, 17p, 20p	76	dona	dret	B2	no	no	sí	no
118	8q, 13, 20q	4p, 8p, 18q	51	dona	esquerre	C2	sí	sí	no	no
122	3p, 7, 8q, 13, 17q, 20q	8p, 15, 17p, 18	62	home	esquerre	C1	no	no	sí	no
124	7, 13, 16, 20q	4, 17p, 18, 20p	74	dona	esquerre	C2	sí	sí	sí	no
127	7p, 8q, 16, 20	18	30	dona	esquerre	C2	sí	no	sí	no
135	-	-	63	dona	esquerre	B2	sí	sí	sí	no
138	8q, 13, 17q, 20q	8p, 17p, 18q	80	dona	esquerre	B2	no	no	sí	no

143	-	-	37	dona	dret	C1	sí	no	sí	no
144	8q, 13, 20q	8p	72	dona	esquerre	B2	no	no	no	no
145	7, 8q, amp8q24, 9, 13, 20	8p, 14, 15, 17p, 18q	73	dona	dret	B2	no	no	sí	no
146	3q, 7pter-q22, 11, amp20q	4	65	dona	esquerre	B1	no	no	sí	no
147	5p, 6p, 7, 8q24-qter, amp8q24, 9, 20q	3p14, 4, 11p, 12p, 14, 17p, 18	76	dona	esquerre	C2	no	no	no	no
151	1pter-p32, 1q, 5p, 9q31-qter, 11p, 12, 17, 18p, 19, 20	4, 5q13-q31, 10, 11q, 15, 18q	79	home	dret	B2	no	no	no	no
154	2pter-p16, 3pter-p14, 7p, 9q, 13, 19, 20, Xq	4, 5q, 8p, 12p, 17p, 18q	54	home	esquerre	C2	sí	sí	no	no
165	7, 8q, amp8q24, amp10q22, 12q43-qter, 16, 17q, 19, amp20q, X	8p, 17p, 18, 20p	76	home	dret	C2	no	no	no	no
171	amp5q31, 7, 8q, 13, 16p, 20, X	8p, 15	85	home	esquerre	C2	no	no	sí	no
175	-	-	46	dona	esquerre	B2	sí	sí	no	no
186	amp11q13, 20q	-	56	home	esquerre	B2	no	no	no	no
187	7, 8q, amp11q13, 13, 16, 17q, 19, amp20q, X	4, 8p, 14, 15, 17p, 18	63	home	esquerre	B2	no	no	no	no
223	amp1p36, 8q, amp8q24, amp11q13, 16, 17q, 20q	8p, 10q22-qter, 14q24-qter, 15, 17p, 18	75	home	esquerre	B2	no	no	no	no
271	amp11q13, 13, 17q, 19, 20	14, 15, 17p, 18	67	home	esquerre	B2	sí	no	no	no
282	8q, amp11q13, 12q24-qter, 13q12-q14, 19, 20q	4, 8p, 9p, 12p, 14q12-q22, 15q15-q23, 18	78	home	esquerre	C2	no	sí	no	no

- (1) edat en anys
- (2) segons la classificació d'Astler-Coller
- (3) inestabilitat de microsatèl·lits

**Taula 23.** Alteracions cromosòmiques recurrents detectades almenys al 15% dels casos. Ordenades de major a menor freqüència.

	<b>cromosoma</b>	<b>casos (N)</b>	<b>freqüència (%)</b>
<b>pèrdues</b>	<b>18q</b>	<b>33</b>	<b>66</b>
	17p	24	48
	18p	20	40
	8p	19	38
	4p	16	32
	4q	15	30
	14	11	22
	15	9	18
	<b>guanys</b>	<b>20q</b>	<b>37</b>
8q		23	46
13		22	44
7p		19	38
7q		16	32
17q		16	32
20p		13	26
11q		11	22
19		11	22
Xq		10	20
Xp		9	18

#### d) Grups de tumors en funció de les alteracions cromosòmiques

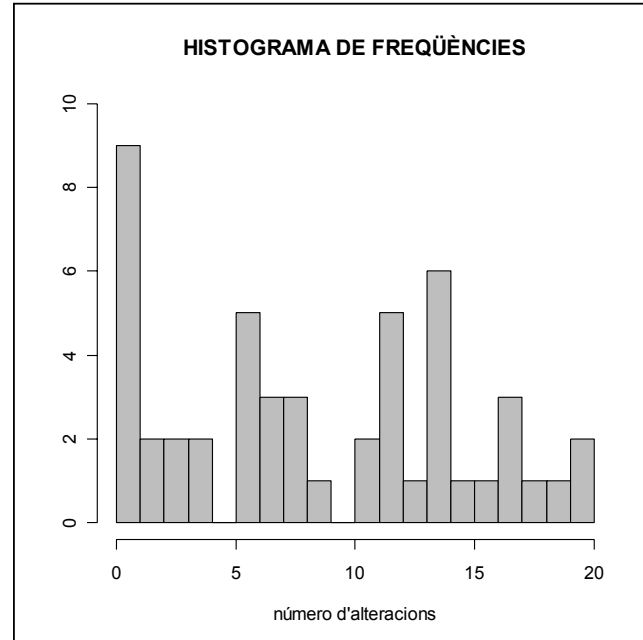
Amb l'objectiu de classificar les mostres en diferents grups a partir de les dades obtingudes per CGH es van utilitzar diferents criteris:

1. El **número d'alteracions**. És una característica que ja ha estat utilitzada en estudis anteriors per delimitar grups de casos amb comportament diferent (Chan et al., 2001).
2. Si estan **perduts els braços 17p i 18q**. Ens hem basat en la classificació de Dutrillaux (Dutrillaux, 1995) on definia un grup anomenat *monosòmic* que tenia com a característica principal que el 100% dels casos havien perdut el cromosoma 18 i el 17p (vegeu l'apartat de la *Introducció*).
3. **Agrupaments jeràrquics** que s'utilitzen per classificar casos en funció de les distàncies mesurades en valors numèrics (en el nostre cas, guanys o pèrdues).

##### *d1) Classificació segons el número d'alteracions*

Com podem veure a la Figura 28, el número d'alteracions oscil·la entre 0 i 20 per cas. En una anàlisi inicial determinem si existeix una relació entre les característiques clinicopatològiques i moleculars dels tumors i el número d'alteracions (Taula 24). Com es pot observar, els tumors amb amplificacions

gèniques i els tumors amb mutacions a TP53 tendeixen a tenir un número més alt d'alteracions cromosòmiques (ressaltats amb negreta). També veiem un resultat significatiu pel què fa al grup MSI, però tractant-se només de tres casos li donem una importància relativa.



**Figura 28.** Histograma de freqüències del nombre d'alteracions cromosòmiques de la sèrie de 50 casos.

**Taula 24.** Nombre d'alteracions cromosòmiques en referència a característiques clinicopatològiques i moleculars.

paràmetre	categories	N	nombre d'alteracions	P
<b>sexe</b>	home	26	9,88 ±6,50	0,206
	dona	24	7,63 ±5,91	
<b>estadi de Dukes</b>	B	30	8,30 ±6,70	0,495
	C	20	9,55 ±5,64	
<b>localització</b>	esquerra	35	8,46 ±6,00	0,560
	dreta	15	9,60 ±7,01	
<b>supervivència</b>	SÍ	35	8,32 ±5,79	0,439
	NO	15	9,81 ±7,29	
<b>TP53</b>	wt	15	5,13 ±6,10	<b>0,006</b>
	mutat	35	10,37 ±5,73	
<b>KRAS</b>	wt	33	8,06 ±6,07	0,249
	mutat	17	10,24 ±6,59	
<b>MSI</b>	SÍ	3	1,67 ±1,53	<b>0,041</b>
	NO	47	9,26 ±6,18	
<b>amplificacions<sup>a</sup></b>	SÍ	19	12,37 ±4,95	<b>0,001</b>
	NO	31	6,61 ±6,04	

<sup>a</sup> Les amplificacions no s'inclouen en el còmput total d'alteracions cromosòmiques.

Segons el número d'alteracions es poden fer diverses classificacions. En base a la distribució observada podríem parlar de: tumors sense alteracions (N=9), amb molt poques (menys de 5) (N=6), tumors amb un número intermig d'alteracions (entre 5 i <10) (N=12) i finalment tumors amb moltes alteracions ( $\geq 10$ ) (N=25).

Observem com el nombre d'alteracions no està associat a cap paràmetre clinicopatològic (Taula 24). Les úniques associacions que trobem són un major nombre d'alteracions en els tumors amb mutació a TP53, i la presència d'amplificacions gèniques.

#### *d2) Classificació segons si estan alterats els braços 17p i 18q*

Un altre criteri alhora de classificar els tumors, és decidir quins són aquells canvis que poden definir un conjunt de casos amb un sentit biològic. Segons els grups que va fixar Dutrillaux (vegeu l'apartat de la *Introducció*), el grup dels *monosòmics* es caracteritzava per tenir pèrdues dels cromosomes 17p i 18 en el 100% dels casos. Com que a la nostra sèrie també hi ha molts casos amb aquesta alteració vam definir un grup com a tal. És a dir, tots aquells casos que tinguessin perduts els braços 17p i 18q alhora, formen part d'un grup que anomenem *inestables cromosòmicament monosomic like (ICML)*. La resta els subdividim en: aquells casos que no tenen cap alteració a nivell cromosòmic, que anomenem **normals**; i els casos que presenten alguna alteració diferent a -17p i -18q, i que els anomenem *inestables cromosòmicament restants (ICR)*.

Si comparem les característiques clinicopatològiques en funció dels tres grups creats trobem diferències significatives en funció de la presència d'amplificacions i de TP53 mutat (Taula 25).

Podem dir, en general, que el grup **normal** és un grup heterogeni pel què fa a la supervivència, presenten poques mutacions al gen TP53, i es localitzen al còlon esquerre (a excepció d'un sol cas).

Pel què fa al grup que anomenem **ICML**, podem dir que és el que té un nombre d'alteracions més elevat, i que a nivell qualitatiu tenen els cromosomes 17 i 18 afectats.

Finalment, el grup **ICR**, tenen un nivell d'alteracions intermig i molt heterogeni. Destaquen quatre casos que presenten guany del braç 17p, contràriament al que passa amb la gran majoria de tumors colorectals. Pel què fa a la resta de característiques clinicopatològiques són semblants al grup anterior per ser molt heterogènies.

**Taula 25.** Resum de les característiques de cada grup basat en els braços 17p i 18q.

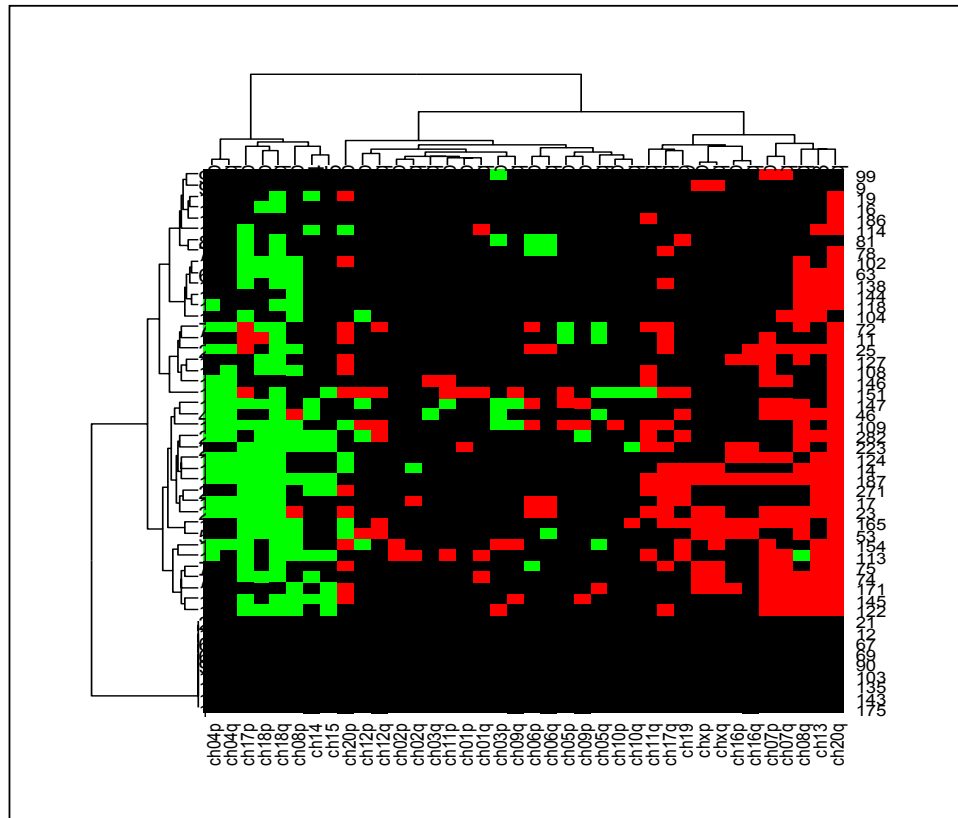
característica	normals	ICML	ICR	P
número de casos (N)	8	22	17	-
número d'alteracions total	0	282	153	-
mitjana d'alteracions $\pm$ SD	0	12,8 $\pm$ 0,9	9,3 $\pm$ 0,9	-
casos amb amplifícacions	0	11	8	<b>0,04</b>
TP53 mutat	2	17	15	<b>0,003</b>
KRAS mutat	3	7	7	0,83
estadi de Dukes B/C	5/3	13/9	9/8	0,88
còlon esquerre/dret	7/1	16/6	12/5	0,64
sense recidiva <sup>(1)</sup>	5	14	8	0,55

<sup>(1)</sup> Pacients lliures de malaltia després de 4 anys de seguiment.

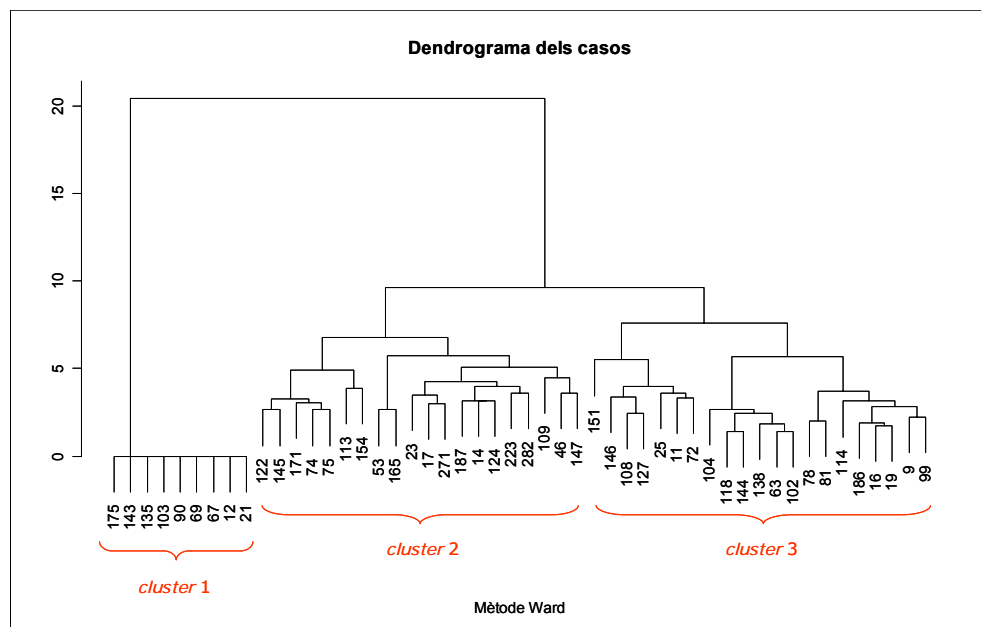
### d3) Classificació segons els agrupaments jeràrquics

Una de les maneres d'agrupar les mostres a partir d'una matriu de dades són els mètodes d'agrupament jeràrquics (o bé, *clusters* jeràrquics). Resumint, vam utilitzar el mètode Ward (vegeu l'apartat de *Materials i mètodes*) per veure com s'ordenaven els casos usant els resultats de la CGH expressats en una matriu de distàncies amb els valors  $-1$ ,  $0$  i  $1$ , segons si aquell braç està perdut, igual, o guanyat. La manera de representar les distàncies és mitjançant dendrogrames. Si dibuixem el dendrograma dels casos com a files i el dels cromosomes com a columnes, i després es pinten les caselles resultants en funció de si hi ha hagut un guany o una pèrdua en colors diferents, obtenim el que s'anomenen *heat maps* (Figura 29).

Si ens fixem en el dendrograma dels casos veiem tres grups principals que anomenem de la següent manera: *cluster 1* (són els casos que no tenen cap alteració), *cluster 2* (són els casos que tenen més alteracions i la majoria presenten pèrdues del braç  $-17p$  i  $-18q$ ), i *cluster 3* (són aquells casos que tenen un nombre d'alteracions intermig).



**Figura 29. Heat map dels 50 casos.** Cada fila correspon a un cas, i cada columna a un tros de cromosoma (dividits en el braç curt i el llarg segons s'explica a l'apartat de Materials i Mètodes). Amb vermell es representen els guanys, amb verd les pèrdues, i amb negre quan no hi ha canvi.



**Figura 30. Dendrograma dels 50 casos.**

Si tornem a mirar les mateixes característiques clinicopatològiques com hem fet abans veiem les següents diferències entre els tres grups:

**Taula 26.** Resum de les característiques de cada grup basat en els dendrogrames.

característica	cluster 1	cluster 2	cluster 3	P
número de casos (N)	9	20	21	-
número d'alteracions total	0	290	150	-
mitjana d'alteracions $\pm$ SD	0	14,5 $\pm$ 2,7	7,14 $\pm$ 4,4	-
casos amb amplifacions	0	11	8	<b>0,019</b>
TP53 mutat	2	15	18	<b>0,002</b>
KRAS mutat	3	9	5	0,36
estadi de Dukes B/C	6/3	10/10	14/7	0,50
còlon esquerre/dret	7/2	15/5	13/8	0,56
sense recidiva <sup>(1)</sup>	6	11	12	0,84

<sup>(1)</sup> Pacients lliures de malaltia després de 4 anys de seguiment.

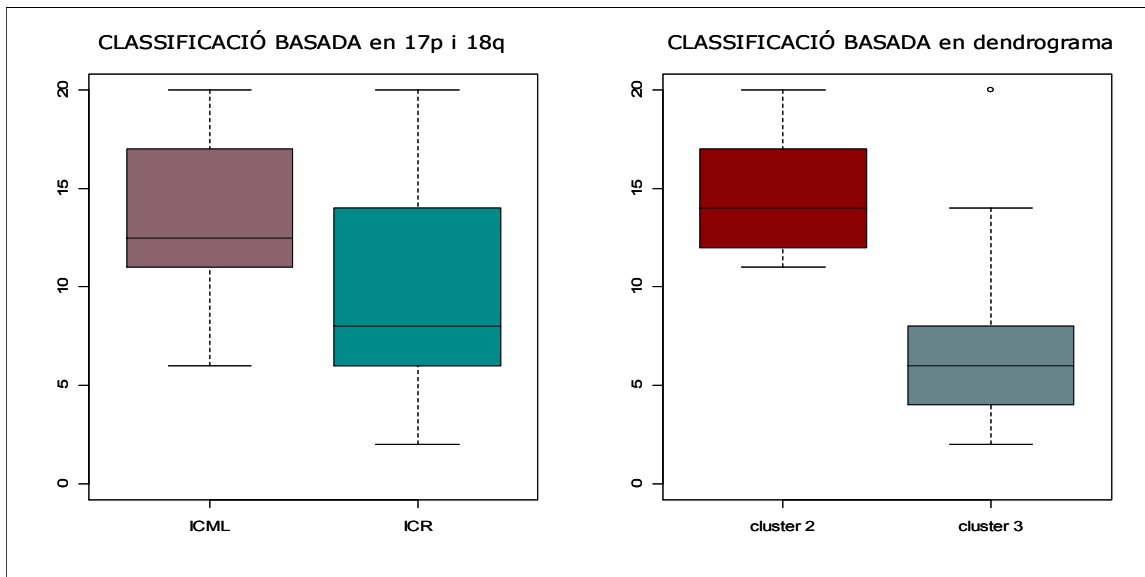
Altres veiem que les diferències significatives són les amplifacions i la mutació del gen TP53.

Igual que a la segona classificació, aquí veiem com el *cluster 2* és el grup amb més alteracions, però amb una diferència més gran del *cluster 3* que en la classificació anterior.

Comparant els casos del grup que hem anomenat *ICML* (basat en els braços 17p i 18q) amb els del grup *cluster 2* (basat en el dendrograma), veiem que la majoria coincideixen. Per tant, la nostra pregunta és: són equivalents els grups *ICML* i *cluster 2*? Si mirem les diferències cas per cas veiem que:

- els casos 46, 171, i 282, del *cluster 2*, tenen pèrdua del braç 18q i no del 17p i per tant no serien del grup *ICML*, però tots tres tenen més de 10 alteracions.
- per altra banda els casos 63, 78, 81, 102, i 138, són del *cluster 3*, i tots presenten pèrdues dels braços 17p i 18q. Però a diferència d'abans, aquests tenen menys de 10 alteracions.

Per tant, tot i tenir unes característiques molt semblants, les dues classificacions no agrupen els mateixos casos. Podem afirmar, que a la classificació basada en els dendrogrames estem separant dos grups amb un número d'alteracions més dispar, mentre que amb la classificació basada en 17p i 18q (qualitativa) hi ha un solapament major del número d'alteracions entre els dos grups (Figura 31).



**Figura 31.** Nombre d'alteracions de cada un dels grups definits seguint dues classificacions independents. A l'esquerra basada en dues alteracions qualitatives (-17p i -18q), i a la dreta basada en la classificació del dendrograma. La línia mitja de dins les caixes correspon a la mediana, el límit inferior de la caixa és el primer quartil, i el límit superior correspon al tercer quartil. Les barres s'allarguen fins als valors extrems.

Resumint, veiem que sigui quin sigui el criteri seguit, les característiques dels 3 grups resultants no varien gaire, i són pocs els casos que ballen d'un grup a l'altre. Per tant, a l'hora de valorar la idoneïtat de les classificacions considerem que la més bona és la qualitativa per dos motius. Primer de tot, perquè és la més senzilla i fàcil d'utilitzar. Només cal mirar els braços 17p i 18q per decidir si un cas pertany a aquell grup o no. I segon, perquè en el cas dels dendrograms només la presència d'un cas nou fa variar les agrupacions i no és un sistema molt pràctic per ser aplicat de manera rutinària.

Per tant, vam arribar a la conclusió que els grups més idonis, eren els sorgits de la classificació basada en criteris qualitius. Així resumint ens quedem amb els següents grups:

- I) el grup MSI (3 casos)
- II) el grup normal a nivell cromosòmic (8 casos)
- III) el grup ICML (22 casos)
- IV) el grup ICR (17 casos)

I) A la nostra sèrie només hi ha 3 casos (6%) que presenten **MSI**. Tal i com s'ha vist a la literatura aquest grup té un pronòstic més bo i estan localitzats al còlon dret. Com és un grup molt ben definit, i representen un tant per cent baix els vam considerar a part. Vegeu la taula a continuació per veure les característiques del grup.

**Taula 27.** Característiques del grup amb *inestabilitat de microsatèl·lits*.

cas	nº d'alteracions	TP53	còlon	Dukes
9	1	mutat	dret	B2
69	0	no mutat	dret	B2
99	2	no mutat	dret	B2

II) El grup **normal** el formen vuit casos (16%), set dels quals estan localitzats al còlon esquerre. És de destacar, que tot i no tenir alteracions hi ha tres casos que presenten recidiva i mort per la malaltia. D'aquests tres, dos tenen el gen TP53 mutat (Taula 28). L'interès d'aquest grup és determinar quina via de progressió segueixen. Perquè no tenen un genoma inestable, i per tant la seva via de progressió és diferent.

**Taula 28.** Característiques del grup de CGH *normal*.

cas	nº d'alteracions	TP53	còlon	Dukes
12	0	no mutat	esquerre	B2
21	0	no mutat	esquerre	B1
67	0	no mutat	esquerre	C2
90	0	mutat	esquerre	C2
103	0	no mutat	esquerre	B2
135	0	no mutat	esquerre	B2
143	0	no mutat	dret	C1
175	0	mutat	esquerre	B2

III) El grup **ICML** es caracteritza perquè presenten pèrdues dels braços dels cromosomes -17p i -18q. Aquest grup amb 22 casos (44%) també es caracteritza per tenir un nombre alt d'altres alteracions cromosòmiques i les següents característiques clinicopatològiques (Taula 29).

De forma general veiem que la majoria de casos tenen el gen TP53 mutat (77%) i que es localitzen al còlon esquerre (73%). Pel què fa al seu estadi de desenvolupament, un cas es troba en l'estadi de Dukes B1, un 54,5% es troben en l'estadi de Dukes B2, un altre cas està a l'estadi C1, i la resta en el C2. I com ja s'ha dit la supervivència és una mica més bona que al grup *ICR*.

IV) En el grup **ICR** hi ha la resta de casos amb alteracions cromosòmiques. És un calaix de sastre on hi ha disset casos (34%) amb diferents característiques, i amb un nombre d'alteracions cromosòmiques intermig (Taula 30).

**Taula 29.** Característiques del grup de CGH *ICML*.

cas	nº d'alteracions	TP53	còlon	Dukes
14	11	no mutat	esquerre	B1
17	9	mutat	esquerre	C2
23	11	mutat	esquerre	B2
53	10	no mutat	esquerre	C2
63	6	mutat	esquerre	C2
74	9	no mutat	esquerre	C2
75	8	mutat	esquerre	B2
78	5	mutat	esquerre	B2
81	5	mutat	dret	B2
102	5	mutat	dret	B2
109	16	mutat	dret	B2
113	13	mutat	dret	B2
122	10	no mutat	esquerre	C1
124	8	mutat	esquerre	C2
138	7	mutat	esquerre	B2
145	10	mutat	dret	B2
147	13	mutat	esquerre	C2
154	14	mutat	esquerre	C2
165	13	mutat	dret	C2
187	15	mutat	esquerre	B2
223	11	no mutat	esquerre	B2
271	9	mutat	esquerre	B2

**Taula 30.** Característiques del grup de CGH *ICR*.

cas	nº d'alteracions	TP53	còlon	Dukes
11	6	no mutat	esquerre	B2
16	2	mutat	esquerre	C2
19	3	mutat	esquerre	B1
25	10	no mutat	esquerre	B1
46	10	mutat	esquerre	B2
72	9	no mutat	esquerre	C2
104	6	mutat	esquerre	C2
108	5	mutat	esquerre	C2
114	6	mutat	dret	B2
118	6	mutat	dret	C2
127	5	mutat	dret	C2
144	4	mutat	dret	B2
146	5	no mutat	esquerre	B1
151	16	mutat	esquerre	B2
171	9	mutat	esquerre	C2
186	2	mutat	dret	B2
282	13	mutat	esquerre	C2

De forma general veiem que la majoria de casos tenen el gen TP53 mutat (76,5%) i que es localitzen al còlon esquerre (71%). Pel què fa al seu estadi de desenvolupament, un 53% són de l'estadi de Dukes B, i la resta del C2. Com veiem a les corbes de Kaplan Meier de la Figura 33 la supervivència és més dolenta que el grup *ICML*.

A nivell dels cromosomes afectats dels dos grups amb inestabilitat cromosòmica veiem diferències en les freqüències d'alteracions en diversos cromosomes, sempre amb una major recurrència en el grup *ICML* que el *ICR*, excepte el 5 (Taula 31).

**Taula 31.** Alteracions típiques del grup *ICML* comparat amb el grup *ICR*.

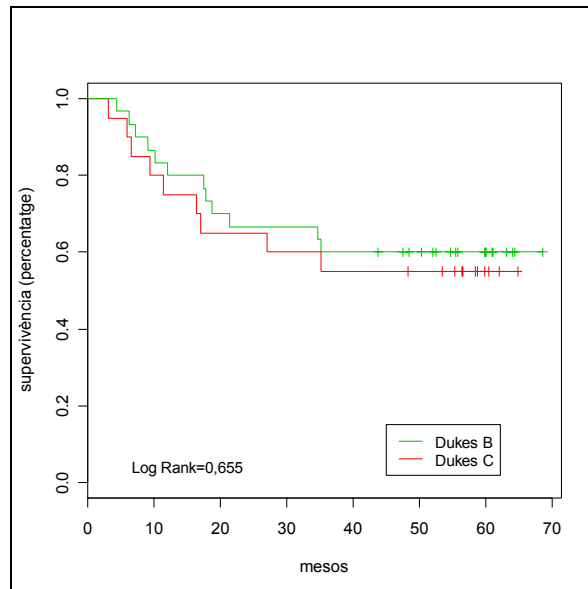
alteració	<i>ICML</i> (N=22)	<i>ICR</i> (N=17)	P
-17p	22 (100%)	2 (12%)	0,00
+17q	12 (54,5%)	4 (23,5%)	0,09
-18p	15 (68%)	5 (29,4%)	0,04
-18q	22 (100%)	11 (65%)	0,00
+X	7 (32%)	1 (6%)	0,05
-5q	1 (4,5%)	4 (23,5%)	0,15

### Supervivència dels diferents grups

Un dels principals objectius alhora de definir grups és valorar la capacitat que tenim per distingir l'agressivitat d'aquests tumors, així com identificar els factors amb possible aplicabilitat pronòstica dins de cada grup. Primerament estudiem la supervivència que tenen els casos en funció del grup que els hem adjudicat. Aquesta supervivència es representa mitjançant les corbes de Kaplan-Meier. En aquest tipus de gràfica l'eix de les abscisses indica el temps de seguiment en mesos, i el de les ordenades la proporció de casos sense malaltia respecte el total en aquell moment. Així, cada esglaó a la gràfica representa un cas de recidiva, i les creus identifiquen aquells casos que surten del seguiment per qualsevol motiu que no sigui la recidiva.

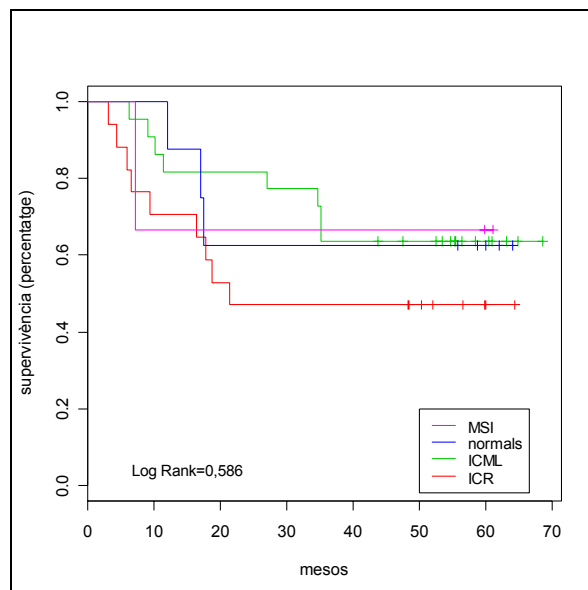
Les corbes de supervivència es comparen mitjançant el test de Log Rank. Aquest test calcula el *p* valor considerant com a hipòtesi nul·la que les corbes de supervivència són idèntiques en els diferents grups. Si aquesta assumpció és certa, el *p* valor és la probabilitat de seleccionar a l'atzar individus que tinguin una corba de supervivència diferent a la observada. Normalment es considera significatiu un valor de Log Rank inferior a 0,05.

Una de les característiques clíniques que tenen un valor pronòstic més clar és l'estadi de Dukes (vegeu l'apartat de la *Introducció*). En aquest conjunt de mostres però, no és així donat que intencionadament hem inclòs casos de característiques semblants (Figura 32), i per tant és més necessari la utilització de marcadors de pronòstic addicionals als ja tradicionals.



**Figura 32.** Corbes de Kaplan-Meier de la supervivència lliure de malaltia en funció de l'estadi de Dukes.

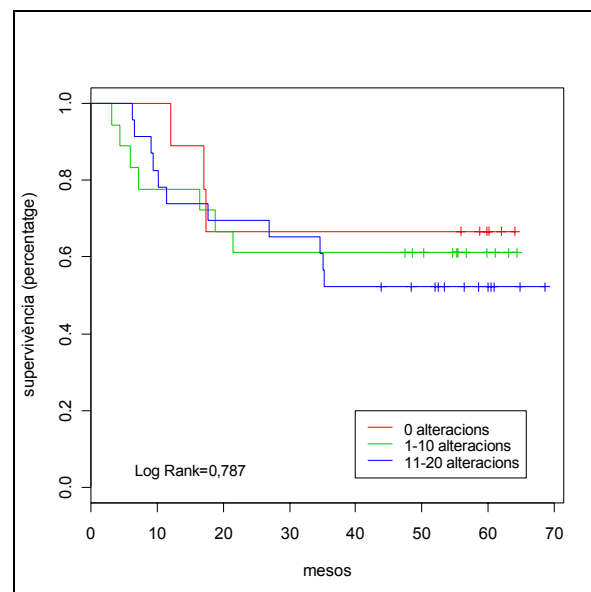
Per valorar la utilitat pronòstica de la classificació basada en les dades de CGH representem les corbes de Kaplan-Meier pels quatre grups que hem definit (vegeu l'apartat de *Resultats*). A la Figura 33 comprovem que només hi ha un grup que té un pronòstic més dolent que la resta: el que hem anomenat ICR; on quasi el 60% presenten recidiva després de 5 anys de seguiment. Els tres grups restants, tenen una supervivència similar. Així en el grup ICML al final del seguiment un 40% tenen recidiva. Al grup dels *normals* veiem que 3 dels vuit casos presenten recidiva, i això representa un 37,5% del total. Finalment tenim el grup amb MSI on només hi ha tres casos, dels quals un presenta recidiva.



**Figura 33.** Corbes de Kaplan-Meier de la supervivència lliure de malaltia per a la classificació basada en els grups de la CGH.

Un altre dels nostres interrogants era veure si en funció del nombre d'alteracions variava la supervivència. És ben sabut que a mesura que els tumors progressen van acumulant alteracions, ja sigui en forma de mutacions puntuals, o per un increment del dany genòmic en general (vegeu l'apartat de la *Introducció*). Però el que no queda clar és si això representa una pitjor supervivència. Per tal de comprovar-ho fem tres grups en funció del nombre d'alteracions a nivell de CGH. Així separem aquells casos que no tenen cap alteració (N=9), dels que tenen d'1 a 10 alteracions (N=18), dels que en tenen d'11 a 20 (N=23). A la Figura 34 podem observar que el grup de casos amb més alteracions és el que té més mal pronòstic, perquè 11 casos dels 23 presenten recidiva (47,8%). El grup amb un nivell intermig d'alteracions també té una supervivència intermitja: 7 dels 18 casos presenten recidiva (38,9%). I finalment els que tenen una supervivència més bona són els que no presenten alteracions a nivell de CGH: 3 dels 9 casos presenten recidiva (33,3%).

Cal destacar que el test Log Rank no és significatiu. És evident que estem barrejant mostres que pertanyen a grups molt diferents, com és el cas dels casos que tenen MSI. Per això s'han de valorar els factors pronòstics del nombre d'alteracions independentment dins de cada grup de mostres que hem definit. Perquè és fàcil pensar, que pot ser tant o més important el tipus d'alteracions com el número d'alteracions.



**Figura 34.** Corbes de Kaplan-Meier de la supervivència lliure de malaltia pels diferents rangs en el número d'alteracions a nivell de la CGH.

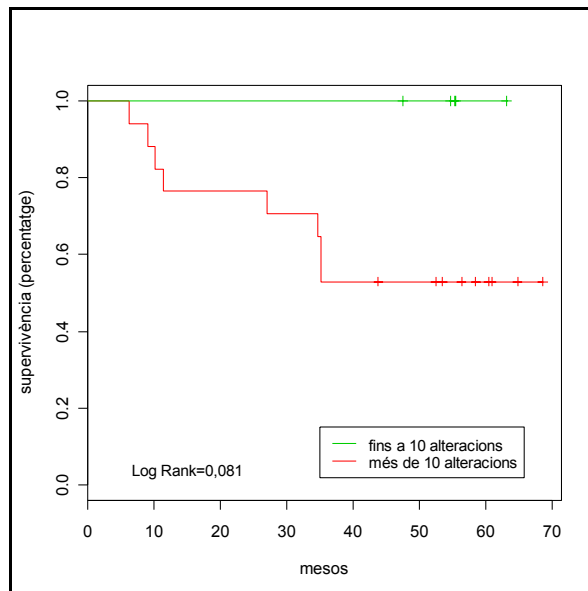
Sembla doncs evident que les classificacions fetes fins ara en base a les alteracions cromosòmiques no milloren considerablement la nostra capacitat de predicció pronòstica. Però a continuació ens vàrem plantejar que si els grups

realment representen diferents entitats o vies de progressió, els factors pronòstics poden ser diferents a cada via.

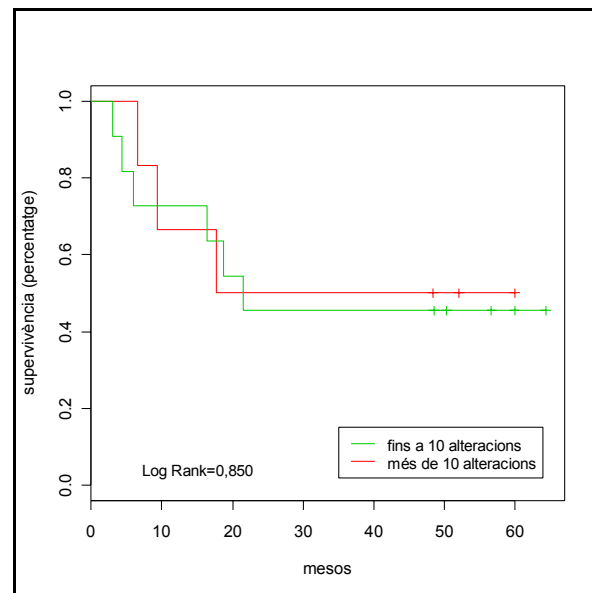
Per això vam intentar esbrinar què passava dins els grups ICML i ICR. Si agafem el llindar de les 10 alteracions, observem que pel grup ICML els casos amb més alteracions tenen una supervivència més baixa (de 17 casos, 8 presenten recidiva). I dels 5 que tenen menys de 10 alteracions no n'hi ha cap que presenti recidiva (Figura 35). Contràriament en el grup *ICR*, no hi ha diferències destacables quan separem els casos en funció del número d'alteracions (Figura 36).

Aquest resultat ens indica que dins del grup ICML l'acumulació d'alteracions cromosòmiques aniria paral·lela a l'increment de l'agressivitat, i en aquest sentit hipotèticament seria la inestabilitat cromosòmica el motor de la progressió en aquest subgrup. Pel contrari, en el grup de tumors ICR el comportament biològic vindria més condicionat pel tipus d'alteracions.

En resum podem dir que la classificació dels tumors en grups diferents segons la via de progressió és absolutament necessària a l'hora d'identificar quins són els marcadors més útils per a cada cas.



**Figura 35.** Corbes de Kaplan-Meier de la supervivència lliure de malaltia pel conjunt de casos del grup ICML en funció del número d'alteracions.



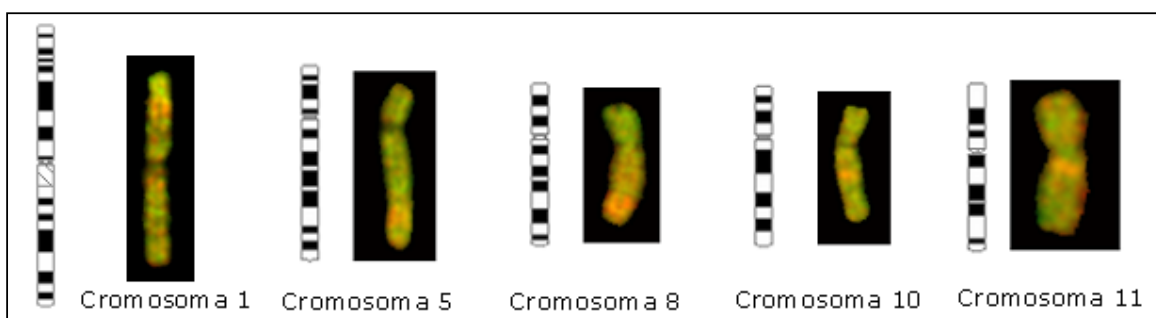
**Figura 36.** Corbes de Kaplan-Meier de la supervivència lliure de malaltia pel conjunt de casos del grup ICR en funció del número d'alteracions.

### Amplificacions

Com ja hem explicat abans (vegeu l'apartat de la *Introducció*), les amplificacions són regions on hi ha un increment notable del nombre de còpies de DNA. La CGH no és una tècnica amb prou resolució per poder quantificar el guany net del tros amplificat, però és interessant perquè ens el situa cromosòmicament. A la sèrie dels 50 casos vam trobar sis regions diferents amplifiades: 20q33, 11q13, 8q24, 5q31-q33, 10q22, i 1p36 (vegeu la Taula 32 i la Figura 37).

**Taula 32.** Amplificacions de DNA de la sèrie de 50 casos.

cromosoma	número de casos	possibles gens afectats
20q	10	E2F (factor de transcripció), MMP9
11q13	8	CCND1 (ciclina D1)
8q24	5	MYC
5q31-q33	1	FGF1 (factor de creixement de fibroblast)
10q22	1	FRAT1 ( <i>Frequently rearranged in T-cell lymphoma</i> )
1p36	1	TP73 (Tumor protein p73)



**Figura 37.** Exemple de cinc tipus d'amplificacions trobades a la sèrie de 50 casos.

### Associacions amb supervivència de les alteracions cromosòmiques

Hi ha molts estudis centrats en la recerca d'associacions entre determinades alteracions cromosòmiques i característiques clinicopatològiques o moleculars. Un dels objectius més buscats és el de trobar marcadors moleculars de pronòstic, per la seva utilitat a l'hora de tractar amb els pacients i per a poder estudiar possibles dianes terapèutiques.

Per això vam estudiar la correlació entre cada una de les alteracions i la supervivència dels pacients. Fixant-nos en cada un dels braços dels cromosomes vam trobar que hi ha dos cromosomes que tenen una tendència significativa. Així, a la Taula 33 mostrem alguns dels resultats obtinguts.

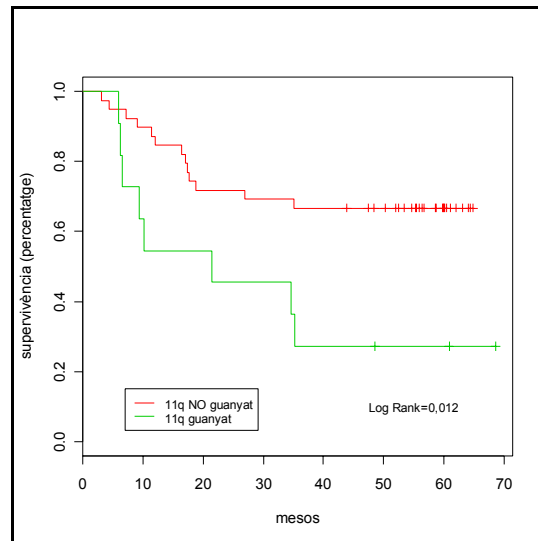
**Taula 33.** Correlació entre l'alteració de cada un dels cromosomes i la supervivència lliure de malaltia dels 50 casos mirat amb el test de Fisher. Només es donen els resultats que difereixen d'un p-valor=1.

guany	p valor	pèrdua	p valor
		- 4p	0,543
+1p	0,171	- 4q	0,759
+2p	0,171	- 5q	0,638
+5p	0,565	- 6q	0,254
+6q	0,254	- 8p	0,255
+8p	0,254	- 8q	0,420
+8q	0,779	- 9p	0,420
+9q	0,565	- 10p	0,420
+10q	0,420	- 10q	0,171
+11p	0,565	- 11p	0,420
<b>+11q</b>	<b>0,036</b>	- 11q	0,420
+12q	0,223	<b>- 12p</b>	<b>0,026</b>
+19	0,328	- 14	0,166
+20p	0,346	- 15	0,140
+Xp	0,716	- 17p	0,578
		- 20p	0,380

Només hi ha dos cromosomes on trobem un resultat significatiu: el guany del braç 11q, i la pèrdua del braç 12p. Si ho mirem amb detall trobem que, pel què fa al cromosoma 11, vuit dels onze casos que tenen el braç llarg guanyat presenten recidiva a causa de la malaltia (Taula 34 i Figura 38), per tant sembla un factor de mal pronòstic. Per això suposem que hi deu haver algun oncogèn que està actuant com a factor de selecció positiu en aquelles cèl·lules. El més conegut en aquesta regió és el gen CCND1 (Brison, 1993), però ja es discutirà amb més detall a l'apartat de la *Discussió*.

**Taula 34.** Relació entre la supervivència lliure de malaltia i el guany del braç 11q.

	sense recidiva	amb recidiva
11q no guanyat	26	13
11q guanyat	3	8

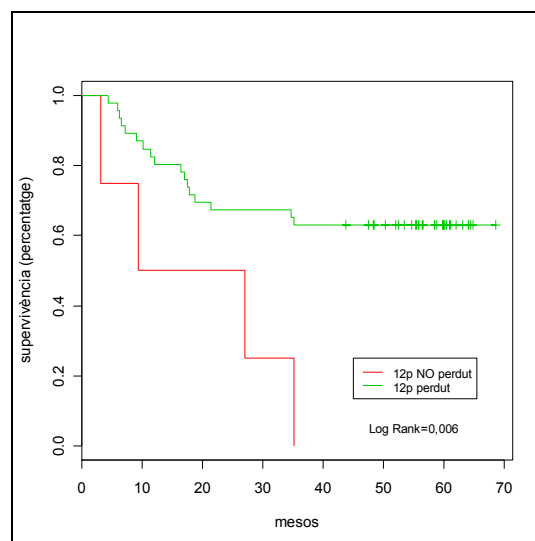


**Figura 38.** Corbes de Kaplan-Meier de la supervivència lliure de malaltia en funció dels guanys al braç 11q.

Pel què fa al cromosoma 12, trobem que els quatre casos que tenen el braç 12p perdut presenten recidiva (Taula 35 i Figura 39). Per tant ens indicaria la presència d'algun gen supressor de tumors.

**Taula 35.** Relació entre la supervivència lliure de malaltia i la pèrdua del braç 12p (no s'inclouen els casos que guanyen el 12p).

	sense recidiva	amb recidiva
12p no perdut	29	17
12p perdut	0	4



**Figura 39.** Corbes de Kaplan-Meier de la supervivència lliure de malaltia en funció de la pèrdua del braç 12p.

### Alteracions numèriques i estructurals

Amb la tècnica de la CGH només podem detectar aquelles alteracions cromosòmiques estructurals que no siguin balancejades, per tant estem perdent la informació de les alteracions estructurals on no hi ha ni guany ni pèrdua de material genètic. Tot i això, volem comprovar si aquests dos tipus d'alteracions succeeixen de forma independent. Mitjançant el test de Pearson observem que hi ha una correlació positiva ( $r=0,489$  i  $p>0,0001$ ) entre les alteracions numèriques i les estructurals. En conseqüència no sembla que de forma general, aquests dos tipus d'alteracions es donin de forma independent dins de la nostra sèrie.

### Mutacions a TP53 i LOHs a 17p

Una altra de les associacions a considerar és la pèrdua del braç 17p i la mutació del gen TP53, que es localitza a la regió 17p13.1. Com que es tracta d'un gen supressor de tumors, és d'esperar que en alguns casos una còpia s'hagi perdut mitjançant l'aberració cromosòmica i que l'altra estigui mutada (mecanisme del *double impacte* dels gens supressors de tumors) (Knudson, 1993). Així, dels vint-i-quatre casos que tenen el braç 17p perdut, dinou presenten també la mutació del gen TP53 (Taula 36). Però el valor de significació no arriba a 0,05 perquè hi ha molts casos que no han perdut el braç 17p i també presenten la mutació.

**Taula 36.** Relació entre la mutació de TP53 i la pèrdua del braç 17p (no s'inclouen els casos que guanyen el 17p).

	TP53 no mutat	TP53 mutat
17p no perdut	10	16
-17p	5	19
Fisher	0,224	

### Mutacions a TP53 i inestabilitat cromosòmica

Si comparem la mutació del gen TP53 i el número d'alteracions a nivell de CGH dels 50 casos (sense contar les amplificacions) trobem que la gran majoria de tumors amb mutacions a p53 tenen alteracions cromosòmiques (Taula 37).

**Taula 37.** TP53 i alteracions cromosòmiques (no s'han inclòs les amplificacions).

	TP53 no mutat	TP53 mutat
sense alteracions	7	2
amb alteracions	8	33
Fisher	0,002	

## TP53 i amplificacions

Alhora, també trobem correlació entre la presència d'alguna amplificació i la mutació del gen TP53. En conjunt, aquestes dades demostren de forma clara que la presència de mutacions en el gen TP53 va associada a una major inestabilitat cromosòmica.

**Taula 38.** TP53 i les amplificacions.

	TP53 no mutat	TP53 mutat
sense amplificació	13	17
amb amplificació	2	18
Fisher	0,026	

## 5. 2. Amplificació del gen de la ciclina D1 (CCND1)

Mitjançant la tècnica de la PCR a temps real o PCR quantitativa (QPCR), vam quantificar l'amplificació de la regió 11q13, on se situa el gen de la ciclina D1. De les regions amplificades vam escollir aquesta per ser una de les que afectava a més casos (onze tumors amb guany o amplificació) i perquè ja s'havia descrit anteriorment que un dels gens responsable de l'amplificació a aquesta regió era la ciclina D1 (Brison, 1993).

Vam utilitzar el sistema *Taqman* (vegeu l'apartat de *Materials i mètodes*). Com a control de l'amplificació vam escollir el gen de la  $\beta$ 2-MICROGLOBULINA (B2M), que es localitza a la regió 15q21-q22. És un gen dels anomenats *housekeeping* que sovint s'utilitza com a control intern en experiments d'expressió. En el nostre cas però, estem quantificant el DNA i el que ens interessa és que aquella regió del cromosoma no estigui afectada. De la sèrie de 50 casos tenim cinc casos que tenen el braç 5q perdut, i tots ells presenten una amplificació a 11q13. Per tant, per aquests cinc casos s'ha de tenir en compte el possible efecte de la pèrdua de la B2M. Les reaccions es van per triplicat.

Per poder establir els llindars de canvi per aquests dos gens vam fer un experiment de reproductibilitat amb 7 teixits tumorals que no presentaven canvis en els cromosomes 11 i 15 (Taula 39). Vam calcular la mitjana i la desviació estàndard (SD) dels 10 casos, i vam establir com a llindar el valor de  $1 \pm 5 \times SD$ , és a dir cinc vegades la desviació estàndard. A la Taula 39 estan resumits els càlculs.

Per tant, i en funció d'aquests resultats, decidim establir com a llindar de canvi  $5 \times 0,08 = 0,4$ . És a dir que parlarem d'amplificació a partir d'un valor de 1,4 i de pèrdua a partir de 0,6.

**Taula 39.** Resultat de la QPCR del gen de la ciclina D1 en 7 teixits controls que no presenten canvis als cromosomes 11 i 15.

cas	Ct CIC D1	Ct B2M	$\delta Ct$	$\delta\delta Ct$	$2^{-\delta\delta Ct}$
11T	25,24	25,66	-0,42	0	1
16 T	25,76	26,37	-0,61	-0,19	0,88
99 T	24,75	25,45	-0,71	-0,29	0,82
102 T	25,79	26,35	-0,55	-0,13	0,91
127 T	25,26	25,79	-0,53	-0,11	0,92
143 T	23,95	24,31	-0,36	0,06	1,05
144 T	24,78	25,28	-0,50	-0,08	0,95
147 T	25,30	26,02	-0,72	-0,30	0,81
				<b>mitjana</b>	<b>0,91</b>
				<b>SD</b>	<b>0,08</b>

A la Taula 40 hi ha resumits els resultats de la QPCR. Observem tres casos que presenten una amplificació al locus de la CCND1, la resta no arriben al llindar establert a 1,4 vegades. El cas 11 es va fer servir de referència de no amplificació, i es va incloure un control negatiu (el cas 127). Notis que en el cas referència el número de còpies és 1, perquè s'ha utilitzat per restar-li al mateix cas l'increment de Ct ( $\delta Ct$ ). També hi ha dos casos dels que tenen perdut el cromosoma 15 que presenten un número de còpies del gen CCND1 superior a 1,4, però no podem confirmar el guany. Perquè pot ser que el gen de la CCND1 no estigui guanyat i al perdre's el gen de la B2M, estem sobrevalorant el guany. Si tot i així és negatiu, podem assegurar que el gen de la CCND1 no és el responsable de l'amplificació.

També es va incloure el cas 151T, que presenta una pèrdua al braç 11q. Veiem però, que amb la QPCR no es detecta el canvi considerant com a llindar de pèrdua el valor 0,6.

**Taula 40.** Resultats de la CGH i QPCR del gen de la ciclina D1.

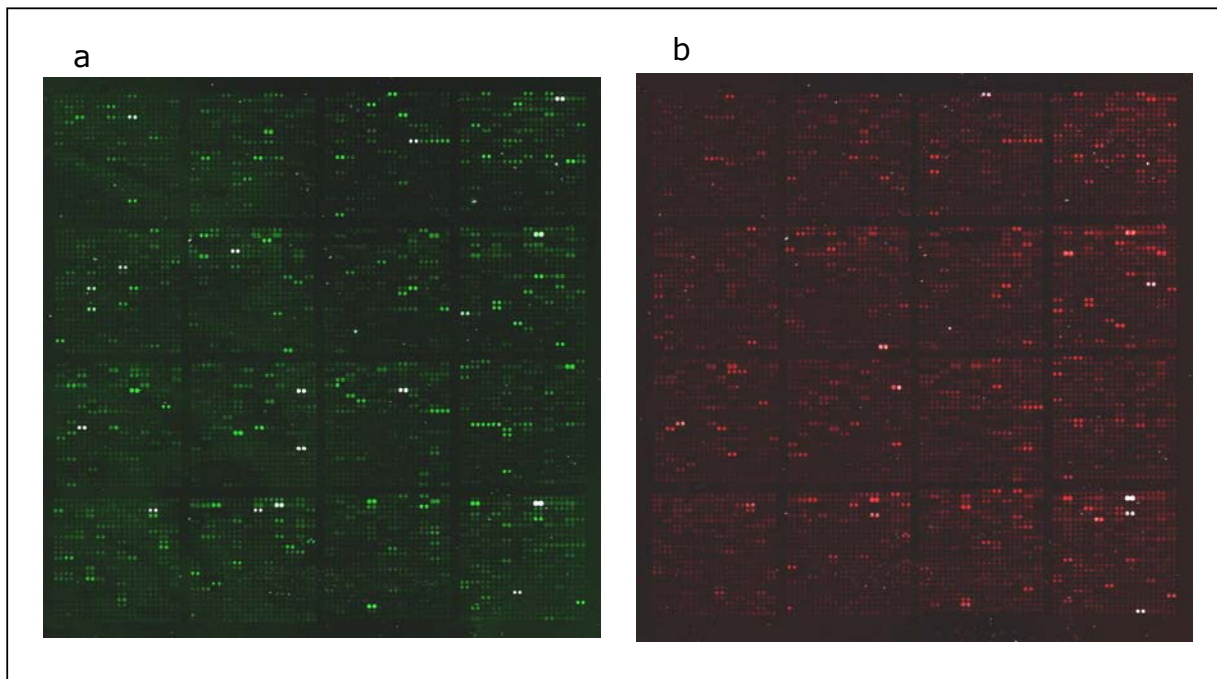
cas	guany per CGH	pèrdua per CGH	$2^{-\delta\delta Ct}$	log2 del canvi
11 T	-	-	1	0
23 T	amp11q13	-	<b>13,6*</b>	<b>3,8</b>
72 T	amp11q13	-	1,2	0,2
108 T	amp11q13	-	0,8	-0,2
109 T	11q	-	<b>1,4</b>	<b>0,5</b>
113 T	11	-15q	1,1**	0,07
127 T	-	-	1,1	0,1
146 T	11	-	1,2	0,2
151 T	-	-11q	0,8	-0,3
186 T	amp11q13	-	<b>1,4</b>	<b>0,4</b>
187 T	amp11q13	-15	0,9**	-0,2
223 T	amp11q13	-15	1,4**	0,4
271 T	amp11q13	-15	1,5**	0,6
282 T	amp11q13	15q15-q23	1,2**	0,2

\* Valors llindars per guany i pèrdua: 1,4 i 0,6 respectivament. \*\* Casos on s'ha perdut el cromosoma 15q.

### 5. 3. Expressió gènica diferencial

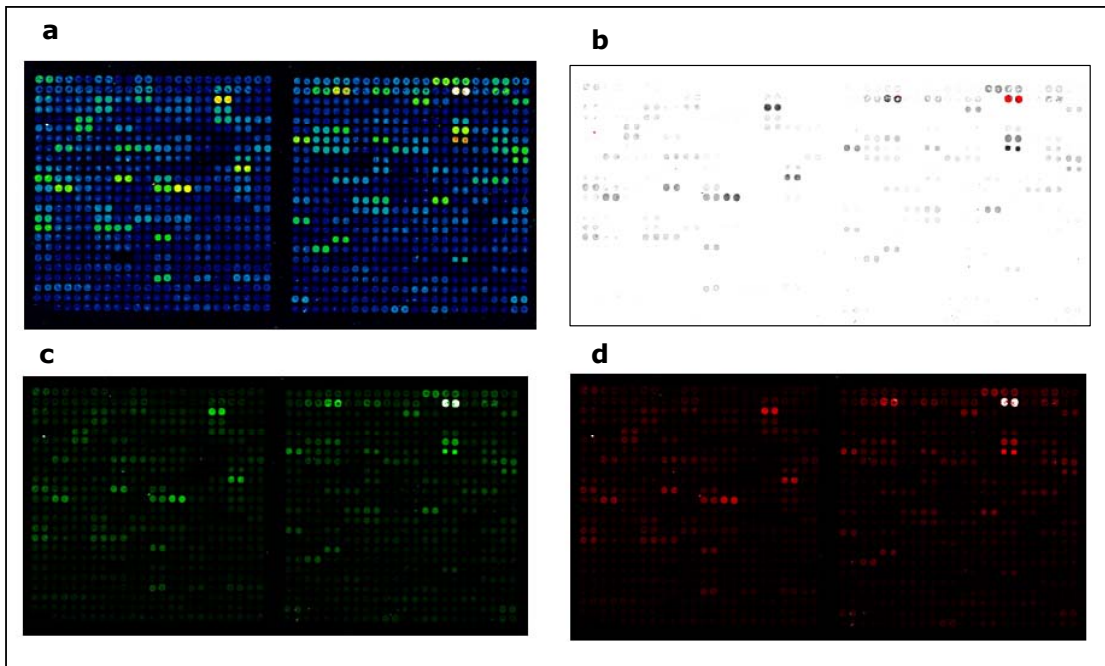
#### Obtenció de les imatges

Per cada hibridació s'obtenen dues imatges que corresponen a cada un dels canals associats als fluorocroms utilitzats: Cy3 a una longitud d'ona de 550nm, i Cy5 a una longitud d'ona de 650 nm (Figura 40).



**Figura 40.** (a) Imatge obtinguda a l'escanear a una longitud d'ona de 550 nm. La paleta de colors és arbitrària: s'ha escollit un gradient que va del negre (poc senyal) a verds més intensos (més senyal). Els punts blancs indiquen els píxels saturats. El material hibridat és la RAP-PCR del POOLN. (b) Imatge obtinguda a l'escanear a una longitud d'ona de 650 nm. La paleta de colors és arbitrària: s'ha escollit un gradient que va del negre (poc senyal) a vermells més intensos (més senyal). Els punts blancs indiquen els píxels saturats. El material hibridat és la RAP-PCR del tumor número 67.

Per facilitar l'avaluació ràpida de la qualitat de la hibridació sovint es visualitza la imatge amb diferents paletes de colors que transformen els valors numèrics d'intensitat a colors arbitraris que nosaltres escollim. Així, a part del verd i del vermell també es pot utilitzar una paleta amb blanc i negre (útil per comparar intensitats globals entre experiments), o una amb tots els colors de l'arc de sant martí (útil per determinar la relació senyal/soroll) (Figura 41).



**Figura 41.** Detall de la mateixa àrea d'una hibridació amb quatre paletes de colors diferents. **(a)** L'escala de colors segueix la següent pauta de menys a més intensitat: negre < blau < verd < groc < vermell < blanc (saturat). **(b)** Imatge en blanc i negre, on el color blanc correspon a intensitats baixes, i el negre a intensitats altes. El color vermell indica els píxels saturats. **(c)** Paleta de color verd. **(d)** Paleta de color vermell.

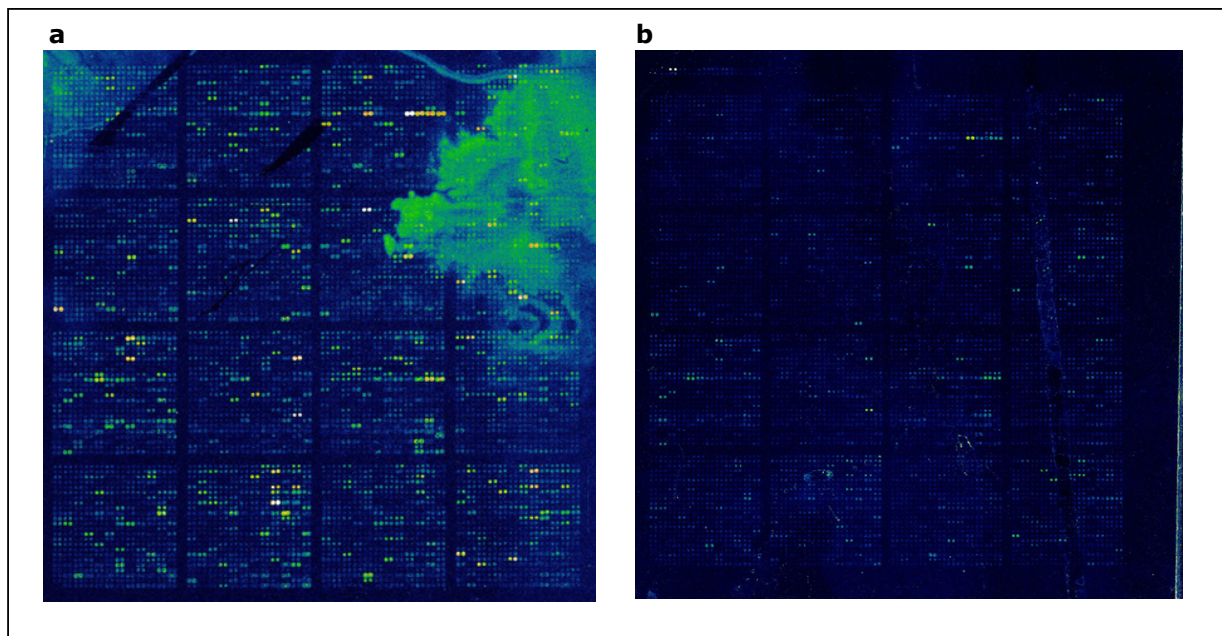
#### Selecció dels experiments: anàlisi de la qualitat

Per analitzar la qualitat de les hibridacions fèiem una inspecció visual dels dos canals de la hibridació, el Cy3 i el Cy5. Els paràmetres que es tenien en compte per considerar una bona hibridació són els següents:

- el senyal absolut del conjunt dels punts estigui dins del rang de lectura, és a dir, que es vegin prou punts hibridats específicament.
- la relació del senyal respecte el fons sigui elevada.
- la hibridació sigui homogènia en tota la seva àrea, és a dir, que no hi hagi àrees amb més senyal que d'altres.

Si bé es poden utilitzar criteris objectius a l'hora de determinar la qualitat de la hibridació basant-nos en tots aquests criteris, la inspecció visual de la imatge utilitzant l'escala de colors de l'arc de sant martí descrita a la Figura 41a, és adient per tal que l'investigador pugui avaluar la qualitat global de l'experiment d'una forma ràpida i relativament objectiva. Evidentment cal un cert entrenament per a la interpretació de les imatges, però en base a la nostra experiència, l'ús d'una paleta tipus arc de Sant Martí és un dels mètodes que millors resultats dona quan comparem les interpretacions de diferents investigadors sobre el mateix experiment. És a dir, el nombre de discordances és menor.

La major part de les imatges de qualitat deficient es van obtenir al canal del fluorocrom Cy3. El problema que es va trobar amb més freqüència va ser l'aparició d'un núvol que emmascarava el fons de la imatge i no es podien veure els punts (Figura 42a), com si els rentats no haguessin retirat de forma eficient la sonda sobrera. En d'altres ocasions el senyal era molt fluix (Figura 42b), i semblaria que el marcatge podia haver fallat o el fluorocrom havia perdut la seva fluorescència. Si bé també s'aconseguien imatges de qualitat equivalent a les del Cy5, es van haver de descartar una proporció significativa de les imatges del Cy3. En principi hem atribuït les dificultats per obtenir bones imatges amb Cy3 a la inestabilitat d'aquest fluorocrom en comparació al Cy5. En estudis previs hem observat que quan s'obtenen bones hibridacions el fluorocrom no té efecte (o és menyspreable) en el resultat. Donat que el principal objectiu del nostre treball era comparar el patró d'expressió dels tumors entre sí, es va optar per utilitzar el Cy5 pel marcatge del tumor (a fi de garantir el màxim de dades d'aquest teixit), i el Cy3 pel marcatge dels teixits normals i POOLN controls. D'aquesta manera vam minimitzar la pèrdua d'informació sense haver d'incrementar significativament el nombre d'experiments.



**Figura 42.** Imatge del canal Cy3 de dues hibridacions de mostres de teixit normal. **(a)** A l'àrea de la dreta es pot detectar una zona bruta on la sonda no s'ha rentat eficaçment. **(b)** Veiem com quasi no hi ha senyal específica.

En el disseny inicial es van fer les següents hibridacions:

- 50 hibridacions N vs T
- 50 hibridacions POOLN vs T

Després de l'anàlisi visual, tal i com hem explicat, ens vam quedar amb les següents dades:

- canal Cy5 dels 50 teixits tumorals
- canal Cy3 de 13 hibridacions de la mescla de teixits normals (POOLN)

#### Filtratge dels punts vàlids

Com ja s'ha explicat a l'apartat de *Materials i mètodes* es van establir uns llistats *deterministes* a l'hora de descartar els punts no analitzables. Així vam imposar un llistat 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 podem analitzar entre un 90 i un 99,5% de tot els punts. I representa que ens quedem de mitjana amb 4400 gens dels 4608 (un 95,5%).

#### Protocol d'anàlisi

Vam treballar amb les dades del canal Cy5 corresponent al teixit tumoral independentment del canal Cy3 del teixit normal aparellat. Així tenim els valors d'intensitat absoluts per als 50 teixits tumorals, i dels 13 POOLN. A continuació és necessari un procés de normalització de les dades.

### **Normalització i transformació de les dades**

Una vegada feta la selecció dels punts que són analitzables cal normalitzar les dades. El fet d'hibridar cada mostra en un xip diferent, juntament amb les variacions que es produeixen en l'eficiència del marcatge fa que el senyal de fluorescència obtingut en cada experiment no es pugui comparar directament (vegeu l'apartat de *Materials i mètodes*). Cal doncs, un pas previ per transformar les dades crues de manera que siguin comparables, i així estudiar les intensitats de canals d'hibridacions diferents. Aquest procés s'anomena normalització i existeixen diferents maneres per fer-ho.

En el nostre cas, a l'hora de normalitzar vam seleccionar les 13 hibridacions del *pool* de teixits normals (POOLN) que oferien una més bona qualitat, i vam calcular la mitjana de cada gen. Això ens donà una determinada distribució que es considera com a referència per normalitzar la resta d'hibridacions (Figura 23, de l'apartat de *Materials i mètodes*).

Les intensitats es van normalitzar per quantils respecte a aquesta hibridació referència. De manera que les intensitats dels 4608 gens dels 50 casos seguissin

una distribució igual que la distribució de referència. Així totes les hibridacions tenen un rang de variació igual i es poden comparar entre elles.

Si bé les dades normalitzades ja es poden utilitzar directament per fer totes les anàlisis bioinformàtiques, en el nostre cas hem transformat els valors d'intensitat a valors relatius, utilitzant les dades de referència com a base. Per això calculem la raó entre el cas i el POOLN de les intensitats de cada gen segons la fórmula:

$$\text{raóT} = \log_2 \left( \frac{(\text{INT GEN}_i \text{ TUMOR}_n)}{(\text{INT GEN}_i \text{ POOLN})} \right)$$

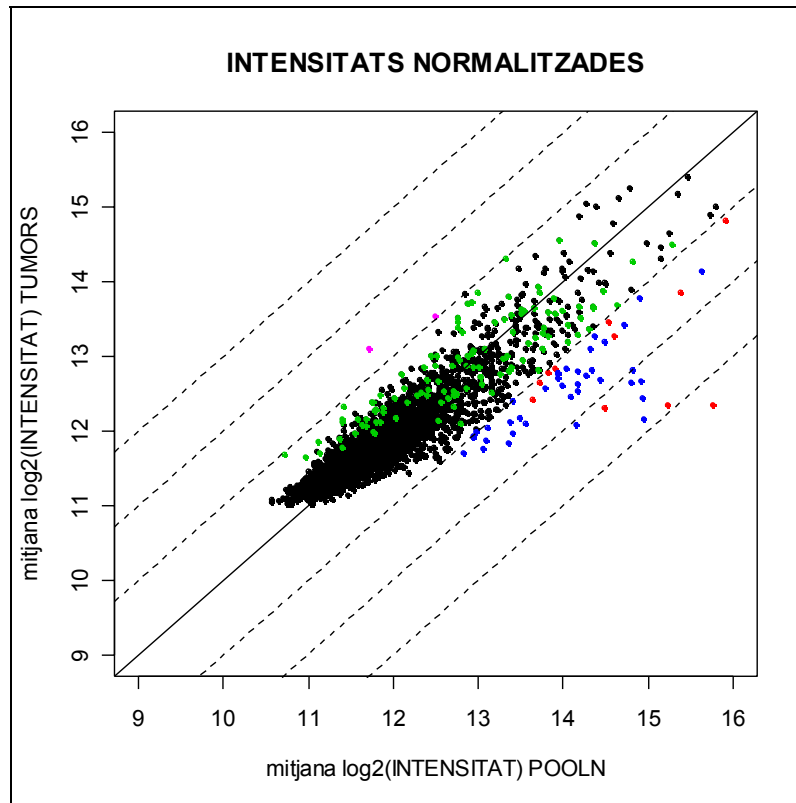
Així, estem fent un càlcul de la diferència del tumor respecte un *pool* de teixits normals. A l'aplicar logaritmes ens disminueix la variància i a més existeix simetria dels valors de sobreexpressió respecte als de infraexpressió, degut a la propietat:  $\log x = -\log 1/x$

Així, els valors per damunt de zero corresponen a sobreexpressions, i valors per sota de zero són infraexpressions, del teixit tumoral respecte un *pool* de normals. I més concretament, com que calculem el logaritme en base 2, diem que quan la raóT és igual a 1 hi ha un canvi del doble, i quan la raóT és igual a -1 hi ha un canvi de la meitat.

Hem de tenir en compte però que aquestes variacions no seran necessàriament proporcionals a la quantitat inicial d'RNA, perquè en el pas de la normalització hem igualat tots els rangs de variació de les hibridacions. El sentit del canvi és més important que no pas la magnitud d'aquest.

Una manera de representar les dades resumides és mirant la intensitat mitjana dels 50 casos per cada gen en relació a la intensitat del POOLN (Figura 43). El codi de colors seguit és el mateix que a la Figura 44. Com es pot veure, la majoria de gens que tenen de mitjana més diferència d'expressió tendeixen a tenir una intensitat més elevada (punts blaus, vermells i liles). Observem que hi ha una desviació cap a la infraexpressió del tumor respecte el POOLN.

Donat el gran nombre de variables considerades en aquest tipus d'estudi, abans d'entrar en l'anàlisi estadística (associacions i correlacions entre variables, classificacions, anàlisi de supervivència, etc.), cal explorar els resultats de forma general per a decidir l'estratègia més adient per seleccionar les variables a estudiar. La reducció del nombre de variables ens permetrà guanyar potència estadística i simplificar la interpretació de les possibles associacions que trobem.

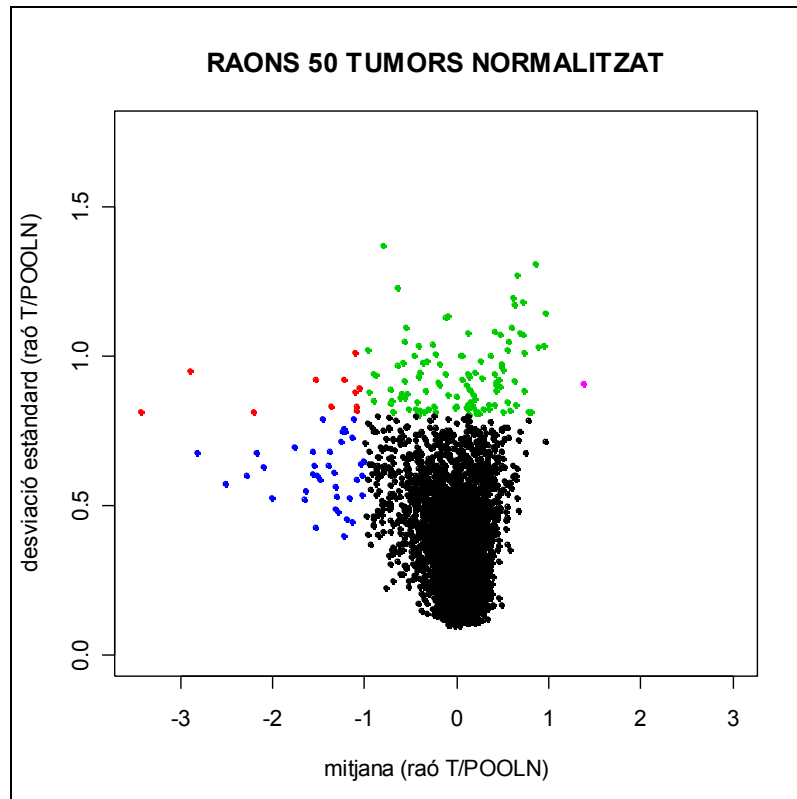


**Figura 43.** Intensitat mitjana dels 4608 gens de les 13 hibridacions del POOLN (eix de les abscisses) i de les 50 hibridacions dels teixits tumorals (eix de les ordenades). El codi de colors s'explica a la Figura 44.

En un primer pas representem la mitjana de les raons dels 4608 gens en els 50 tumors respecte la desviació estàndard (SD) (Figura 44).

D'aquesta manera podem analitzar simultàniament el nivell de representació mitjana de cada gen en els tumors respecte al POOLN (teixit normal), i la dispersió en el grau de representativitat de cada gen.

La mitjana de la raóT ens indica la tendència cap a la infra- o sobreexpressió, i la desviació estàndard la variabilitat d'expressió dins dels tumors. Quan més gran sigui aquesta desviació estàndard esperem que aquests gens siguin més informatius a l'hora de fer classificacions, o buscar marcadors associats amb característiques clinicopatològiques, per exemple.



**Figura 44.** A l'eix de les abscisses es representa la mitjana de la raó dels 4608 gens pels 50 casos (les dades estan expressades en logaritme en base 2). A l'eix de les ordenades es representa la SD dels 4608 gens pels 50 casos. Els gens representats de color blau són aquells que tenen una SD  $\leq 0,8$  i una raó mitjana  $< -1$  (N=38), els de color vermell són aquells que tenen una SD  $> 0,8$  i una raó mitjana  $> 1$  (N=11), els de color verd són aquells que tenen una SD  $> 0,8$  i una raó mitjana entre  $-1$  i  $1$  (N=115), els de color lila són aquells que tenen una SD  $> 0,8$  i una raó mitjana  $> 1$  (N=2), i finalment la resta representats de color negre són aquells que varien poc (SD  $< 0,8$ ) i tenen una raó mitjana entre  $-1$  i  $1$  (N=4442).

A la Figura 44 veiem com els gens representats de color blau (N=38), tenen una variabilitat baixa o moderada (SD  $\leq 0,8$ ), però tenen una mitjana de la raó  $< -1$ . Aquests gens es podria considerar que de forma recurrent tenen tendència a estar infraexpressats en els tumors respecte la referència normal. Els gens de color vermell (N= 11) representen aquells que tenen una variabilitat elevada (SD  $> 0,8$ ) i alhora es tendeixen a infraexpressar.

D'altra banda, veiem una altra sèrie de gens representats de color verd (N=115), que tenen una elevada variabilitat (SD  $> 0,8$ ), però la gran majoria no presenten una clara tendència cap a la infra- o sobreexpressió, el que suggereix un ampli espectre de canvis.

Sorprenentment, només hi ha 2 gens (pintats de color lila) que tinguin una elevada variabilitat (SD  $> 0,8$ ) i alhora es tendixin a sobreexpressar (mitjana de la raó  $> 1$ ). És evident l'asimetria del núvol de punts que demostra una tendència a la infraexpressió dels 4608 gens.

Finalment, hi ha una gran quantitat de gens (N=4442) pintats de color negre que corresponen a aquells que no presenten una clara tendència cap a la infra- o sobreexpressió i alhora presenten una desviació estàndard més baixa de 0,8.

### **Selecció dels gens**

Un dels objectius de la present tesi és poder classificar els 50 casos en diferents grups en funció del seu patró d'expressió. És fàcil pensar que no tots els gens contribuiran a definir els grups, i tampoc resulta pràctic treballar amb els 4608 gens. Per a fer la selecció dels gens més interessants vam escollir un llinar en funció de la desviació estàndard (SD). Aquest valor ens indica la variabilitat d'aquell gen en les 50 mostres, i per tant, a l'hora de fer grups aquests gens seran els que amb més probabilitat contribuiran a marcar diferències entre els tumors. Segons el valor de SD que determinem el nombre de gens que queda per sobre és diferent. Així:

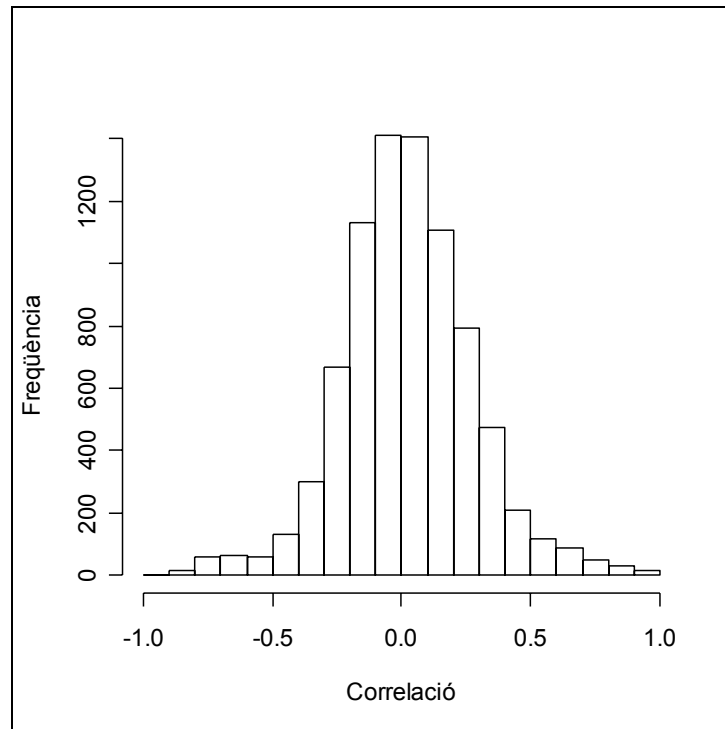
- hi ha 32 gens que tenen una SD més gran que 1.
- hi ha 128 gens que tenen una SD més gran que 0,8.
- hi ha 889 gens que tenen una SD més gran que 0,5.

De forma arbitrària vam escollir el valor de SD de 0,8, perquè d'aquesta manera ens quedem amb un nombre de gens manejable i a priori candidats a ser útils en els sistemes de classificació. A la Figura 44 corresponen als gens pintats de color vermell (N= 11), verd (N= 115), i lila (N=2).

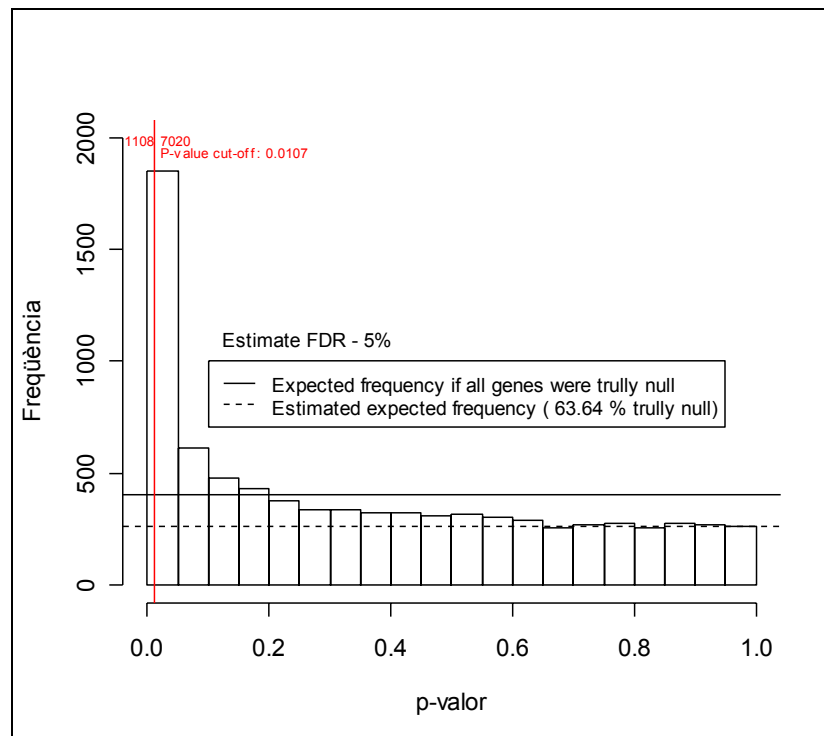
### **Descriptiva dels 128 gens**

Tot i que aquests 128 gens són els que a priori ens poden donar més informació, cal que estudiem la relació que hi ha entre ells. Perquè pot ser que alguns d'ells estiguin correlacionats i ens donin informació redundant. Per això els comparem 2 a 2, fent un total de 8128 parelles de comparacions possibles. A la Figura 45 es representa un histograma de freqüències de les correlacions entre els gens.

Veiem com una part de les parelles de gens no estan correlacionades, i se centren al zero. Una petita proporció tenen correlacions negatives i positives. Per a la detecció de les parelles de gens amb correlació significativa podem emprar una anàlisi basada en el *False Discovery Rate* (Storey i Tibshirani, 2003). A la Figura 46 mostrem l'histograma dels p-valors de les correlacions entre tots els gens.



**Figura 45.** Histograma de freqüències de les correlacions entre les 8128 parelles de gens.



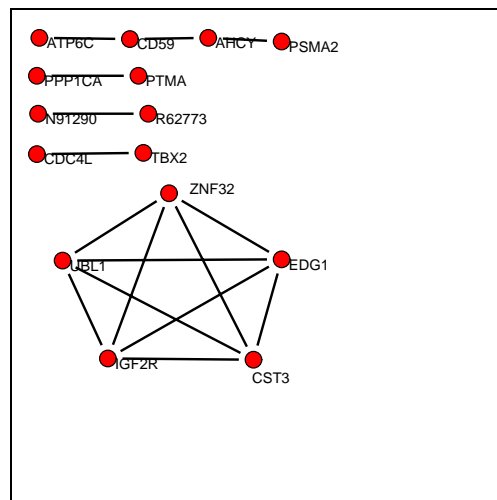
**Figura 46.** Histograma de freqüències dels 8128 p-valors dels 128 gens agafats 2 a 2. La línia horitzontal contínua correspon a la distribució esperada si la correlació entre tots ells fos nul·la. La línia discontinua és a l'alçada de la distribució estimada.

Si no hi hagués cap correlació (l'expressió de tots els gens és independent entre ells), hauríem d'obtenir una distribució uniforme de tots els p-valors (línia horitzontal contínua de la Figura 46). Observem que realment hi ha una mescla de dues distribucions: la de les parelles de gens no correlacionades (uniforme) i la de les parelles de gens correlacionades (amb p-valors propers a zero).

A partir d'aquesta distribució i utilitzant la metodologia d'Storey i Tibshirani (2003), obtenim una sèrie de paràmetres per estimar tant les freqüències de correlacions nul·les com les significatives. En el nostre cas s'estima que un 63,6% de les parelles de gens no estan correlacionades. Així, fixant un llindar a partir del p-valor 0,0107 obtenim 1108 parelles de gens amb correlació significativa (Figura 46). Dels que cal esperar que un 5% siguin falsos positius, és a dir, 55 parelles de gens.

Mirant amb més detall aquelles parelles de gens correlacionades, podem buscar si els gens que es comporten de la mateixa manera tenen alguna relació funcional (ja sigui perquè formen part d'una mateixa via de transmissió de senyal, d'una via metabòlica, de diferents vies que tinguin funcions complementàries, etc.).

Una manera de mostrar les correlacions entre els gens és mitjançant representacions topològiques de xarxes formades per nodes (gens) i arestes (entre els gens correlacionats). Per començar agafem aquelles parelles de gens que tenen una correlació més gran de 0,9 en valor absolut (N=16) i que mostrem a la Figura 47.

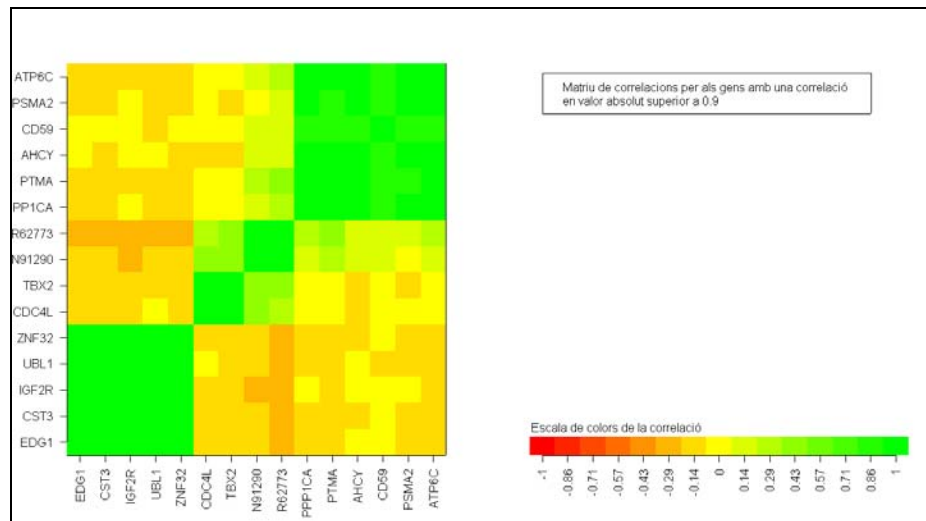


**Figura 47. Graph.** Representació de les correlacions superiors a 0,9.

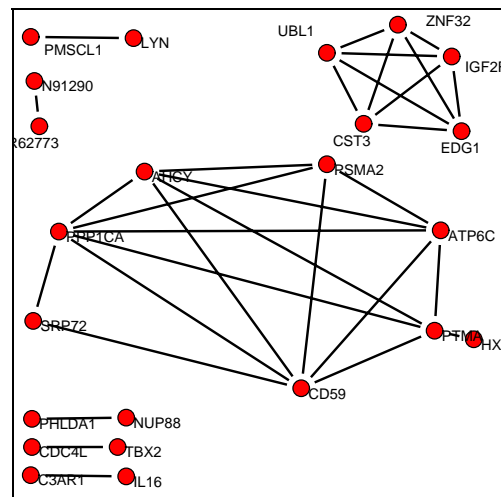
És de destacar que totes les correlacions d'aquests grups de gens tenen sentit positiu (els valors exactes es troben a l'annex I, a la Taula 49). Hi ha cinc gens que formen una xarxa de correlacions positives de tots entre ells. A la Taula 41 es detallen les funcions de cada un d'ells. Una altra manera de representar els

resultats és mitjançant una taula de colors (*heatmap*) on es dibuixen els diferents valors de correlacions seguint l'escala de colors indicada a la llegenda (Figura 48). En aquesta gràfica observem quatre estructures, i no cinc com en el *graph*, ja que el grup de gens de l'ATP6V1C1 (ATP6C) té una elevada correlació amb els gens PPP1CA i PTMA.

A continuació, si agafem un llindar més baix estudiant així les correlacions majors de 0,85 en valor absolut trobem que apareixen noves xarxes de gens, alguns dels quals presenten correlacions negatives. Ara tenim 32 parelles de gens correlacionats, 16 més que abans. A la Figura 49 es poden veure les xarxes de gens.

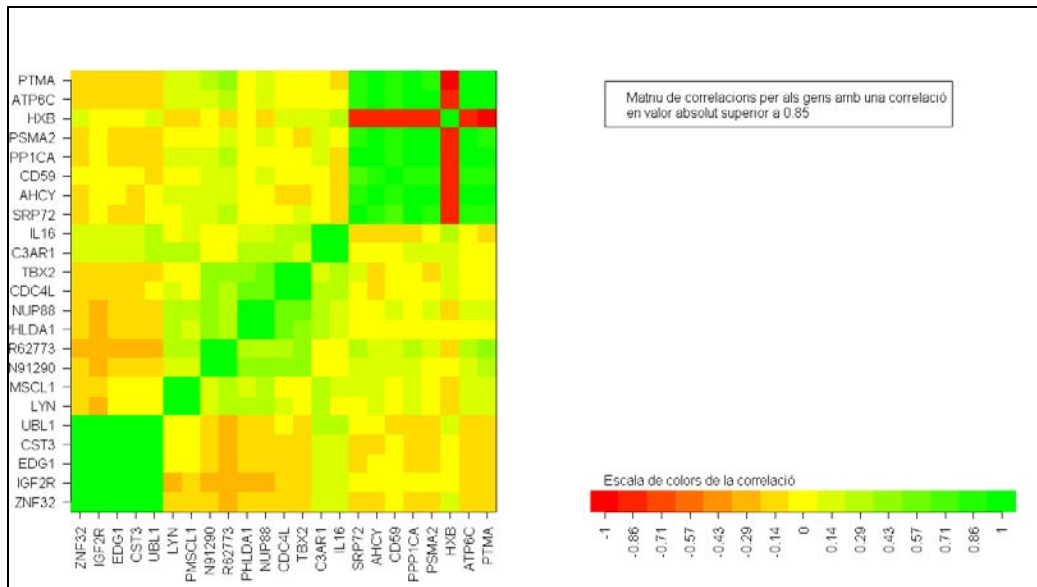


**Figura 48.** *Heat map* de les correlacions superiors a 0,9. Els valors propers a 1, que corresponen a una correlació positiva, apareixen de color verd, i els valors propers a  $-1$ , que corresponen a una correlació negativa, apareixen de color vermell. És de destacar que totes les correlacions més fortes són positives.



**Figura 49.** *Graph*. Representació de les correlacions superiors a 0,85.

Si mirem amb més detall les xarxes de gens trobem que el grup dels 5 gens UBL1 (SUMO1), ZNF32, IGF2R, EDG1, CST3, que apareixia a l'agafar el llindar de 0,9 es manté i no se n'hi afegeix cap de nou. Per tant aquest grup és força robust. En canvi, dos gens que apareixien sols amb el llindar de correlació establert a 0,9 (PPP1CA i PTMA) ara estan entrelligats amb una xarxa més complexa de nous gens. I a més, trobem que algunes de les correlacions són negatives (els valors exactes es troben a l'annex I, a la Taula 49). A la Figura 50 es veu el sentit de la correlació agafant el codi de colors descrit a l'escala.



**Figura 50.** Heat map de les correlacions superiors a 0,85. Els valors propers a 1, que corresponen a una correlació positiva, apareixen de color verd, i els valors propers a -1, que corresponen a una correlació negativa, apareixen de color vermell.

Anar més enllà en l'estudi d'aquestes correlacions s'allunya dels objectius de la present tesi, així i tot volíem veure a títol il·lustratiu la informació que es pot extreure de les dades. Per això, a partir d'aquestes correlacions vam intentar trobar alguna relació funcional. En el cas del grup de 5 gens amb una correlació més alta hem intentat esbrinar quina funció té cada gen i quin podria ser el punt comú entre tots ells. A la Taula 41 hi ha el nom complet i la funció d'aquests gens. Veiem que no pertanyen a una mateixa via, però sí tenen funcions comunes. Hi ha dos gens que codifiquen per proteïnes receptores, i tres gens que codifiquen per proteïnes amb funcions reguladores. Però és difícil trobar un lligam directe entre les seves funcions, i per tant no vam buscar-hi més interpretacions, ja que s'allunyava del nostre objecte d'estudi.

**Taula 41.** Llistat de gens que presenten una correlació major de 0,9.

Símbol	ACC <sup>(1)</sup>	Nom del gen	Funció <sup>(2)</sup>
<a href="#">IGF2R</a>	<a href="#">NM_000876</a>	insulin-like growth factor 2 receptor	<i>Transport of phosphorylated lysosomal enzymes from the Golgi complex and the cell surface to lysosomes. Lysosomal enzymes bearing phosphomannosyl residues bind specifically to mannose-6-phosphate receptors in the Golgi apparatus and the resulting receptor-ligand complex is transported to an acidic prelysosomal compartment where the low pH mediates the dissociation of the complex. This receptor also binds insulin growth factor II. Similarity: Contains 1 fibronectin type II domain.</i>
<a href="#">SUMO1 (UBL1)</a>	<a href="#">NM_003352</a>	SMT3 suppressor of mif two 3 homolog 1 (yeast)	<i>Associates with RAD51/RAD52. Involved in targeting RANGAP1 to the nuclear pore complex protein RANBP2. Similarity: Belongs to the ubiquitin family. SMT3 subfamily.</i>
<a href="#">ZNF32</a>	<a href="#">NM_001005368</a>	zinc finger protein 32 (KOX 30)	<i>May be involved in transcriptional regulation. Similarity: Belongs to the krueppel C2H2-type zinc-finger protein family.</i>
<a href="#">EDG1</a>	<a href="#">NM_001400</a>	endothelial differentiation, sphingolipid G-protein-coupled receptor, 1	<i>Receptor for the lysosphingolipid sphingosine 1-phosphate (S1P). S1P is a bioactive lysophospholipid that elicits diverse physiological effect on most types of cells and tissues. This inducible epithelial cell G-protein-coupled receptor may be involved in the processes that regulate the differentiation of endothelial cells. Seems to be coupled to the G(i) subclass of heteromeric G proteins. Similarity: Belongs to family 1 of G-protein coupled receptors.</i>
<a href="#">CST3</a>	<a href="#">NG_000839</a>	cystatin C (amyloid angiopathy and cerebral hemorrhage)	<i>As an inhibitor of cysteine proteinases, this protein is thought to serve an important physiological role as a local regulator of this enzyme activity. Similarity: Belongs to the cystatin family.</i>

<sup>(1)</sup>NCBI Reference Sequences (RefSeq): <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi> (Search: Nucleotide)

<sup>(2)</sup>GeneCards: <http://bioinfo.weizmann.ac.il/cards/index.shtml> . Búsqueda fetes el 20 de diciembre de 2004.

### **Anàlisi de components principals: ACP**

L'ACP és un mètode estadístic que a partir d'una matriu de dades de  $N$  individus x  $P$  variables, analitza quines són les variables redundants i genera un conjunt de  $P$  eixos factorials (o simplement factors) que recullen les redundàncies.

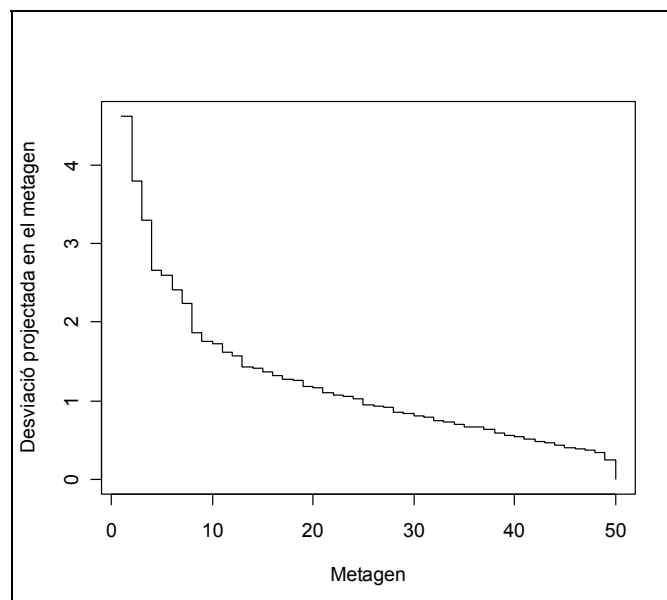
En el nostre cas té molt sentit aplicar una anàlisi de components principals, ja que les nostres variables, els gens, és ben sabut que estan correlacionats. D'aquesta manera, generarem uns nous factors que anomenarem **metagens**, que recolliran en una única nova variable un mateix patró d'expressió comú per a un grup de gens.

Tècnicament, l'anàlisi de components principals realitza una projecció ortogonal de les dades sobre un nou espai vectorial, de forma que els eixos factorials són

simplement combinacions lineals de les variables originals. Per a fer aquesta projecció es procedeix a fer una descomposició de valors singulars a partir de la matriu de variàncies i covariàncies o bé a partir de la matriu de correlacions. En el nostre cas, hem optat per la matriu de variàncies i covariàncies, ja que d'aquesta manera cada un dels gens tindrà en l'ACP un pes proporcional a la seva variància. D'altra banda, el fet que la projecció sigui ortogonal implica que els eixos factorials seran independents entre sí. Aquest aspecte és molt important i té una lectura bona en el nostre context d'anàlisi, ja que determinarem patrons d'expressió, que seran comuns en grups de gens, però que alhora seran independents entre ells. Finalment, la projecció ortogonal es realitza de forma que la variància recollida per tots els eixos és, òbviament, igual a la variància de tots els gens, però a més a més, els eixos estan ordenats de major a menor variància. Així, els darrers eixos són residuals i expliquen comportaments molt particulars. En canvi, els primers eixos factorials solen ser molt informatius i recullen els principals patrons d'expressió presents en el conjunt de dades.

L'anàlisi de les dades de microarrays mitjançant ACP s'ha demostrat que és un bon mètode en estudis anteriors (Armstrong et al., 2002; Hilsenbeck et al., 1999; Wikman et al., 2002).

El gràfic que mostrem a continuació visualitza com decreix la variància de cadascun dels eixos factorials obtinguts amb les nostres dades de microarrays (Figura 51).



**Figura 51.** Desviacions recollides pels metagens. Observem que a partir del desè metagen, aproximadament, la desviació explicada es podria considerar residual.

Degut a la ortonormalitat de la projecció, els valors dels gens projectats a cada un dels metagens, es poden interpretar directament com a la correlació dels gens

sobre el metagen. D'altra banda, la suma de les correlacions al quadrat, per a un mateix metagen, és igual a la unitat. Per tant, les correlacions al quadrat, anomenades contribucions, mesuren com contribueixen cadascun dels gens per a cada metagen, el que ens ajuda enormement a interpretar quins són els gens que participen en el patró d'expressió proposat per cada metagen. A continuació mostrem els gens que acumulen un 50% de la contribució per cada metagen (Taula 42). A l'annex I hi ha un llistat on es troben les contribucions exactes d'aquests gens.

**Taula 42.** Gens que contribueixen fins a un 50% en els 10 primers eixos factorials.

<b>EIX</b>	<b>GENS que contribueixen fins a un 50%</b>
1	TNC (HXB), ATP6V1C1 (ATP6C), PTMA, NAB2, AHCY, SRP72, RBM4, MAPK6, CD59, BNC, TLH6
2	ARL6IP, H96850, ITGB1, NCBP1, KIAA0107, LYN, EXOSC9 (PMSCL1)
3	CD164, PAI-RBP1, ARL6IP, MGC15407, N91290, PHLDA1, PPP2R1B, LRBA (CDC4L), DDX1, ALB, ENG, RBBP2, KIAA0107, R62773, TCF7L2, LOC51738, TOP2A, NUP88, N52568, N78103
4	KIAA0265, LYN, BIN1, CSE1L, EXOSC9 (PMSCL1), DAZAP2, ADPRTL1, FUBP3, HPR6.6, R88904, CloneFBD3, TBX2, N52568, HOXD9
5	MESDC2, KIAA0265, IGF2, ARL6IP, WDR31, KIAA0203, HUMAUANTIG, NUP155, LAMA5, SAT, ZNF220, ENG
6	ZNF32, SUMO1 (UBL1), EDG1, CST3
7	1LBG, N52568, REV1L, LOC51738, PRKAG2, IL16, C3AR1, R88904, KIAA0210, HINT, NUP155
8	FKBP5, PHLDA1, TPR, CloneFBD3, GSTTLp28, HINT, BIN1, C5, NUP155, R62773, CAPZA2, C20ORF1
9	CAPZA2, HARS, LYN, HUMAUANTIG, CD9, FKBP5, GMFB, RBBP2, LTBP1, EXOSC9 (PMSCL1), LYPLA1, PPP2R1B, VCAM1
10	PGM5, FUBP3, MTMR6, KPNA1, HPR6.6, N91290, R62773, EIF4A2, LTBP1, HUMAUANTIG, C20ORF1, LCN2, ZNF220, TUBB, WBSCR14

El que resulta interessant és analitzar fins a quin punt els metagens són capaços de distingir cadascuna de les tipologies de les mostres definides per les variables clíniques separatament. La manera més senzilla de respondre a aquesta pregunta és calculant les diferents mitjanes de cadascun dels valors projectats en els eixos factorials i avaluar si existeix cap diferència, emprant una prova ANOVA. A continuació mostrem la llista dels metagens amb un p-valor inferior a 0,05 per les diferents variables clinicopatològiques.

Un cop arribats a aquest punt, el següent pas seria fer una validació per estudiar la robustesa d'aquests metagens. És a dir, caldria comprovar si tornant a fer el mateix anàlisi amb un grup de mostres extern es confirmen els mateixos resultats (en el millor dels casos). O bé utilitzant el mateix grup de mostres deixant-ne una a fora de l'estudi (*leave-one-out*) i comprovant si els resultats que s'obtenen són els mateixos. Això es pot fer repetides vegades i al final s'obté una taxa d'error.

En la present tesi no s'aprofundeix en aquest estudi perquè aquesta feina és més per a un bioestadístic. Es tracta d'una tasca llarga que requereix la col·laboració de moltes disciplines (biòlegs, informàtics, i estadístics,). En un futur continuarem aprofundint en aquest aspecte per tal d'elaborar les hipòtesis que associen els metagens amb cada una de les característiques estudiades.

Nosaltres ens hem centrat en els resultats directes de les hibridacions, buscant patrons d'expressió que ens separessin diferents grups.

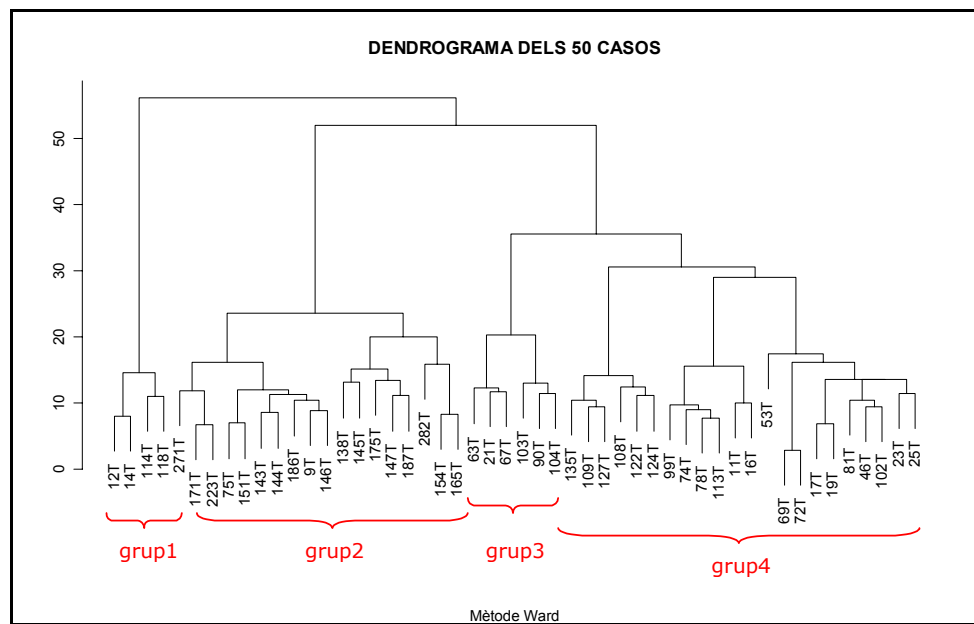
**Taula 43.** *p*-valors inferiors a 0,05 pels metagens que distingeixen les diferents característiques clinicopatològiques.

<b>metagen</b>	<b>supervivència lliure de malaltia</b>
4	0,002
31	0,019
23	0,026
49	0,030
<b>metagen</b>	<b>localització</b>
15	0,016
<b>metagen</b>	<b>grups CGH</b>
12	0,005
48	0,029
41	0,049
<b>metagen</b>	<b>Dukes</b>
12	0,011
3	0,018
7	0,043
<b>metagen</b>	<b>mutació TP53</b>
24	0,006
<b>metagen</b>	<b>mutació RAS</b>
18	0,004
2	0,044

### **Agrupaments jeràrquics a partir de les dades d'expressió**

Amb l'objectiu d'interpretar les dades d'expressió dels 128 gens de forma senzilla i pràctica es va fer un anàlisi multivariable de classificació. Resumint, vam utilitzar el mètode Ward (vegeu l'apartat de *Materials i mètodes*) per veure com

s'ordenaven els casos usant els resultats d'expressió dels 128 gens. La manera de representar les distàncies és mitjançant dendrogrames. A les figures 54 i 55 es dibuixen els dendrogrames dels casos i dels gens, respectivament. Podem veure com es poden identificar diferents grups segons quina sigui l'alçada del nivell de tall. Així, en el dendrograma dels casos (Figura 52), si tallem per sota la tercera ramificació ens queden quatre grups. El primer format per quatre casos (*grup 1*), el segon format per divuit casos (*grup 2*), el tercer format per sis casos (*grup 3*), i el quart format per vint-i-dos casos (*grup 4*).



**Figura 52.** Dendrograma dels 50 casos utilitzant com a distància les diferències d'expressió de 128 gens.

Pel què fa al dendrograma dels gens (Figura 53), veiem que es diferencien clarament dos grups, uns que tendeixen a sobreexpressar-se i uns altres que tendeixen a infraexpressar-se. A la Taula 44 hi ha uns quants exemples dels gens amb una tendència més clara a la infraexpressió i a la sobreexpressió.

Un cop calculades les distàncies entre gens i casos independentment, si dibuixem el dendrograma dels casos com a columnes, i el de les raons d'expressió com a files, pintant les caselles resultants en funció de si hi ha hagut sobreexpressió (de color vermell) o infraexpressió (de color verd), obtenim el que s'anomenen *heat maps* (Figura 54).



**Taula 44.** Mitjana de la raó dels gens més diferencialment expressats en els 50 casos.

Acc	Símbol	Chr	Mitjana	Nom del gen
H12279	PGM5	9	-3.42	<i>phosphoglucomutase 5</i>
W07367	FUBP3		-2.89	<i>far upstream element (FUSE) binding protein 3</i>
W52186	SYNPO2	4	-2.20	<i>synaptopodin 2</i>
N76867	TCF7L2	10	-1.53	<i>transcription factor 7-like 2 (T-cell specific, HMG-box)</i>
AA429297	ZFP103	2	-1.35	<i>zinc finger protein homologous to Zfp103 in mouse</i>
AA402960	RNF5	6	-1.23	<i>ring finger protein 5</i>
H17504	MAPK6	15	-1.11	<i>mitogen-activated protein kinase 6</i>
H94262	EHM2	9	-1.09	<i>EHM2 gene</i>
N50247	MTMR6	13	-1.09	<i>myotubularin related protein 6</i>
AA504656	LTBP1	2	-1.08	<i>latent transforming growth factor beta binding protein 1</i>
N73030	C5	9	-1.05	<i>complement component 5</i>
N57872	TLH6		-0.96	<i>alanine-glyoxylate aminotransferase homolog</i>
AA598572	SYK	9	-0.95	<i>spleen tyrosine kinase</i>
R11586	REV1L	2	-0.90	<i>REV1-like (yeast)</i>
H96850	H96850		0.72	-
W67174	ITGB1	19	0.73	<i>Homo sapiens integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12) (ITGB1), mRNA</i>
AA406601	ABLIM	10	0.74	<i>actin binding LIM protein 1</i>
AA258396	PHLDA1	12	0.75	<i>pleckstrin homology-like domain, family A, member 1</i>
AA441895	GSTTLp28	10	0.79	<i>glutathione-S-transferase like, glutathione transferase omega</i>
AA488609	NUP88	17	0.82	<i>nucleoporin 88kD</i>
AA083228	CAPZA2	7	0.87	<i>capping protein (actin filament) muscle Z-line, alpha 2</i>
AA278749	NCBP1	9	0.89	<i>nuclear cap binding protein subunit 1, 80kD</i>
AA487912	GNB1	1	0.95	<i>guanine nucleotide binding protein (G protein), beta polypeptide 1</i>
AA434487	NAB2	12	0.97	<i>NGFI-A binding protein 2 (ERG1 binding protein 2)</i>
H20652	ARL6IP	16	1.03	<i>ADP-ribosylation factor-like 6 interacting protein</i>
H73329	C20ORF1	20	1.38	<i>chromosome 20 open reading frame 1</i>

Chr: cromosoma

Així comprovem, com els quatre grups que separàvem amb el dendrograma de les mostres tenen uns patrons d'expressió distingibles entre ells. Per exemple veiem com el *grup 2* de casos té un conjunt de gens que tendeixen a la sobreexpressió, i al *grup 3* en canvi estan infraexpressats. Són aquestes diferències les que ens defineixen els grups.

A continuació vam explorar amb més detall aquests quatre grups per veure si tenien alguna similitud amb els grups definits mitjançant la CGH.

A la Taula 45 veiem que els grups definits a partir dels resultats de la CGH no es corresponen als que hem obtingut per l'expressió dels arrays. És difícil avaluar el

grup amb MSI ja que només es tracta de tres casos. Pel què fa als casos que no tenien cap canvi a nivell cromosòmic estan representats principalment a la branca del *grup 3*, tot i que hi ha casos també a la resta de branques. Els grups que anomenem ICML i ICR estan repartits quasi al 50% entre les branques del grup 2 i 4 que són les més nombroses.

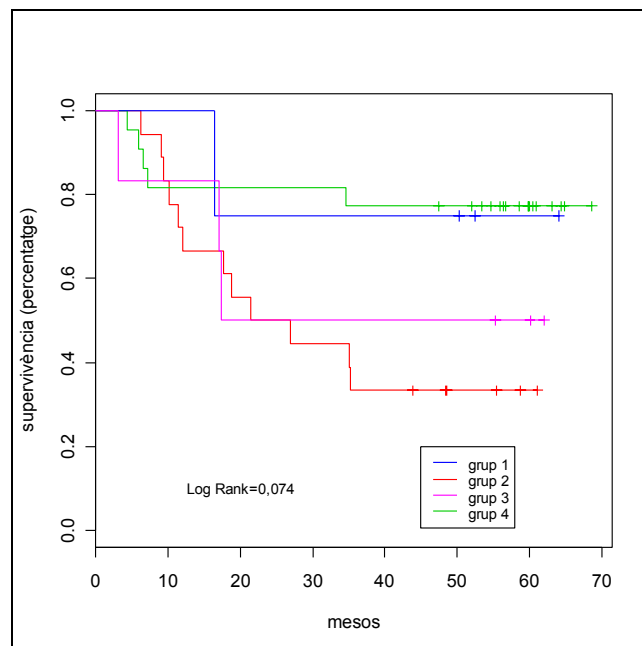
Per tant podem concloure que la informació que obtenim a partir de la CGH i dels arrays no és equivalent, i que per tant és complementària.

**Taula 45.** Comparació entre la classificació de la CGH i dels arrays.

	grup1	grup2	grup3	grup4
<b>MSI</b>	0	1	0	2
<b>NORMAL</b>	1	2	<b>4 (50%)</b>	1
<b>ICML</b>	1	9 (41%)	1	<b>11 (50%)</b>
<b>ICR</b>	2	6 (35%)	1	<b>8 (47%)</b>

### Supervivència lliure de malaltia

Amb l'objectiu expressat reiteradament en aquesta tesi de fer grups per valorar la capacitat de distingir l'agressivitat dels tumors, vam estudiar les corbes de Kaplan-Meier per aquests quatre grups (Figura 55). Veiem com el *grup 2* és el que té una supervivència lliure de malaltia més baixa, i el *grup 4* és el que té un més bon pronòstic. Per tant, el patró d'expressió d'aquests 128 gens ens defineix dos grups de pacients que tenen tumors de diferent agressivitat.



**Figura 55.** Corbes de Kaplan-Meier de la supervivència lliure de malaltia dels 4 grups definits per microarrays.

### Amplificació i expressió del gen CCND1

Per últim vam comparar els resultats de l'amplificació del gen CCND1 amb els d'expressió de l'RNA. Només hi ha un cas que presenti amplificació clara d'aquest gen, el 23, però com es pot veure a la Taula 46 el trobem infraexpressat a nivell d'RNA quan ho comparem amb el POOLN. Pel què fa a la resta de casos tampoc es veu cap mena de correlació. Els resultats d'expressió i amplificació no coincideixen, però això ja s'ha descrit sovint a la literatura: ja que més quantitat de DNA no sempre va acompanyada d'una major quantitat d'RNA, ni tampoc una elevada quantitat d'mRNA implica més proteïna, o sigui funció.

**Taula 46.** Comparació entre els resultats dels arrays i la PCR quantitativa.

cas	guany per CGH	pèrdua per CGH	log2 (QPCR)	log2 (raóT/POOLN)
23 T	amp11q13	-	3,8	-0,4
72 T	amp11q13	-	0,2	-0,32
108 T	amp11q13	-	-0,2	0,22
109 T	11q	-	0,5	0,32
113 T	11	-15q	0,07	0,42
146 T	11	-	0,2	0,57
186 T	amp11q13	-	0,4	0,42
187 T	amp11q13	-15	-0,2*	0,17
223 T	amp11q13	-15	0,4*	0,44
271 T	amp11q13	-15	0,6*	0,22
282 T	amp11q13	15q15-q23	0,2*	0,31

\*Casos on s'ha perdut el cromosoma 15q.

