

VI. DISCUSSIÓ

Com hem vist el procés que porta a una cèl·lula a convertir-se en cancerosa és llarg i complex. Malgrat el gran nombre d'estudis que s'han fet fins ara, l'heterogeneïtat de les alteracions i el gran desconeixement que tenim sobre el paper que juguen moltes d'elles fa que encara no tinguem una imatge clara de quins son els paràmetres més importants, ni com s'han d'abordar, tant pel què fa a l'avaluació del pronòstic, com pel disseny d'una teràpia coadjuvant més eficaç.

Qualsevol anàlisi global de biòpsies humanes ens dóna una panoràmica a vista d'ocell de la situació d'un conjunt de cèl·lules en un moment determinat: el de l'extirpació del tumor. Es tracta de fer una descripció detallada dels tumors (a nivell del genoma, el transcriptoma, i el proteoma) per obtenir un diagnòstic el màxim acurat i així classificar els tumors amb un objectiu principal: que tingui un valor clínic. És a dir, que ens serveixi com a indicador de pronòstic, i alhora que ens sigui útil per decidir el tractament més efectiu possible.

Abans d'entrar en l'anàlisi i interpretació de les nostres dades cal fer algunes consideracions referents a les mostres. En abordatges com el que recull aquesta tesi, una de les limitacions més importants alhora d'avaluar els resultats és que es parteix d'un nombre de casos reduït. A més també s'ha de determinar si el disseny de la sèrie és adient per adreçar les qüestions o objectius plantejats. En el nostre cas, el grup de mostres escollit havia de ser suficientment gran per poder fer grups i definir predictors de classe, donat que aquesta és una de les premisses plantejades inicialment. Una mida mostral de 50 casos es va considerar raonable. Per donar consistència a les conclusions, però, s'haurien de poder fer estudis de validació creuada sense que això representés una feina inassequible, tant pel què fa a aspectes econòmics com logístics. Amb mides mostrals inferiors, és impossible validar els resultats sense tenir un conjunt de mostres independent per dues raons principals. La primera, perquè l'error estimat té una variància molt elevada en conjunts de mostres petites; i la segona, perquè les mostres utilitzades per construir el predictor en l'estudi inicial potser no són prou representatives per reflectir totes les característiques de la població (Simon et al., 2003). Cal remarcar que el nostre estudi és de caràcter exploratori, i que abans d'aplicar les conclusions obtingudes s'ha de fer una validació amb un conjunt prou gran de mostres independents.

Hem de destacar que els casos analitzats en la present tesi són dels estadis de Dukes B i C amb la voluntat de poder distingir dos grups que tot i que a nivell clinicopatològic estan molt ben definits, el grau d'incertesa referent a l'evolució clínica i resposta al tractament a nivell individual és molt alta (Figura 32, de

l'apartat de *Resultats*). En conseqüència l'addició de qualsevol nou paràmetre predictor serà d'enorme utilitat.

Classificació dels tumors en base al patró d'alteracions cromosòmiques

La primera aproximació que hem fet ha estat considerar el nombre i tipus d'alteracions cromosòmiques per obtenir subclassificacions dels tumors. Ja a començaments del segle XX Theodor Boveri (Boveri, 1914) va proposar la idea que l'acumulació d'alteracions podia ser responsable del creixement i la progressió dels tumors. En el cas del càncer colorectal està a bastament descrit que a més alteracions el pronòstic empitjora. Així, quan s'estudien el número d'alteracions cromosòmiques mitjançant cariotips (Bardi et al., 2004), o el grau d'aneuploidia (Risques et al., 2001), trobem que hi ha una correlació amb mal pronòstic. També s'ha descrit que a mesura que els adenomes progressen cap a carcinomes acumulen més alteracions cromosòmiques a nivell de CGH (Ried et al., 1999).

En la nostra sèrie, la mitjana d'alteracions per cas és de 8,8 (amb un rang de variació de 0 fins a 20 alteracions). Aquests resultats no difereixen gaire amb els publicats per De Angelis (De Angelis et al., 1999) on trobaven una mitjana de 7 alteracions (amb un rang de 0-19), i per Korn (Korn et al., 1999) que descriuen 10,2 alteracions de mitjana per cas (amb un rang de 1-20).

En base a la distribució del nombre d'alteracions es poden fer varies classificacions (Figura 28, de l'apartat de *Resultats*), però donat el baix nombre de casos de què disposem i que no existeixen criteris objectivables, hem preferit no definir subgrups de tumors en funció del nombre d'alteracions, ja que aquestes classificacions són poc comparables i "exportables" a d'altres estudis. Això ve reforçat pel fet que el nombre d'alteracions com a tal no està associat a cap paràmetre clinicopatològic (Taula 24, de l'apartat de *Resultats*). Les úniques associacions que s'observen són: un major nombre d'alteracions en els tumors amb mutació en el gen TP53, i la presència d'amplificacions gèniques (que es discuteix més endavant).

Aquesta associació entre la inactivació del gen TP53 i l'acumulació d'alteracions a nivell de cromosomes ja ha estat àmpliament descrita tant en estudis in vitro (Williams et al., 1997; Yin et al., 1992) com sobre mostres tumorals (De Angelis et al., 1999; Georgiades et al., 1999; Leslie et al., 2003; Risques et al., 2003; Sugai et al., 2003).

Un cop vist que el nombre i el tipus (com discutirem més endavant) d'alteracions trobades és semblant al que descriuen la resta d'estudis, ens plantegem fer la

classificació dels tumors. Aquesta classificació ens ha de servir com a base per analitzar les associacions clinicopatològiques i moleculars amb la supervivència lliure de malaltia. Una de les classificacions de tumors colorectals basades en el cariotip i que hem considerat com a més sòlida és la de Dutrillaux (Dutrillaux, 1995). La nostra classificació recull els grups de tumors NORMALS i ICMT que tindrien correspondència directa amb dos dels grups de Dutrillaux: *Normal Type* i *Monosomic Type*, respectivament. Falta esbrinar si el nostre tercer grup (ICR) és un autèntic calaix de sastre o correspon al grup *Trisomic Type*.

En base als coneixements que ja teníem, i a la informació que hem generat en la present tesi hem classificat els tumors segons el patró d'alteracions genòmiques en els següents grups:

Grup de tumors amb MSI

Com ja s'ha descrit en anteriors treballs els casos amb inestabilitat de microsatèl·lits presenten molt poques alteracions a nivell cromosòmic (Dutrillaux, 1995; Eshleman et al., 1998; Remvikos et al., 1995), si bé estudis utilitzant la CGH d'alta resolució (Nakao et al., 2004) i tècniques de *fingerprinting* (Risques et al., 2003) sí que són capaces de trobar desequilibris cromosòmics en aquests tumors, encara que amb una menor proporció que la resta. En el nostre grup de mostres només hi ha tres casos amb MSI, que representen un 6% del total, i per tant no podem treure'n cap conclusió donada la seva baixa freqüència. El que sí confirmem és que tenen poques o cap alteració cromosòmica (0, 1 i 2 cromosomes alterats), i també es localitzen al còlon dret, una característica ben establerta per d'altres estudis (Ionov et al., 1993; Thibodeau et al., 1993). Una altra característica d'aquest grup de tumors és que són menys agressius i presenten més bon pronòstic (Gonzalez-Garcia et al., 2000; Gryfe et al., 2000; Hemminki et al., 2000; Watanabe et al., 2001), si bé aquesta dada no és possible confirmar-la en la nostra sèrie donat el baix nombre de casos.

Grup de tumors amb cariotip normal

En aquesta sèrie trobem vuit casos que no tenen inestabilitat de microsatèl·lits, ni tampoc presenten cap alteració a nivell de CGH. Això ens indica que els canvis que puguin tenir a nivell genòmic afectaran petites regions (menors de 10 Mb, que és la resolució màxima de la tècnica). Per tant, suposem que aquests tumors progressen per un altre camí i acumulen alteracions a un altre nivell: ja sigui mitjançant alteracions en el patró de metilació, mutacions puntuals, translocacions balancejades, o no balancejades (delecions i guanys) petites. Dels vuit casos que formen aquest grup tres moren per la malaltia. A excepció d'un cas tots es localitzen al còlon esquerre, i només dos d'ells tenen el gen TP53 mutat. Aquest grup de tumors és molt interessant per la seva capacitat de

progressar sense presentar una inestabilitat genòmica evident. Aquesta observació també ha estat feta per altres autors (Georgiades et al., 1999; Goel et al., 2003; Jones et al., 2005; Risques et al., 2003), i per tant sembla necessari considerar aquest grup a part. Si ens fixem en estudis anteriors, Dutrillaux (Dutrillaux, 1995), per exemple, va separar un grup de tumors normals, però no va distingir els que tenien MSI dels que no.

Altres autors han definit un grup amb poques alteracions a nivell de CGH posant el llindar a les 6 alteracions (Chan et al., 2001). Aquests autors defineixen el grup amb el nom de MACS (de l'anglès, *microsatellite and chromosome stable*), que també es caracteritzen per no tenir inestabilitat de microsatèl·lits, perquè la majoria es troben al còlon esquerre, i per presentar una histologia moderadament diferenciada. Al nostre parer, no s'haurien de barrejar els casos amb fins a 6 alteracions a nivell de CGH, amb aquells que no en tenen cap. Ja que el grup sense cap alteració té prou entitat per ell mateix i representa un 16% dels casos de la nostra sèrie. També pensem que aquells casos que ja han acumulat algunes alteracions a nivell de CGH podrien tractar-se de tumors en estadis inicials que podrien evolucionar cap al grup de CIN, com els mateixos autors reconeixen (Chan et al., 2001). És per això que creiem molt més encertat i exacte considerar el grup sense cap canvi independentment.

En aquest moment no sabem quins poden ser els mecanismes moleculars que es troben darrera de la progressió tumoral en el grup normal. Si bé es pot plantejar que alteracions a nivell epigenètic o subcromosòmic poden ser el motor de la progressió en aquest grup, calen estudis específics per identificar-ne les bases moleculars.

Grup de tumors ICML

Aquest grup recull el percentatge més alt de casos (44%) i es caracteritza per presentar un alt nombre d'alteracions, pèrdues dels braços dels cromosomes 17p i 18q, alta freqüència de mutacions en TP53 i localització preferent al còlon esquerre. Aquestes característiques fan pensar que aquest grup correspon als definits prèviament per Dutrillaux (Dutrillaux, 1995) com a *Monosomic Type*, al *Neartriploid* de Bardi (Bardi et al., 2004), o al *High aneuploid* de Risques (Risques et al., 2003).

Referent a la supervivència, dins d'aquest grup que en conjunt té una supervivència lliure de malaltia del 60% al final del seguiment, hem comprovat que hi ha una gran diferència entre aquells casos que presenten fins a 10 alteracions cromosòmiques dels que en tenen més de 10 (vegeu la Figura 35). Aquestes diferències ens podrien indicar que dins del grup ICML, l'acumulació

d'alteracions cromosòmiques està associada a l'agressivitat i per tant seria un bon indicador de pronòstic.

Grup de tumors ICR

És evident que ara per ara, no podem considerar que aquest grup tingui entitat pròpia, donat que els tumors no presenten característiques diferenciadores de la resta. No obstant, és probable que tingui un notable solapament amb d'altres descrits com a *Trisomic type* (Dutrillaux, 1995), *Hyperdiploid* (Bardi et al., 2004), o *Low aneuploid* (Risques et al., 2003). En el nostre estudi aquest és el grup amb una pitjor supervivència en conjunt, quasi el 60% dels pacients recauen després de 5 anys de seguiment. A diferència del grup ICML, el número d'alteracions no fa cap diferència en quant a supervivència en els ICR (vegeu la Figura 36), el que indica que serien altres factors els responsables de la progressió i la malignitat, tal com ja s'ha suggerit en d'altres estudis (Bardi et al., 2004; Jones et al., 2005; Risques et al., 2003).

Alteracions específiques

Les alteracions cromosòmiques recurrents van associades a la presència de gens rellevants pel procés tumoral, i per tant s'han de considerar també de forma independent. En el nostre estudi diferenciem entre amplificacions, que afecten regions relativament petites incloent un o uns pocs gens, i les que afecten a regions cromosòmiques importants i que consisteixen en guanys i pèrdues.

Alteracions estructurals

En aquest estudi no hem pogut estudiar a fons les alteracions estructurals, ja que mitjançant la tècnica de la CGH només podem detectar canvis estructurals quan afecten a un tros de cromosoma de forma no equilibrada. Tots aquells canvis equilibrats ens passaran per alt. Encara i així, vam intentar comparar les alteracions numèriques de les estructurals que es podien identificar, i vam trobar que presentaven correlació entre elles. És a dir, a més nombre d'alteracions estructurals també hi ha més alteracions numèriques. Això indica que alguns dels mecanismes responsables d'ambdós processos són comuns, ja sigui a nivell que presentin alteracions que donin la permissivitat perquè es puguin donar a terme (de forma indirecta). O bé perquè el mecanisme actiu que provoca l'alteració sigui el mateix. Però no vam aprofundir més en aquest tema perquè com que no estem identificant totes les alteracions estructurals no valia la pena. Només destacar que en un treball de Bardi (1995), utilitzant la citogenètica clàssica, troben que els tumors més diferenciats presenten alteracions numèriques o cariotips normals, en canvi tumors més desdiferenciats presentaven més alteracions estructurals.

Guanys i pèrdues

Els guanys i pèrdues es consideren variacions d'una còpia (o d'unes poques en el cas de guanys). Nosaltres vam definir com a alteracions recurrents aquelles que trobem en més d'un 15% dels casos. D'aquesta manera hem trobat alterats en forma de guanys recurrents els cromosomes 20q (74%), 8q (46%), 13 (44%), 7 (32%), 17q (32%), 20p (26%), 19 (22%) i X (18%). Pel què fa a les pèrdues, els cromosomes més freqüentment alterats són 18 (40%), 17p (48%), 8p (38%), 4 (32%), 14 (22%) i 15 (18%). A la Taula 47 hi ha un resum comparatiu dels nostres resultats i de quatre autors més que també van utilitzar la CGH, més un estudi fet mitjançant citogenètica clàssica (Muleris et al., 1990). Si ho mirem amb detall veiem que l'estudi fet per De Angelis té una proporció de guanys molt semblant als nostres 50 casos. Notis que aquest estudi es va fer amb 45 casos, per tant la mida mostral s'acosta a la d'aquesta tesi. I a més, la majoria de casos que ells utilitzen són dels estadis de Dukes B i C, més 8 del D. Contràriament, en el cas de l'estudi fet per Korn (Korn et al., 1999) es van utilitzar metàstasis de càncer colorectal, per això no és d'estranyar que el tant per cent d'alteracions és més elevat, tot i que els cromosomes afectats són els mateixos.

Paral·lelament, trobem que un dels cromosomes amb el resultat més dispar és el braç 1p, on nosaltres quasi no trobem canvis (només en 2 casos hi ha pèrdua), i altres autors l'arriben a trobar perdut en un 56% de les metàstasis i en un 53% de tumors estudiant els cariotips mitjançant citogenètica clàssica.

Taula 47. Alteracions cromosòmiques recurrents en càncer colorectal.

Tesi (N=50)	Chr	Angelis (N=45)	Korn (N=27) METASTASIS	Aragane (N=30)	Ried (N=16)	Muleris (N=100)
guanys						
38%	+7p	33%	44%	36,7% (p21)	18,75%	33% ^a
32%	+7q	-	26%	30% (q31-36)	-	-
46%	+8q	33%	44%	43% (q23-24)	46,75%	40%
44%	+13	45%	48%	-	50%	51%
26%	+20p	-	33%	33,3%	-	-
74%	+20q	67%	85%	63,3%	75%	43% ^a
18%	+Xp	-	26%	-	18,75%	50% ^a
pèrdues						
4%	-1p	27%	56%	-	6,25%	53% del(1)(p3)
32%	-4p	-	37%	-	-	-
30%	-4q	36%	48%	-	18,75% (sencer)	44%
10%	-5q	27%	26%	-	6,25%	47% del(5)(q2)
38%	-8p	27%	59%	-	43,75%	51%
18%	-15q	27%	30%	-	-	42% ^a
48%	-17p	22%	41%	-	12,5%	76%
40%	-18p	-	48%	-	-	78% ^a
66%	-18q	49%	89%	36,7% (q21)	12,5%	78% ^a

^a Cromosoma sencer. Chr: cromosoma

A quasi tots els cromosomes el número de canvis que hem trobat cau dins del rang de variació dels cinc estudis agafats com a referència, a excepció dels cromosomes 7q, 13, Xp, 1p, i 15q; on trobem valors per sota o per sobre dels descrits però amb una diferència petita.

Pèrdues de cromosomes

Existeixen nombrosos estudis on descriuen que moltes d'aquestes alteracions, sobretot a nivell de pèrdues al·lèliques, s'associen a un pronòstic diferencial en CCR (Arribas et al., 1999; Bardi et al., 2004; Castells et al., 1999; Choi et al., 2002; Diep et al., 2003; Halling et al., 1999; Zhou et al., 2002).

Pèrdues a 12p

Tan sols quatre tumors mostren pèrdues a 12p, però tots ells presenten recidiva. Per tant, al braç 12p hi ha un possible gen supressor de tumors. Dels molts gens que hi ha en aquest braç un dels més interessants és el gen CDKN1B (anteriorment anomenat p27). Aquest gen és membre de la família de proteïnes CDKI (de l'anglès, *cyclin-dependent kinase inhibitor*). S'ha postulat que és un gen supressor de tumors, perquè a part de la seva funció com a CDKI, també regula la resistència a drogues en tumors sòlids, promou l'apoptosi, actua com a protector davant la inflamació per ferides, i té un paper en la diferenciació cel·lular (Lloyd et al., 1999). Per tant és un molt bon candidat com a marcador tumoral. Així mateix, en CCR s'ha demostrat que el gen CDKNB1 és un marcador de pronòstic independent (Loda et al., 1997), i que té una correlació significativa amb el grau del tumor (Ciaparrone et al., 1998).

Pèrdues a 18q

D'entre els 50 casos analitzats l'alteració més recurrent pel què fa a les pèrdues la trobem al braç 18q. En molts estudis s'ha vist que la pèrdua d'heterozigositat de 18q s'associa a un pronòstic dolent, però els resultats no són del tot evidents. En un treball de Lanza (Lanza et al., 1998) només miren tres marcadors situats a 18q i troben que hi ha correlació amb la supervivència en un grup de tumors d'estadis de Dukes B i C, en el sentit que aquells casos que perden els tres marcadors tenen un pitjor pronòstic. En canvi, en un estudi fet per Martínez-López (Martinez-Lopez et al., 1998) també mirant 3 marcadors troben que s'associen a mal pronòstic en pacients d'estadi de Dukes C, i no en pacients en estadi de Dukes B. D'altres autors, com Ookawa (Ookawa et al., 1993), comparen les alteracions que apareixen de nou a metàstasis aparellades amb els seus tumors primaris, i troben que de manera significativa aquesta alteració es troba més en les metàstasis, per tant parlen d'un bon marcador de la progressió del tumor. Pel què fa als nostres resultats, tenint en compte que mirem la pèrdua del braç 18q per CGH, no hem vist que hi hagi cap diferència significativa entre

aquesta i la supervivència, tant si mirem tots els casos conjuntament, com si separem els pacients en estadi de Dukes B i C.

El que queda clar és que al braç petit del cromosoma 18 hi ha un o més gens supressors de tumors claus en la progressió del CCR (Tarafa et al., 2000).

Amplificacions

Com ja s'ha comentat a la introducció, les amplificacions gèniques són característiques de tumors avançats. En el nostre estudi hem trobat un total de sis regions diferents amplificades en almenys un tumor. Cinc de les sis amplificacions es localitzen en regions que rarament es perden en els tumors. Només en el cas del braç 5q, on hem trobat un cas amb amplificació, també hi ha cinc casos que presenten aquest braç alterat en forma de pèrdua. Cal destacar que cap de les amplificacions detectades presenta associacions significatives amb supervivència, el que probablement té relació amb el fet que succeeixen en fases avançades.

Un fet a destacar és la correlació entre el número d'alteracions cromosòmiques i les amplificacions. Per tal d'estudiar aquesta relació vam considerar les amplificacions de forma independent als guanys, i trobem que quan més alteracions cromosòmiques presenta un cas, més probabilitat hi ha que tingui alguna amplificació ($p=0.003$). Aquest fet es pot interpretar de la següent manera: la inestabilitat cromosòmica és un prerequisit per tenir capacitat d'amplificar (Limoli et al., 1997; Paulson et al., 1998). D'altra banda i com ja s'havia descrit anteriorment, trobem que la presència de mutació en el gen TP53 s'associa a la presència d'amplificacions. En el sentit que d'entre els casos que tenen alguna amplificació, la majoria també té el gen TP53 mutat. Per això és fàcil pensar que degut a la capacitat del gen TP53 de vetllar per la integritat del genoma, la seva actuació s'estén al control de la reparació associada als processos d'amplificació. Experiments amb fibroblasts demostren que si introduïm el gen TP53 salvatge en línies que no tenen aquest gen es veu una reducció del nombre d'amplificacions (Yin et al., 1992). Però també s'han descrit gliomes que no tenen TP53 mutat i presenten amplificació del gen MDM2 (Reifenberger et al., 1993). Per tant, la mutació del gen TP53 és un dels camins, però no l'únic. Existeixen altres vies alternatives per permetre els mecanismes d'amplificació.

Seguidament passem a analitzar cadascuna de les amplificacions en detall.

Amplificacions a 20q

La regió més freqüentment amplificada és el braç 20q, en 10 casos (20%). És de destacar que aquesta regió no es troba mai perduda i, ara per ara, el gen candidat amb més números de tenir algun poder de selecció positiu en aquesta

regió és el factor de transcripció E2F1, situat a 20q11.2. El complex DRTF1/E2F1 actua controlant la progressió del cicle cel·lular per passar de la fase G1 a la S. E2F1 està unit a la proteïna RB1 quan és inactiu. Quan es fosforila RB1, s'allibera E2F1 i pot mediar la proliferació cel·lular (vegeu l'apartat d'*Introducció*).

Recentment, però, en un estudi fet amb arrays de CGH, troben que la regió que es guanya amb més freqüència és 20q13.3, en un 81% de casos de CRC esporàdics (Douglas et al., 2004), però no esmenten cap gen candidat. Un altre marcador descrit en aquest braç i situat a 20q11.2-q13.1 és MMP9 (de l'anglès *matrix metalloprotease-9*). Alts nivells d'aquest gen s'han associat a la presència de matàstasi en CCR (Zeng et al., 1996).

És de destacar que si ens fixem en el total de guanys i pèrdues, el braç 20q és la regió que presenta més alteracions en aquesta sèrie i totes en el mateix sentit: hi ha 27 casos que tenen aquest cromosoma guanyat.

Amplificacions a 11q13

L'amplificació de la regió 11q13 és molt comuna en un gran ventall de tumors diferents (mama, esòfag, cap i coll, ovari, bufeta, melanomes, etc.) (Brison, 1993), i en el nostre estudi afecta a set tumors (14%). Són molts els gens que trobem en aquesta regió, des del gen CCND1 (cyclina D1), a factors de creixement com FGF3 (*fibroblast growth factor 3*), FGF4 (*fibroblast growth factor 4*), VEGFB (*VEGFB vascular endothelial growth factor B*), o FOSL1 (*FOS-like antigen 1*). En el cas de càncer colorectal, l'amplificació del gen CCND1 s'ha descrit entre un 3 i un 6% dels casos (Brison, 1993). Després d'estudiar per PCR quantitativa l'amplificació de CCND1 d'aquesta sèrie només es va confirmar l'amplificació del gen en tres casos (que representen un 6% de la sèrie).

La CCND1 és essencial per a la progressió a través de la fase G1, per tant és un marcador de l'avenç del cicle cel·lular i s'expressa en aquelles cèl·lules que estan en proliferació. Les cèl·lules de teixit normal no expressen CCND1. Però en teixit tumoral de càncer colorectal s'ha vist que hi ha una sobreexpressió de la proteïna estudiada mitjançant immunohistoquímica des d'un 10% i fins un 45% dels casos segons diversos autors (McKay et al., 2000).

Per valorar el valor pronòstic d'aquests resultats hi ha molts estudis que comparen la supervivència en funció del grau d'expressió de la proteïna i la seva localització (nucli o citoplasma). Els resultats publicats són molt contradictoris i van des de sèries de casos on sembla que els nivells baixos d'expressió de CCND1 s'associen a un mal pronòstic (Holland et al., 2001), i d'altres que troben que una expressió més elevada s'associa a un pitjor pronòstic (Maeda et al., 1998; McKay et al., 2000). També s'ha descrit l'alteració del braç 11q en les

metàstasis de CCR. Korn estudien 27 metàstasis i 6 tumors primaris sincrònics i troben el guany del braç 11q a dues de les metàstasis i en canvi a cap tumor primari. Ells apunten que per això és un marcador de la progressió a metàstasi. En la nostra sèrie hem trobat amplificat el gen CCND1 en tres pacients. Dos dels quals no han recaigut i un d'ells sí, el 186. Al tractar-se de tant pocs casos és difícil fer conjectures a nivell de l'evolució de la malaltia en funció de la presència d'aquest gen amplificat. En un treball fet amb carcinoma escamós de cap i coll troben que un terç dels casos (21/67) presenten aquest gen amplificat. I els pacients amb aquesta alteració tenen un pronòstic més dolent (Nimeus et al., 2004).

Altrament s'ha de dir que no és necessària l'amplificació a nivell de DNA perquè hi hagi sobreexpressió a nivell d'RNA o proteïna. En una sèrie de 45 casos de carcinoma escamós de cap i coll s'ha descrit que el 22% presenten amplificació del gen CCND1 i el 53% de casos el sobreexpressen (Kyomoto et al., 1997). Així i tot a nivell pronòstic l'amplificació té valor, en canvi la sobreexpressió no en aquesta mateixa sèrie. I tampoc és cert que el fet de trobar el gen amplificat aquest se sobreexpressi (Kanda et al., 1994). Fins i tot s'ha descrit que la sobreexpressió del gen CCND1 afavoreix l'amplificació d'aquest (Zhou et al., 1996). En la present tesi, però, no ens hem centrat en el mecanisme pel qual un gen es pot amplificar.

Ens queda per veure quin és el gen responsable de l'amplificació en els quatre casos on no hem confirmat que sigui la CCND1. Per tant, hi ha d'haver algun altre gen en aquesta regió amb capacitat oncogènica. Si repassem la literatura trobem molts gens que es localitzen en aquesta regió (vegeu l'apartat de la *Introducció*). Per tant, qualsevol dels gens esmentats anteriorment, FGF3, FGF4, VEGFB i FOSL1, podrien estar implicats en el procés tumoral. És per això que creiem necessari posar més èmfasi en aquesta regió, perquè de ben segur hi ha algun oncogèn marcador de CCR important.

Amplificacions a 8q24

El gen MYC es localitza a 8q24. L'amplificació d'aquest gen s'ha descrit en molts tumors (vegeu l'apartat de la *Introducció*). En aquesta sèrie de casos trobem 5 pacients (10%) que presenten amplificació a la regió 8q24, que suggereix la implicació del gen MYC com a responsable d'aquesta. En estudis anteriors s'ha confirmat que MYC estava amplificat entre un 10-15% de tumors primaris de càncer colorectal (Finley et al., 1989; Rochlitz et al., 1996). També s'ha descrit que l'amplificació no es correlaciona amb la sobreexpressió a nivell transcripcional, perquè aquesta última té lloc amb molta més freqüència (70%) (Augenlicht et al., 1997). Tot i que nosaltres no hem confirmat que aquest sigui

el gen amplificat, és de destacar que la freqüència amb la que s'ha trobat l'amplificació de 8q24 és del 10%, igual que en altres sèries de tumors descrites a la literatura. També cal senyalar que els guanys a 8q són molt freqüents, i que per tant és molt probable la presència de més d'un gen amb característiques oncogèniques.

Amplificacions a 5q31-q33, 1p36 i 10q22

Finalment també vam trobar amplificades, en un sol cas cada una, les regions 5q31-q33, 1p36 i 10q22, on els gens candidats implicats en el procés tumoral són FGF1 (*fibroblast growth factor 1*), TP73 (*tumor protein p73*) i FRAT1 (*frequently rearranged in advanced T-cell lymphomas*), respectivament.

FGF1 és una proteïna de la família de factors de creixement que s'uneixen a l'heparina i són agents angiogènics *in vivo*, i que també s'han descrit com a mitògens potents en una gran varietat de cèl·lules *in vitro*. És de destacar la presència de cinc casos on es perd el braç 5q, però a la regió 5q21-q22 on se situa el gen APC, un conegut gen supressor de tumors (vegeu l'apartat de la *Introducció*), cosa que explicaria aquesta tendència a la pèrdua.

Pel què fa a TP73, la seva sobreexpressió s'ha descrit en una gran varietat de tumors (melanoma, neuroblastoma, *hepatocellular*, bufeta, colorectal, pulmó, mama, i pròstata) (Dominguez et al., 2001). La funció exacta d'aquesta proteïna no està del tot establerta, però sembla que podria regular la funció de TP53 inhibint la seva unió al DNA. Per tant la seva sobreexpressió seria un mecanisme d'inactivació de TP53 (Vikhanskaya et al., 2000). Així i tot, els resultats a la literatura són molt contradictoris, perquè també hi ha molts estudis que associen la pèrdua d'expressió de TP73 amb la progressió dels tumors de bufeta (Moll i Petrenko, 2003). Una altra explicació de la relació amb l'avenç dels tumors és que quan tenim TP53 mutada, aquesta s'uneix en forma d'heterodímers defectius a TP73, i per tant provoca de manera indirecta la sobreexpressió d'aquesta. La seva funció com a supressor tumoral quedava parcialment contradita amb l'absència de cap fenotip de càncer en ratolins KO per TP73. Però, sembla ser que hi ha dues isoformes diferents de TP73 (p73 i p63), i que també existeix una mutació dominant negativa que acaba de complicar el panorama (Moll i Petrenko, 2003).

Per altra banda, FRAT1 (de l'anglès, *Frequently rearranged in advanced T-cell lymphomas 1*) és un regulador positiu de la via de WNT mitjançant la inhibició de GSK3B que condueix a l'estabilització de la B-catenina (vegeu l'apartat de la *Introducció*). La inhibició de GSK3B és de vital importància durant el desenvolupament, ja que s'ha descrit que és responsable del destí cel·lular durant l'embrionès de *Drosophila* (Perrimon, 1994).

Consideracions sobre la metodologia de microarrays

En l'estudi d'aquest conjunt de tumors, la segona aproximació ha estat l'anàlisi de la regulació genòmica a nivell transcripcional. Ja des dels inicis de la utilització de la tècnica de les microarrays es va apuntar la seva aplicació en l'estudi del càncer. Són molts els autors que l'han utilitzada per agrupar diferents tipus de càncers (vegeu l'apartat d'*Introducció*).

Abans d'entrar a discutir els nostres resultats creiem convenient fer unes consideracions sobre l'abordatge metodològic a nivell de l'anàlisi de dades. Això és important perquè tant la presentació de les dades com la interpretació dels resultats ve fortament condicionada per la metodologia d'aquest anàlisi. Mitjançant les microarrays generem una gran quantitat de variables que s'han de destriar bé i *manipular* per tal de generar uns resultats assequibles. En estudis d'arrays, la mida mostral és uns quants ordres de magnitud inferior al número de variables (en aquest cas gens). Per tant és impossible des d'un punt de vista matemàtic utilitzar totes les variables per generar predictors de classe, ja que existeix el problema de la *sobrecàrrega*. Això succeeix degut a que si per exemple tenim 10 pacients dels quals analitzem 5000 variables, és molt possible que a l'atzar puguem trobar dos patrons d'expressió diferents que ens identifiquin dos subgrups en funció del color del cabell. Ara bé, aquest patró només té cabuda en el conjunt dels 10 individus pels quals s'ha testat, però no ens permetria discriminar un conjunt de mostres de validació nou. En anglès aquest concepte és anomenat *overfitting*, i es defineix per la capacitat de trobar un patró discriminant a l'atzar en un conjunt de mostres degut a la utilització d'una gran quantitat de variables.

Per tant, per desenvolupar un bon mètode de classificació de les mostres en estudi cal:

- primer, seleccionar un conjunt de gens.
- segon, aplicar els algorismes per definir un nombre de gens més reduït que ens permeti un esquema de classificació acurat.

I també a nivell biològic és més interessant perquè ens basem en la idea que només un nombre reduït de gens és rellevant pel desenvolupament dels tumors. Idealment pocs gens haurien de ser suficients per a classificar els tumors i, a més a més, aquests haurien de tenir més implicacions de cara a aclarir els mecanismes moleculars que duen al seu desenvolupament.

Però a l'hora de decidir quins gens seleccionem ens enfrontem amb dos problemes: el primer, decidir el nombre de gens amb els que treballem; i el segon, quin és el criteri de selecció. No hi ha cap mètode ideal, i a la literatura

cada autor escull una manera en funció de l'objectiu que tingui darrera. Així trobem autors que estableixen un llindar d'intensitat mínim i a partir d'aquí treballen amb tots els gens, altres escullen un llindar de la SD (Tarte et al., 2003), o bé seleccionen els gens en base a una característica determinada i es queden amb els gens significatius per T-test (Glebov et al., 2003). En aquest últim cas cal fer alguna correcció com la de Bonferroni, per no *sobrecarregar* el test.

En la present tesi vam escollir un total de 128 gens que tenen una SD de la mitjana de la ràtio pels 50 casos major que 0,8, límit que s'ha establert de forma arbitrària. Com que un dels objectius era la definició de grups de tumors amb perfils d'expressió diferents, la variable de la SD ens permet escollir aquells gens que presentin una major variabilitat entre tots els casos i que conseqüentment seran els més adients a l'hora de trobar grups de comportament diferent. Si estudiem les correlacions entre aquests 128 gens trobem que 1108 parelles de gens (de les 8128 parelles possibles) estan correlacionades de forma significativa (utilitzant la correcció d'Stoey i Tibshirani) (Stoey i Tibshirani, 2003).

Quan mirem amb més detall aquests gens correlacionats observem xarxes de gens que es comporten de la mateixa manera. L'última conseqüència d'aquest tipus d'estudis és la possibilitat de construir i fer un model de la xarxa de regulacions transcripcionals completa d'un organisme. Però la nostra aproximació no és apropiada per aquest últim objectiu, ja que partim d'una representació dels RNA missatgers limitada, tant pel nombre de gens representats a la microarray (uns 4600), com per la selecció posterior per la RAP-PCR. En qualsevol cas, sí que és possible a nivell exploratori utilitzar aquestes dades per obrir noves vies d'estudi i veure quins gens formen part de la mateixa via o de diferents vies que es coregulen de la mateixa manera. Per exemple, el grup de cinc gens que presenta una correlació superior a 0,9 (UBL1, ZNF32, IGF2R, EDG1, CST3) no formen part de la mateixa via funcional, però es podria aprofundir en el seu estudi per comprovar en quins processos estan actuant i veure com es coregulen conjuntament.

Un cop seleccionats els gens es poden escollir infinitat de mètodes per agrupar les mostres. No volem fer aquí un recull de totes les possibilitats però sí apuntar unes quantes aproximacions com poden ser els agrupaments jeràrquics, els mètodes de SVM (de l'anglès, *support vector machines*), o els mètodes d'anàlisi de components principals (PCA). Nosaltres hem utilitzat els anàlisis de components principals donat que són adients per a integrar tota la informació que ens donen les hibridacions sobre microarrays. S'han utilitzat per a una gran varietat d'estudis: des de diferenciar tumors limfàtics (Armstrong et al., 2002), a

adenocarcinomes pulmonars (Wikman et al., 2002), línies cel·lulars de càncer de mama (Hilsenbeck et al., 1999), etc. És una tècnica que ens ajuda a trobar patrons dins d'una matriu de dades de gran dimensió. L'objectiu és comprimir les dades de manera que reduïm les dimensions i ens quedem amb la informació que explica un elevat percentatge de la variabilitat. Perquè els gens correlacionats són informació redundant, i el que fem és rebaixar la dimensió de variables sense perdre informació de manera significativa. És una aproximació descriptiva que intenta facilitar la interpretació de les dades perquè així en sigui més pràctica. Aquestes noves variables les anomenem *metagens*, tal i com va fer Huang (Huang et al., 2003). Cada metagen és un *eix factorial*, o també anomenat *component principal*, i ens dóna informació nova que ens serà útil per classificar els tumors. Cada metagen és independent de la resta, i estan ordenats de manera que el primer explica la màxima quantitat de variància, el segon una mica menys, i així successivament de forma decreixent.

La dificultat d'aquestes anàlisis rau en el fet que la interpretació del significat dels metagens no és directa. Cal estudiar més a fons tots els gens que formen part dels metagens d'interès. Per exemple, hem vist que alguns metagens correlacionen amb la supervivència lliure de malaltia (vegeu l'apartat de *Resultats*).

Un cop determinats aquests marcadors d'expressió (gens) com a importants a l'hora de predir l'evolució dels tumors, el següent pas seria dissenyar una estratègia perquè fos aplicable de manera senzilla a la clínica diària, com podria ser la PCR quantitativa o mitjançant una microarray reduïda, que aquests gens tenen un comportament diferent en funció de l'agressivitat del tumor; i a continuació es podria fer un *test* encaminat a estudiar l'estat d'expressió d'aquest conjunt de gens per decidir si aquell pacient cal tractar-lo amb agents quimioteràpics o no, independentment de l'estadi de Dukes (B o C) de la biòpsia. Aquest test podria ser una plataforma similar a la utilitzada per nosaltres, però sense la necessitat de que hi hagi tants gens impresos a l'array. En qualsevol cas, la gran limitació d'aquest tipus d'anàlisi és la impossibilitat de comparar estudis independents, donat que manca una plataforma estàndard i per tant els metagens que cada laboratori selecciona són diferents.

Classificació dels tumors en base al patró d'expressió

A partir de les dades d'expressió hem pogut definir quatre grups. Dos dels quals estan formats per molt pocs casos i per tant és difícil d'interpretar la seva significació. Els altres dos estan formats per 22 i 18 casos respectivament, i representen el 80% de tots els tumors de còlon analitzats en aquest estudi. En qualsevol cas, els grups presenten una supervivència lliure de malaltia diferent i

per tant es pot hipotetitzar que representen grups amb característiques i curs clínic diferenciats. En conseqüència sembla adient proposar la realització d'un estudi independent i en una sèrie més àmplia per tal de confirmar las tendències observades i la possible utilitat d'aquest tipus d'anàlisis. Estudis duts a terme per altres investigadors amb dissenys semblants al nostre arriben a resultats en la mateixa línia. Per exemple, en un estudi de càncer gàstric (Tay et al., 2003), distingeixen un grup de tumors amb pitjor supervivència, després d'analitzar els resultats de la hibridació sobre microarrays de 13K i 18K. Altres autors han trobat diferents patrons d'expressió en càncer colorectal específics de tumors del còlon dret i l'esquerre (Glebov et al., 2003). També és de destacar un estudi d'expressió de CCR fet per Karin Birkenkamp-Demtroder (Birkenkamp-Demtroder et al., 2002), on analitzant els resultats de les hibridacions sobre microarrays mitjançant clusters jeràrquics distingeixen tres grups: el primer, format per teixits normals i casos en estadi de Dukes A; el segon, format per casos en estadi de Dukes B i C; i un tercer, format per casos en estadi de Dukes D. Per tant, sembla que els estadis de Dukes B i C presenten un gran solapament a nivell molecular.

Cribratge de les dades: els gens que despunten

Alhora, gràcies al cribratge global de l'expressió de molts gens mitjançant les hibridacions sobre microarrays hem identificat alguns possibles gens supressors de tumors i oncogens quan comparem la ràtio mitjana dels 50 tumors respecte el *pool* de normals (POOLN). Tots ells estan probablement implicats en el procés tumoral, i seria necessari confirmar el seu estat sobreexpressat o infraexpressat mitjançant alguna altra tècnica, així com valorar la seva implicació funcional en el procés de formació del tumor. Tot i així, ens és possible hipotetitzar el mecanisme pel qual poden ser importants.

Un dels gens que apareix infraexpressat en molts tumors és el **TCF7L2** (de l'anglès, *transcription factor 7-like 2*). Conegut anteriorment amb el nom de TCF4, és membre de la via de senyalització de WNT (vegeu l'apartat de la *Introducció*). Aquest gen codifica per un factor de transcripció que s'uneix a β -catenina i activa la transcripció de gens diana. Per tant, a priori, la seva repressió farà que el tàndem β -catenina/TCF sigui inactiu i s'aturi la proliferació. Ara bé, estudis funcionals indiquen que també pot actuar com a inhibidor endogen d'aquesta via en alguns contextos, ja que una de les seves isoformes s'uneix a β -catenina però no ho fa al DNA (Douglas et al., 2001). D'aquesta manera disminueix la capacitat de β -catenina d'activar els gens diana. És per tant un tema complex que s'hauria d'estudiar amb més detall.

Pel què fa als gens sobreexpressats, és de destacar que les diferències són més petites. Un dels primers de la llista (Taula 44 de l'apartat de *Resultats*) és el gen **ITGB1** (de l'anglès, *integrin beta 1*). Es tracta d'un gen que codifica per una proteïna de la família de les integrines beta, que tenen un paper cabdal en la interacció entre la matriu extracel·lular i el citoesquelet. Les integrines estan formades per heterodímers d'una subunitat alfa i una beta, formant un domini extracel·lular, un domini transmembrana, i una cua citoplasmàtica curta (Hynes i Lander, 1992). Pot formar dímers amb diversos subtipus de cadenes alfa, i així reconèixer a una gran varietat de molècules extracel·lulars (col·lagen, fibronectina, laminina, etc.). S'han descrit dues isoformes diferents de cadena beta: la A i la B. S'ha vist que la isoforma B interfereix amb la A fent que actuï com a un dominant negatiu. D'aquesta manera s'inhibeix l'adhesió cel·lular i la migració *in vitro* (Balzac et al., 1994). Altre cop cal fer l'advertència que estem mirant expressió de l'RNA i que això no vol dir funció.

Integració de les dades de CGH i arrays

Els resultats obtinguts per arrays i CGH són difícils d'integrar directament, ja que la resolució d'ambdues tècniques és diferent: mitjançant la hibridació sobre cromosomes obtenim una resolució de 10 Mb, i en canvi amb els arrays la resolució depèn de la distància a la que es troben en els cromosomes els gens impresos al xip HU4.6K. En algunes zones dels cromosomes serà menor i en d'altres major, segons la densitat de gens que s'hagin escollit. Però en aquest estudi no hem buscat la correlació directa entre els gens guanyats i sobreexpressats perquè no era el nostre objectiu i representa un anàlisi de difícil interpretació. Ens hem centrat en la classificació dels tumors.

Hem trobat que al classificar les 50 mostres mitjançant agrupaments jeràrquics no supervisats en funció dels resultats de la CGH i dels 128 gens, ens generen grups diferents. Per tant podem pensar que la informació generada amb cada una de les tècniques és independent, i que estem mirant coses diferents. Per això podem deduir, que el fet de no tenir alteracions a nivell cromosòmic no determina tenir una desregulació del transcriptoma característica, perquè els 8 casos sense alteracions es reparteixen en els quatre grups definits per arrays. I tampoc observem que unes característiques específiques a nivell de CGH s'associïn a unes determinades alteracions transcripcionals.

Com ja apunta Tay en un estudi de tumors gàstrics (Tay et al., 2003), és difícil correlacionar les alteracions cromosòmiques estudiades per CGH i l'expressió dels gens per arrays. Mirant les amplificacions respecte la sobreexpressió de gens d'aquella regió no troben cap correlació. Ells apunten al problema que la resolució de la CGH és baixa, però pot molt ben ser que utilitzant arrays de CGH de major

resolució tampoc es pugui trobar una bona correlació, perquè com ja s'ha dit hi ha molts altres mecanismes a part de l'amplificació per incrementar l'expressió d'un gen. En aquest treball no aprofundeixen més en els resultats de la CGH, més enllà d'un mer caràcter descriptiu, sense entrar a fer classificacions.

Arribats al final, després de classificar els tumors en base als seus patrons d'alteracions cromosòmiques i d'expressió, el que queda clar és que és possible distingir subgrups de tumors amb perfils moleculars característics i que s'associen a un curs clínic diferenciat. Cal destacar que la informació que obtenim en l'anàlisi molecular és independent de les característiques clíniques i que per tant pot ser d'utilitat a l'hora d'avaluar millor el pronòstic. No obstant, també és evident la necessitat d'estudiar un conjunt de mostres més gran per poder buscar nous factors de pronòstic dins de cada grup independentment.

VII. CONCLUSIONS

- 1.** El 18% dels tumors NO presenten alteracions a nivell cromosòmic estudiades mitjançant la CGH. La resta (82%) tenen entre 1 i 20 alteracions per tumor. I de mitjana trobem $8,8 \pm 6,3$ alteracions per tumor.
- 2.** Les alteracions cromosòmiques més freqüents són +20q (74%), -18q (66%), -17p (48%), +13 (44%) i -18p (40%).
- 3.** La mutació del gen TP53 s'associa a un número més alt d'alteracions cromosòmiques, al contrari que la inestabilitat de microsatèl·lits (MSI).
- 4.** Les amplificacions gèniques es donen en un context de moltes alteracions cromosòmiques.
- 5.** A l'hora de classificar els tumors segons el seu perfil d'alteracions cromosòmiques ens hem decantat per un criteris modificats de Dutrillaux. Definim així quatre grups: el que presenta inestabilitat de microsatèl·lits (6%), el grup Normal sense alteracions cromosòmiques (16%), el grup d'Inestabilitat Cromosòmica Monosomic Like (44%), i el grup d'Inestabilitat cromosòmica Restant (34%).
- 6.** La classificació dels tumors en funció del patró d'alteracions cromosòmiques és poc útil com a factor pronòstic directament, però sí que permet identificar factors de pronòstic específics de grup. Per exemple, el nombre d'alteracions cromosòmiques és un factor de pronòstic important dins del grup ICML.
- 7.** El guany del braç 11q i la pèrdua del braç 12p són indicadors de mal pronòstic en càncer colorectal.
- 8.** La caracterització de les aparents amplificacions observades per CGH a 11q13 suggereix que la diana d'amplificació no és la ciclina D1 en la majoria de casos.
- 9.** A partir de l'expressió de 128 gens que presenten una major variabilitat entre les 50 mostres hem pogut identificar un grup de gens que se sobreexpressen i un altre que s'infraexpressen.
- 10.** Hem identificat quatre grups a partir de les dades d'expressió. Dos d'ells formats per molts pocs casos (4 i 6 respectivament). I els dos restants amb un pronòstic oposat. Un d'ells, format per 22 casos, té una supervivència lliure de malaltia bona, i l'altre, format per 18 casos, dolenta.
- 11.** La informació que obtenim a partir de la CGH i dels arrays no és equivalent, ja que els grups que podem definir utilitzant ambdues tècniques no coincideixen. Es tracta per tant d'informació complementària.

VIII. BIBLIOGRAFIA

A

Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., and et al. (1993). Clues to the pathogenesis of familial colorectal cancer. *Science* *260*, 812-816.

Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., *et al.* (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* *403*, 503-511.

Al-Mulla, F., Keith, W. N., Pickford, I. R., Going, J. J., and Birnie, G. D. (1999). Comparative genomic hybridization analysis of primary colorectal carcinomas and their synchronous metastases. *Genes Chromosomes Cancer* *24*, 306-314.

Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J. (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* *30*, 41-47.

Arribas, R., Ribas, M., Risques, R. A., Masramon, L., Tortola, S., Marcuello, E., Aiza, G., Miro, R., Capella, G., and Peinado, M. A. (1999). Prospective assessment of allelic losses at 4p14-16 in colorectal cancer: two mutational patterns and a locus associated with poorer survival. *Clin Cancer Res* *5*, 3454-3459.

Askew, D. S., Ashmun, R. A., Simmons, B. C., and Cleveland, J. L. (1991). Constitutive c-myc expression in an IL-3-dependent myeloid cell line suppresses cell cycle arrest and accelerates apoptosis. *Oncogene* *6*, 1915-1922.

Astler, V. B., and Collier, F. A. (1954). The prognostic significance of direct extension of carcinoma of the colon and rectum. *Ann Surg* *139*, 846-852.

Augenlicht, L. H., Wadler, S., Corner, G., Richards, C., Ryan, L., Multani, A. S., Pathak, S., Benson, A., Haller, D., and Heerdt, B. G. (1997). Low-level c-myc amplification in human colonic carcinoma cell lines and tumors: a frequent, p53-independent mutation associated with improved outcome in a randomized multi-institutional trial. *Cancer Res* *57*, 1769-1775.

B

Baisse, B., Bouzourene, H., Saraga, E. P., Bosman, F. T., and Benhattar, J. (2001). Intratumor genetic heterogeneity in advanced human colorectal adenocarcinoma. *Int J Cancer* *93*, 346-352.

Baker, S. J., Preisinger, A. C., Jessup, J. M., Paraskeva, C., Markowitz, S., Willson, J. K., Hamilton, S., and Vogelstein, B. (1990). p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res* *50*, 7717-7722.

Balzac, F., Retta, S. F., Albin, A., Melchiorri, A., Koteliansky, V. E., Geuna, M., Silengo, L., and Tarone, G. (1994). Expression of beta 1B integrin isoform in CHO cells results in a dominant negative effect on cell adhesion and motility. *J Cell Biol* *127*, 557-565.

Bardelli, A., Cahill, D. P., Lederer, G., Speicher, M. R., Kinzler, K. W., Vogelstein, B., and Lengauer, C. (2001). Carcinogen-specific induction of genetic instability. *Proc Natl Acad Sci U S A* *98*, 5770-5775.

- Bardi, G., Fenger, C., Johansson, B., Mitelman, F., and Heim, S. (2004). Tumor karyotype predicts clinical outcome in colorectal cancer patients. *J Clin Oncol* 22, 2623-2634.
- Bassam, B. J., Caetano-Anolles, G., and Gresshoff, P. M. (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* 196, 80-83.
- Baudino, T. A., McKay, C., Pendeville-Samain, H., Nilsson, J. A., Maclean, K. H., White, E. L., Davis, A. C., Ihle, J. N., and Cleveland, J. L. (2002). c-Myc is essential for vasculogenesis and angiogenesis during development and tumor progression. *Genes Dev* 16, 2530-2543.
- Bertucci, F., Salas, S., Eysteris, S., Nasser, V., Finetti, P., Ginestier, C., Charafe-Jauffret, E., Lloriod, B., Bachelart, L., Montfort, J., *et al.* (2004). Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene* 23, 1377-1391.
- Betticher, D. C., Heighway, J., Hasleton, P. S., Altermatt, H. J., Ryder, W. D., Cerny, T., and Thatcher, N. (1996). Prognostic significance of CCND1 (cyclin D1) overexpression in primary resected non-small-cell lung cancer. *Br J Cancer* 73, 294-300.
- Bhattacharyya, N. P., Skandalis, A., Ganesh, A., Groden, J., and Meuth, M. (1994). Mutator phenotypes in human colorectal carcinoma cell lines. *Proc Natl Acad Sci U S A* 91, 6319-6323.
- Bingham, S. A. (2000). Diet and colorectal cancer prevention. *Biochem Soc Trans* 28, 12-16.
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., *et al.* (2000). Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 406, 536-540.
- Bodmer, W., Bishop, T., and Karran, P. (1994). Genetic steps in colorectal cancer. *Nat Genet* 6, 217-219.
- Boland, C. R., Sato, J., Appelman, H. D., Bresalier, R. S., and Feinberg, A. P. (1995). Microallelotyping defines the sequence and tempo of allelic losses at tumour suppressor gene loci during colorectal cancer progression. *Nat Med* 1, 902-909.
- Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Rodriguez-Bigas, M. A., Fodde, R., Ranzani, G. N., and Srivastava, S. (1998). A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* 58, 5248-5257.
- Bostick, R. M., Fosdick, L., Wood, J. R., Grambsch, P., Grandits, G. A., Lillemoe, T. J., Louis, T. A., and Potter, J. D. (1995). Calcium and colorectal epithelial cell proliferation in sporadic adenoma patients: a randomized, double-blinded, placebo-controlled clinical trial. *J Natl Cancer Inst* 87, 1307-1315.
- Boveri, T. (1914). *Zur Frage der Entstehung maligner Tumoren* (Jena, Fischer Verlag).
- Brison, O. (1993). Gene amplification and tumor progression. *Biochim Biophys Acta* 1155, 25-41.

C

Cahill, D. P., Kinzler, K. W., Vogelstein, B., and Lengauer, C. (1999). Genetic instability and darwinian selection in tumours. *Trends Cell Biol* 9, M57-60.

Cahill, D. P., Lengauer, C., Yu, J., Riggins, G. J., Willson, J. K., Markowitz, S. D., Kinzler, K. W., and Vogelstein, B. (1998). Mutations of mitotic checkpoint genes in human cancers. *Nature* 392, 300-303.

Castells, A., Ino, Y., Louis, D. N., Ramesh, V., Gusella, J. F., and Rustgi, A. K. (1999). Mapping of a target region of allelic loss to a 0.5-cM interval on chromosome 22q13 in human colorectal cancer. *Gastroenterology* 117, 831-837.

Chan, T. L., Curtis, L. C., Leung, S. Y., Farrington, S. M., Ho, J. W., Chan, A. S., Lam, P. W., Tse, C. W., Dunlop, M. G., Wyllie, A. H., and Yuen, S. T. (2001). Early-onset colorectal cancer with stable microsatellite DNA and near-diploid chromosomes. *Oncogene* 20, 4871-4876.

Chibon, F., Mariani, O., Mairal, A., Derre, J., Coindre, J. M., Terrier, P., Lagace, R., Sastre, X., and Aurias, A. (2003). The use of clustering software for the classification of comparative genomic hybridization data. an analysis of 109 malignant fibrous histiocytomas. *Cancer Genet Cytogenet* 141, 75-78.

Choi, S. W., Lee, K. J., Bae, Y. A., Min, K. O., Kwon, M. S., Kim, K. M., and Rhyu, M. G. (2002). Genetic classification of colorectal cancer based on chromosomal loss and microsatellite instability predicts survival. *Clin Cancer Res* 8, 2311-2322.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162, 156-159.

Ciaparrone, M., Yamamoto, H., Yao, Y., Sgambato, A., Cattoretti, G., Tomita, N., Monden, T., Rotterdam, H., and Weinstein, I. B. (1998). Localization and expression of p27KIP1 in multistage colorectal carcinogenesis. *Cancer Res* 58, 114-122.

Citarda, F., Tomaselli, G., Capocaccia, R., Barcherini, S., and Crespi, M. (2001). Efficacy in standard clinical practice of colonoscopic polypectomy in reducing colorectal cancer incidence. *Gut* 48, 812-815.

D

da Costa, L. T., Liu, B., el-Deiry, W., Hamilton, S. R., Kinzler, K. W., Vogelstein, B., Markowitz, S., Willson, J. K., de la Chapelle, A., Downey, K. M., and et al. (1995). Polymerase delta variants in RER colorectal tumours. *Nat Genet* 9, 10-11.

De Angelis, P. M., Clausen, O. P., Schjolberg, A., and Stokke, T. (1999). Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes. *Br J Cancer* 80, 526-535.

Deans, G. T., Parks, T. G., Rowlands, B. J., and Spence, R. A. (1992). Prognostic factors in colorectal cancer. *Br J Surg* 79, 608-613.

Di Vinci, A., Infusini, E., Peveri, C., Sciutto, A., Orecchia, R., Geido, E., Monaco, R., and Giaretti, W. (1999). Intratumor heterogeneity of chromosome 1, 7, 17, and 18 aneusomies obtained by FISH and association with flow cytometric DNA index in human colorectal adenocarcinomas. *Cytometry* 35, 369-375.

Diep, C. B., Thorstensen, L., Meling, G. I., Skovlund, E., Rognum, T. O., and Lothe, R. A. (2003). Genetic tumor markers with prognostic impact in Dukes' stages B and C colorectal cancer patients. *J Clin Oncol* 21, 820-829.

Dominguez, G., Silva, J., Silva, J. M., Garcia, J. M., Larrondo, F. J., Vargas, J., Sanfrutos, L., Provencio, M., Espana, P., and Bonilla, F. (2001). Different expression of P14ARF defines two groups of breast carcinomas in terms of TP73 expression and TP53 mutational status. *Genes Chromosomes Cancer* 31, 99-106.

Douglas, E. J., Fiegler, H., Rowan, A., Halford, S., Bicknell, D. C., Bodmer, W., Tomlinson, I. P., and Carter, N. P. (2004). Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. *Cancer Res* 64, 4817-4825.

Douglas, K. R., Brinkmeier, M. L., Kennell, J. A., Eswara, P., Harrison, T. A., Patrianakos, A. I., Sprecher, B. S., Potok, M. A., Lyons, R. H., Jr., MacDougald, O. A., and Camper, S. A. (2001). Identification of members of the Wnt signaling pathway in the embryonic pituitary gland. *Mamm Genome* 12, 843-851.

Duesberg, P. H. (2003). Are cancers dependent on oncogenes or on aneuploidy? *Cancer Genet Cytogenet* 143, 89-91.

Dukes (1932). The classification of cancer of the rectum. *The Journal of Pathology* 35, 323.

Dutrillaux, B. (1995). Pathways of chromosome alteration in human epithelial cancers. *Adv Cancer Res* 67, 59-82.

E

Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* 300, 455.

Eshleman, J. R., Casey, G., Kochera, M. E., Sedwick, W. D., Swinler, S. E., Veigl, M. L., Willson, J. K., Schwartz, S., and Markowitz, S. D. (1998). Chromosome number and structure both are markedly stable in RER colorectal cancers and are not destabilized by mutation of p53. *Oncogene* 17, 719-725.

Eshleman, J. R., Lang, E. Z., Bowerfind, G. K., Parsons, R., Vogelstein, B., Willson, J. K., Veigl, M. L., Sedwick, W. D., and Markowitz, S. D. (1995). Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer. *Oncogene* 10, 33-37.

Evan, G. I., Wyllie, A. H., Gilbert, C. S., Littlewood, T. D., Land, H., Brooks, M., Waters, C. M., Penn, L. Z., and Hancock, D. C. (1992). Induction of apoptosis in fibroblasts by c-myc protein. *Cell* 69, 119-128.

F

Fabarius, A., Hehlmann, R., and Duesberg, P. H. (2003). Instability of chromosome structure in cancer cells increases exponentially with degrees of aneuploidy. *Cancer Genet Cytogenet* 143, 59-72.

Fearon, E. R., and Jones, P. A. (1992). Progressing toward a molecular description of colorectal cancer development. *Faseb J* 6, 2783-2790.

Fearon, E. R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* 61, 759-767.

Feinberg, A. P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301, 89-92.

Fels, L. M., Buschmann, T., Meuer, J., Reymond, M. A., Lamer, S., Rocken, C., and Ebert, M. P. (2003). Proteome analysis for the identification of tumor-associated biomarkers in gastrointestinal cancer. *Dig Dis* 21, 292-298.

Fiegler, H., Carr, P., Douglas, E. J., Burford, D. C., Hunt, S., Scott, C. E., Smith, J., Vetrie, D., Gorman, P., Tomlinson, I. P., and Carter, N. P. (2003). DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer* 36, 361-374.

Finley, G. G., Schulz, N. T., Hill, S. A., Geiser, J. R., Pipas, J. M., and Meisler, A. I. (1989). Expression of the myc gene family in different stages of human colorectal cancer. *Oncogene* 4, 963-971.

Flyger, H. L., Larsen, J. K., Nielsen, H. J., and Christensen, I. J. (1999). DNA ploidy in colorectal cancer, heterogeneity within and between tumors and relation to survival. *Cytometry* 38, 293-300.

Fritz, B., Schubert, F., Wrobel, G., Schwaenen, C., Wessendorf, S., Nessling, M., Korz, C., Rieker, R. J., Montgomery, K., Kucherlapati, R., *et al.* (2002). Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. *Cancer Res* 62, 2993-2998.

G

Georgiades, I. B., Curtis, L. J., Morris, R. M., Bird, C. C., and Wyllie, A. H. (1999). Heterogeneity studies identify a subset of sporadic colorectal cancers without evidence for chromosomal or microsatellite instability. *Oncogene* 18, 7933-7940.

Giaccia, A. J., and Kastan, M. B. (1998). The complexity of p53 modulation: emerging patterns from divergent signals. *Genes Dev* 12, 2973-2983.

Giovannucci, E., Stampfer, M. J., Colditz, G. A., Rimm, E. B., Trichopoulos, D., Rosner, B. A., Speizer, F. E., and Willett, W. C. (1993). Folate, methionine, and alcohol intake and risk of colorectal adenoma. *J Natl Cancer Inst* 85, 875-884.

Giovannucci, E., and Willett, W. C. (1994). Dietary factors and risk of colon cancer. *Ann Med* 26, 443-452.

Glebov, O. K., Rodriguez, L. M., Nakahara, K., Jenkins, J., Cliatt, J., Humbyrd, C. J., DeNobile, J., Soballe, P., Simon, R., Wright, G., *et al.* (2003). Distinguishing right from left colon by the pattern of gene expression. *Cancer Epidemiol Biomarkers Prev* 12, 755-762.

Goel, A., Arnold, C. N., Niedzwiecki, D., Chang, D. K., Ricciardiello, L., Carethers, J. M., Dowell, J. M., Wasserman, L., Compton, C., Mayer, R. J., *et al.* (2003). Characterization of sporadic colon cancer by patterns of genomic instability. *Cancer Res* 63, 1608-1614.

Gonzalez-Garcia, I., Moreno, V., Navarro, M., Marti-Rague, J., Marcuello, E., Benasco, C., Campos, O., Capella, G., and Peinado, M. A. (2000). Standardized approach for microsatellite instability detection in colorectal carcinomas. *J Natl Cancer Inst* 92, 544-549.

Grady, W. M., and Markowitz, S. D. (2002). Genetic and epigenetic alterations in colon cancer. *Annu Rev Genomics Hum Genet* 3, 101-128.

Grau, M., Sole, X., Obrador, A., Tarafa, G., Vendrell, E., Valls, J., Moreno, V., Peinado, M. A., and Capella, G. (2005). Validation of RNA arbitrarily primed PCR probes hybridized to glass cDNA microarrays: application to the analysis of limited samples. *Clin Chem* 51, 93-101.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M., and et al. (1991). Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* 66, 589-600.

Gryfe, R., Kim, H., Hsieh, E. T., Aronson, M. D., Holowaty, E. J., Bull, S. B., Redston, M., and Gallinger, S. (2000). Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N Engl J Med* 342, 69-77.

Gumbiner, B. M. (1995). Signal transduction of beta-catenin. *Curr Opin Cell Biol* 7, 634-640.

H

Halling, K. C., French, A. J., McDonnell, S. K., Burgart, L. J., Schaid, D. J., Peterson, B. J., Moon-Tasson, L., Mahoney, M. R., Sargent, D. J., O'Connell, M. J., et al. (1999). Microsatellite instability and 8p allelic imbalance in stage B2 and C colorectal cancers. *J Natl Cancer Inst* 91, 1295-1303.

Heikkila, R., Schwab, G., Wickstrom, E., Loke, S. L., Pluznik, D. H., Watt, R., and Neckers, L. M. (1987). A c-myc antisense oligodeoxynucleotide inhibits entry into S phase but not progress from G0 to G1. *Nature* 328, 445-449.

Hellman, S. (1997). Darwin's clinical relevance. *Cancer* 79, 2275-2281.

Hemminki, A., Mecklin, J. P., Jarvinen, H., Aaltonen, L. A., and Joensuu, H. (2000). Microsatellite instability is a favorable prognostic indicator in patients with colorectal cancer receiving chemotherapy. *Gastroenterology* 119, 921-928.

Herman, J. G., Umar, A., Polyak, K., Graff, J. R., Ahuja, N., Issa, J. P., Markowitz, S., Willson, J. K., Hamilton, S. R., Kinzler, K. W., et al. (1998). Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A* 95, 6870-6875.

Heselmeyer, K., Schrock, E., du Manoir, S., Blegen, H., Shah, K., Steinbeck, R., Auer, G., and Ried, T. (1996). Gain of chromosome 3q defines the transition from severe dysplasia to invasive carcinoma of the uterine cervix. *Proc Natl Acad Sci U S A* 93, 479-484.

Hilsenbeck, S. G., Friedrichs, W. E., Schiff, R., O'Connell, P., Hansen, R. K., Osborne, C. K., and Fuqua, S. A. (1999). Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J Natl Cancer Inst* 91, 453-459.

Holland, T. A., Elder, J., McCloud, J. M., Hall, C., Deakin, M., Fryer, A. A., Elder, J. B., and Hoban, P. R. (2001). Subcellular localisation of cyclin D1 protein in colorectal tumours is associated with p21(WAF1/CIP1) expression and correlates with patient survival. *Int J Cancer* 95, 302-306.

Howe, G. R., Jain, M., and Miller, A. B. (1990). Dietary factors and risk of pancreatic cancer: results of a Canadian population-based case-control study. *Int J Cancer* 45, 604-608.

Huang, E., Ishida, S., Pittman, J., Dressman, H., Bild, A., Kloos, M., D'Amico, M., Pestell, R. G., West, M., and Nevins, J. R. (2003). Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat Genet* 34, 226-230.

Huang, J., Papadopoulos, N., McKinley, A. J., Farrington, S. M., Curtis, L. J., Wyllie, A. H., Zheng, S., Willson, J. K., Markowitz, S. D., Morin, P., *et al.* (1996). APC mutations in colorectal tumors with mismatch repair deficiency. *Proc Natl Acad Sci U S A* 93, 9049-9054.

Hutter, R. V., and Sobin, L. H. (1986). A universal staging system for cancer of the colon and rectum. Let there be light. *Arch Pathol Lab Med* 110, 367-368.

Hynes, R. O., and Lander, A. D. (1992). Contact and adhesive specificities in the associations, migrations, and targeting of cells and axons. *Cell* 68, 303-322.

I

Iacobuzio-Donahue, C. A., Maitra, A., Olsen, M., Lowe, A. W., van Heek, N. T., Rosty, C., Walter, K., Sato, N., Parker, A., Ashfaq, R., *et al.* (2003). Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays. *Am J Pathol* 162, 1151-1162.

Ichii, S., Horii, A., Nakatsuru, S., Furuyama, J., Utsunomiya, J., and Nakamura, Y. (1992). Inactivation of both APC alleles in an early stage of colon adenomas in a patient with familial adenomatous polyposis (FAP). *Hum Mol Genet* 1, 387-390.

Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 363, 558-561.

Iwabuchi, H., Sakamoto, M., Sakunaga, H., Ma, Y. Y., Carcangiu, M. L., Pinkel, D., Yang-Feng, T. L., and Gray, J. W. (1995). Genetic analysis of benign, low-grade, and high-grade ovarian tumors. *Cancer Res* 55, 6172-6180.

J

Jallepalli, P. V., and Lengauer, C. (2001). Chromosome segregation and cancer: cutting through the mystery. *Nat Rev Cancer* 1, 109-117.

Jones, A. M., Douglas, E. J., Halford, S. E., Fiegler, H., Gorman, P. A., Roylance, R. R., Carter, N. P., and Tomlinson, I. P. (2005). Array-CGH analysis of microsatellite-stable, near-diploid bowel cancers and comparison with other types of colorectal carcinoma. *Oncogene* 24, 118-129.

Jones, P. A., and Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 3, 415-428.

K

Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W., and Waldman, F. M. (1994). Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci U S A* 91, 2156-2160.

Kallioniemi, O. P., Kallioniemi, A., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., and Pinkel, D. (1993). Comparative genomic hybridization: a rapid new method for detecting and mapping DNA amplification in tumors. *Semin Cancer Biol* 4, 41-46.

Kanda, Y., Nishiyama, Y., Shimada, Y., Imamura, M., Nomura, H., Hiai, H., and Fukumoto, M. (1994). Analysis of gene amplification and overexpression in human esophageal-carcinoma cell lines. *Int J Cancer* *58*, 291-297.

Kikuchi-Yanoshita, R., Konishi, M., Ito, S., Seki, M., Tanaka, K., Maeda, Y., Iino, H., Fukayama, M., Koike, M., Mori, T., and et al. (1992). Genetic changes of both p53 alleles associated with the conversion from colorectal adenoma to early carcinoma in familial adenomatous polyposis and non-familial adenomatous polyposis patients. *Cancer Res* *52*, 3965-3971.

Kinzler, K. W., and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell* *87*, 159-170.

Kitahara, O., Furukawa, Y., Tanaka, T., Kihara, C., Ono, K., Yanagawa, R., Nita, M. E., Takagi, T., Nakamura, Y., and Tsunoda, T. (2001). Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer Res* *61*, 3544-3549.

Kleibeuker, J. H., Mulder, N. H., Cats, A., van der Meer, R., and de Vries, E. G. (1996). Calcium and colorectal epithelial cell proliferation. *Gut* *39*, 774-775.

Knudson, A. G. (1993). Antioncogenes and human cancer. *Proc Natl Acad Sci U S A* *90*, 10914-10921.

Koehler, A., Bataille, F., Schmid, C., Ruemmele, P., Waldeck, A., Blaszyk, H., Hartmann, A., Hofstaedter, F., and Dietmaier, W. (2004). Gene expression profiling of colorectal cancer and metastases divides tumours according to their clinicopathological stage. *J Pathol* *204*, 65-74.

Korn, W. M., Yasutake, T., Kuo, W. L., Warren, R. S., Collins, C., Tomita, M., Gray, J., and Waldman, F. M. (1999). Chromosome arm 20q gains and other genomic alterations in colorectal cancer metastatic to liver, as analyzed by comparative genomic hybridization and fluorescence in situ hybridization. *Genes Chromosomes Cancer* *25*, 82-90.

Kronborg, O., Fenger, C., Olsen, J., Jorgensen, O. D., and Sondergaard, O. (1996). Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet* *348*, 1467-1471.

Kyomoto, R., Kumazawa, H., Toda, Y., Sakaida, N., Okamura, A., Iwanaga, M., Shintaku, M., Yamashita, T., Hiai, H., and Fukumoto, M. (1997). Cyclin-D1-gene amplification is a more potent prognostic factor than its protein over-expression in human head-and-neck squamous-cell carcinoma. *Int J Cancer* *74*, 576-581.

L

Lanza, G., Matteuzzi, M., Gafa, R., Orvieto, E., Maestri, I., Santini, A., and del Senno, L. (1998). Chromosome 18q allelic loss and prognosis in stage II and III colon cancer. *Int J Cancer* *79*, 390-395.

Lee, J. K., Bussey, K. J., Gwadry, F. G., Reinhold, W., Riddick, G., Pelletier, S. L., Nishizuka, S., Szakacs, G., Annereau, J. P., Shankavaram, U., et al. (2003). Comparing cDNA and oligonucleotide array data: concordance of gene expression across platforms for the NCI-60 cancer cells. *Genome Biol* *4*, R82.

Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1997). Genetic instability in colorectal cancers. *Nature* *386*, 623-627.

Leppert, M., Dobbs, M., Scambler, P., O'Connell, P., Nakamura, Y., Stauffer, D., Woodward, S., Burt, R., Hughes, J., Gardner, E., and et al. (1987). The gene for familial polyposis coli maps to the long arm of chromosome 5. *Science* *238*, 1411-1413.

Leslie, A., Pratt, N. R., Gillespie, K., Sales, M., Kernohan, N. M., Smith, G., Wolf, C. R., Carey, F. A., and Steele, R. J. (2003). Mutations of APC, K-ras, and p53 are associated with specific chromosomal aberrations in colorectal adenocarcinomas. *Cancer Res* *63*, 4656-4661.

Levine, A. J. (1997). p53, the cellular gatekeeper for growth and division. *Cell* *88*, 323-331.

Levy, D. B., Smith, K. J., Beazer-Barclay, Y., Hamilton, S. R., Vogelstein, B., and Kinzler, K. W. (1994). Inactivation of both APC alleles in human and mouse tumors. *Cancer Res* *54*, 5953-5958.

Lijovic, M., and Frauman, A. G. (2003). Toward an understanding of the molecular genetics of prostate cancer progression. *J Environ Pathol Toxicol Oncol* *22*, 1-15.

Limoli, C. L., Kaplan, M. I., Corcoran, J., Meyers, M., Boothman, D. A., and Morgan, W. F. (1997). Chromosomal instability and its relationship to other end points of genomic instability. *Cancer Res* *57*, 5557-5563.

Lloyd, R. V., Erickson, L. A., Jin, L., Kulig, E., Qian, X., Chevillat, J. C., and Scheithauer, B. W. (1999). p27kip1: a multifunctional cyclin-dependent kinase inhibitor with prognostic significance in human cancers. *Am J Pathol* *154*, 313-323.

Loda, M., Cukor, B., Tam, S. W., Lavin, P., Fiorentino, M., Draetta, G. F., Jessup, J. M., and Pagano, M. (1997). Increased proteasome-dependent degradation of the cyclin-dependent kinase inhibitor p27 in aggressive colorectal carcinomas. *Nat Med* *3*, 231-234.

Loeb, K. R., and Loeb, L. A. (1999). Genetic instability and the mutator phenotype. *Studies in ulcerative colitis. Am J Pathol* *154*, 1621-1626.

Loeb, L. A. (1991). Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res* *51*, 3075-3079.

Loeb, L. A., and Christians, F. C. (1996). Multiple mutations in human cancers. *Mutat Res* *350*, 279-286.

Loukola, A., Eklin, K., Laiho, P., Salovaara, R., Kristo, P., Jarvinen, H., Mecklin, J. P., Launonen, V., and Aaltonen, L. A. (2001). Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (hnpcc). *Cancer Res* *61*, 4545-4549.

Lynch, H. T., Smyrk, T., and Lynch, J. F. (1996). Overview of natural history, pathology, molecular genetics and management of HNPCC (Lynch Syndrome). *Int J Cancer* *69*, 38-43.

M

Maeda, K., Chung, Y., Kang, S., Ogawa, M., Onoda, N., Nishiguchi, Y., Ikehara, T., Nakata, B., Okuno, M., and Sowa, M. (1998). Cyclin D1 overexpression and prognosis in colorectal adenocarcinoma. *Oncology* *55*, 145-151.

Malumbres, M., and Barbacid, M. (2003). RAS oncogenes: the first 30 years. *Nat Rev Cancer* 3, 459-465.

Mandel, J. S., Church, T. R., Ederer, F., and Bond, J. H. (1999). Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. *J Natl Cancer Inst* 91, 434-437.

Markowitz, S. D., Dawson, D. M., Willis, J., and Willson, J. K. (2002). Focus on colon cancer. *Cancer Cell* 1, 233-236.

Martinez-Lopez, E., Abad, A., Font, A., Monzo, M., Ojanguren, I., Pifarre, A., Sanchez, J. J., Martin, C., and Rosell, R. (1998). Allelic loss on chromosome 18q as a prognostic marker in stage II colorectal cancer. *Gastroenterology* 114, 1180-1187.

McKay, J. A., Douglas, J. J., Ross, V. G., Curran, S., Murray, G. I., Cassidy, J., and McLeod, H. L. (2000). Cyclin D1 protein expression and gene polymorphism in colorectal cancer. Aberdeen Colorectal Initiative. *Int J Cancer* 88, 77-81.

Mitelman, F., Johansson, B., Mandahl, N., and Mertens, F. (1997). Clinical significance of cytogenetic findings in solid tumors. *Cancer Genet Cytogenet* 95, 1-8.

Moll, U. M., and Petrenko, O. (2003). The MDM2-p53 interaction. *Mol Cancer Res* 1, 1001-1008.

Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., and Kinzler, K. W. (1997). Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* 275, 1787-1790.

Muleris, M., Salmon, R. J., and Dutrillaux, B. (1988). Existence of two distinct processes of chromosomal evolution in near-diploid colorectal tumors. *Cancer Genet Cytogenet* 32, 43-50.

Muleris, M., Salmon, R. J., and Dutrillaux, B. (1990). Cytogenetics of colorectal adenocarcinomas. *Cancer Genet Cytogenet* 46, 143-156.

N

Nagafuchi, A., Ishihara, S., and Tsukita, S. (1994). The roles of catenins in the cadherin-mediated cell adhesion: functional analysis of E-cadherin-alpha catenin fusion molecules. *J Cell Biol* 127, 235-245.

Nakamura, S., Kino, I., and Baba, S. (1993). Nuclear DNA content of isolated crypts of background colonic mucosa from patients with familial adenomatous polyposis and sporadic colorectal cancer. *Gut* 34, 1240-1244.

Nakao, K., Mehta, K. R., Fridlyand, J., Moore, D. H., Jain, A. N., Lafuente, A., Wiencke, J. W., Terdiman, J. P., and Waldman, F. M. (2004). High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis* 25, 1345-1357.

Newcomb, P. A., Norfleet, R. G., Storer, B. E., Surawicz, T. S., and Marcus, P. M. (1992). Screening sigmoidoscopy and colorectal cancer mortality. *J Natl Cancer Inst* 84, 1572-1575.

Nimeus, E., Baldetorp, B., Bendahl, P. O., Rennstam, K., Wennerberg, J., Akervall, J., and Ferno, M. (2004). Amplification of the cyclin D1 gene is associated with tumour subsite, DNA non-diploidy and high S-phase fraction in squamous cell carcinoma of the head and neck. *Oral Oncol* 40, 624-629.

Nishisho, I., Nakamura, Y., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., Koyama, K., Utsunomiya, J., Baba, S., and Hedge, P. (1991). Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* 253, 665-669.

Nowell, P. C. (1986). Mechanisms of tumor progression. *Cancer Res* 46, 2203-2207.

O

Ohue, M., Tomita, N., Monden, T., Fujita, M., Fukunaga, M., Takami, K., Yana, I., Ohnishi, T., Enomoto, T., Inoue, M., and et al. (1994). A frequent alteration of p53 gene in carcinoma in adenoma of colon. *Cancer Res* 54, 4798-4804.

Olivier, M., Eeles, R., Hollstein, M., Khan, M. A., Harris, C. C., and Hainaut, P. (2002). The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum Mutat* 19, 607-614.

Ookawa, K., Sakamoto, M., Hirohashi, S., Yoshida, Y., Sugimura, T., Terada, M., and Yokota, J. (1993). Concordant p53 and DCC alterations and allelic losses on chromosomes 13q and 14q associated with liver metastases of colorectal carcinoma. *Int J Cancer* 53, 382-387.

P

Paulson, T. G., Almasan, A., Brody, L. L., and Wahl, G. M. (1998). Gene amplification in a p53-deficient cell line requires cell cycle progression under conditions that generate DNA breakage. *Mol Cell Biol* 18, 3089-3100.

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747-752.

Perrimon, N. (1994). Signalling pathways initiated by receptor protein tyrosine kinases in *Drosophila*. *Curr Opin Cell Biol* 6, 260-266.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., et al. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20, 207-211.

Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999). Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 23, 41-46.

Potten, C. S., and Loeffler, M. (1990). Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. *Development* 110, 1001-1020.

Potter, J. D. (1999). Colorectal cancer: molecules and populations. *J Natl Cancer Inst* 91, 916-932.

R

Reifenberger, G., Liu, L., Ichimura, K., Schmidt, E. E., and Collins, V. P. (1993). Amplification and overexpression of the MDM2 gene in a subset of human malignant gliomas without p53 mutations. *Cancer Res* 53, 2736-2739.

Remvikos, Y., Muleris, M., Vielh, P., Salmon, R. J., and Dutrillaux, B. (1988). DNA content and genetic evolution of human colorectal adenocarcinoma. A study by flow cytometry and cytogenetic analysis. *Int J Cancer* *42*, 539-543.

Remvikos, Y., Vogt, N., Muleris, M., Salmon, R. J., Malfoy, B., and Dutrillaux, B. (1995). DNA-repeat instability is associated with colorectal cancers presenting minimal chromosome rearrangements. *Genes Chromosomes Cancer* *12*, 272-276.

Ribas, M., Masramon, L., Aiza, G., Capella, G., Miro, R., and Peinado, M. A. (2003). The structural nature of chromosomal instability in colon cancer cells. *Faseb J* *17*, 289-291.

Ried, T., Heselmeyer-Haddad, K., Blegen, H., Schrock, E., and Auer, G. (1999). Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. *Genes Chromosomes Cancer* *25*, 195-204.

Ried, T., Knutzen, R., Steinbeck, R., Blegen, H., Schrock, E., Heselmeyer, K., du Manoir, S., and Auer, G. (1996). Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* *15*, 234-245.

Risques, R. A., Moreno, V., Marcuello, E., Petriz, J., Cancelas, J. A., Sancho, F. J., Torregrosa, A., Capella, G., and Peinado, M. A. (2001). Redefining the significance of aneuploidy in the prognostic assessment of colorectal cancer. *Lab Invest* *81*, 307-315.

Risques, R. A., Moreno, V., Ribas, M., Marcuello, E., Capella, G., and Peinado, M. A. (2003). Genetic pathways and genome-wide determinants of clinical outcome in colorectal cancer. *Cancer Res* *63*, 7206-7214.

Rochlitz, C. F., Herrmann, R., and de Kant, E. (1996). Overexpression and amplification of c-myc during progression of human colorectal cancer. *Oncology* *53*, 448-454.

S

Sakakura, C., Hagiwara, A., Yasuoka, R., Fujita, Y., Nakanishi, M., Masuda, K., Shimomura, K., Nakamura, Y., Inazawa, J., Abe, T., and Yamagishi, H. (2001). Tumour-amplified kinase BTAK is amplified and overexpressed in gastric cancers with possible involvement in aneuploid formation. *Br J Cancer* *84*, 824-831.

Saltz, L. B., Cox, J. V., Blanke, C., Rosen, L. S., Fehrenbacher, L., Moore, M. J., Maroun, J. A., Ackland, S. P., Locker, P. K., Pirotta, N., *et al.* (2000). Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. *N Engl J Med* *343*, 905-914.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning. A laboratory manual.* Cold Spring Harbor Laboratory Press.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* *270*, 467-470.

Schlegel, J., Stumm, G., Scherthan, H., Bocker, T., Zirngibl, H., Ruschoff, J., and Hofstadter, F. (1995). Comparative genomic in situ hybridization of colon carcinomas with replication error. *Cancer Res* *55*, 6002-6005.

- Schoen, R. E. (2002). The case for population-based screening for colorectal cancer. *Nat Rev Cancer* 2, 65-70.
- Schwab, M. (1998). Amplification of oncogenes in human cancer cells. *Bioessays* 20, 473-479.
- Selaru, F. M., Xu, Y., Yin, J., Zou, T., Liu, T. C., Mori, Y., Abraham, J. M., Sato, F., Wang, S., Twigg, C., *et al.* (2002). Artificial neural networks distinguish among subtypes of neoplastic colorectal lesions. *Gastroenterology* 122, 606-613.
- Sheiness, D., Fanshier, L., and Bishop, J. M. (1978). Identification of nucleotide sequences which may encode the oncogenic capacity of avian retrovirus MC29. *J Virol* 28, 600-610.
- Shibata, D., Peinado, M. A., Ionov, Y., Malkhosyan, S., and Perucho, M. (1994). Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nat Genet* 6, 273-281.
- Shih, I. M., Zhou, W., Goodman, S. N., Lengauer, C., Kinzler, K. W., and Vogelstein, B. (2001). Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Res* 61, 818-822.
- Simon, R., Radmacher, M. D., Dobbin, K., and McShane, L. M. (2003). Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 95, 14-18.
- Skibber, J. (1999). Operative intervention for primary tumors in stage IV colorectal cancer. *Ann Surg Oncol* 6, 629-630.
- Smith, A. J., Stern, H. S., Penner, M., Hay, K., Mitri, A., Bapat, B. V., and Gallinger, S. (1994). Somatic APC and K-ras codon 12 mutations in aberrant crypt foci from human colons. *Cancer Res* 54, 5527-5530.
- Smith, G., Carey, F. A., Beattie, J., Wilkie, M. J., Lightfoot, T. J., Coxhead, J., Garner, R. C., Steele, R. J., and Wolf, C. R. (2002). Mutations in APC, Kirsten-ras, and p53--alternative genetic pathways to colorectal cancer. *Proc Natl Acad Sci U S A* 99, 9433-9438.
- Soussi, T., and Beroud, C. (2001). Assessing TP53 status in human tumours to evaluate clinical outcome. *Nat Rev Cancer* 1, 233-240.
- Sparks, A. B., Morin, P. J., Vogelstein, B., and Kinzler, K. W. (1998). Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. *Cancer Res* 58, 1130-1134.
- Spencer, F., Gerring, S. L., Connelly, C., and Hieter, P. (1990). Mitotic chromosome transmission fidelity mutants in *Saccharomyces cerevisiae*. *Genetics* 124, 237-249.
- Steinert, R., Buschmann, T., van der Linden, M., Fels, L. M., Lippert, H., and Reymond, M. A. (2002). The role of proteomics in the diagnosis and outcome prediction in colorectal cancer. *Technol Cancer Res Treat* 1, 297-304.
- Stocks, P. (1957). Methodology of epidemiological studies of cancer. *Acta Unio Int Contra Cancrum* 13, 86-89.
- Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100, 9440-9445.
- Sugai, T., Takahashi, H., Habano, W., Nakamura, S., Sato, K., Orii, S., and Suzuki, K. (2003). Analysis of genetic alterations, classified according to their

DNA ploidy pattern, in the progression of colorectal adenomas and early colorectal carcinomas. *J Pathol* 200, 168-176.

T

Takayama, T., Ohi, M., Hayashi, T., Miyanishi, K., Nobuoka, A., Nakajima, T., Satoh, T., Takimoto, R., Kato, J., Sakamaki, S., and Niitsu, Y. (2001). Analysis of K-ras, APC, and beta-catenin in aberrant crypt foci in sporadic adenoma, cancer, and familial adenomatous polyposis. *Gastroenterology* 121, 599-611.

Tarafa, G., Villanueva, A., Farre, L., Rodriguez, J., Musulen, E., Reyes, G., Seminago, R., Olmedo, E., Paules, A. B., Peinado, M. A., *et al.* (2000). DCC and SMAD4 alterations in human colorectal and pancreatic tumor dissemination. *Oncogene* 19, 546-555.

Tarte, K., Zhan, F., De Vos, J., Klein, B., and Shaughnessy, J., Jr. (2003). Gene expression profiling of plasma cells and plasmablasts: toward a better understanding of the late stages of B-cell differentiation. *Blood* 102, 592-600.

Tay, S. T., Leong, S. H., Yu, K., Aggarwal, A., Tan, S. Y., Lee, C. H., Wong, K., Visvanathan, J., Lim, D., Wong, W. K., *et al.* (2003). A combined comparative genomic hybridization and expression microarray analysis of gastric cancer reveals novel molecular subtypes. *Cancer Res* 63, 3309-3316.

Teh, B. T., Larsson, C., and Nordenskjold, M. (1999). Tumor suppressor genes (TSG). *Anticancer Res* 19, 4715-4728.

Thibodeau, S. N., Bren, G., and Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science* 260, 816-819.

Tischfield, J. A. (1997). Loss of heterozygosity or: how I learned to stop worrying and love mitotic recombination. *Am J Hum Genet* 61, 995-999.

Tomlinson, I., and Bodmer, W. (1999). Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nat Med* 5, 11-12.

Tomlinson, I., Halford, S., Aaltonen, L., Hawkins, N., and Ward, R. (2002a). Does MSI-low exist? *J Pathol* 197, 6-13.

Tomlinson, I., Sasieni, P., and Bodmer, W. (2002b). How many mutations in cancer? *Am J Pathol* 160, 755-758.

Toribara, N. W., and Sleisenger, M. H. (1995). Screening for colorectal cancer. *N Engl J Med* 332, 861-867.

V

van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A. P., *et al.* (2002). The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 111, 241-250.

Van Gelder, R. N., von Zastrow, M. E., Yool, A., Dement, W. C., Barchas, J. D., and Eberwine, J. H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci U S A* 87, 1663-1667.

Vikhanskaya, F., D'Incalci, M., and Broggin, M. (2000). p73 competes with p53 and attenuates its response in a human ovarian cancer cell line. *Nucleic Acids Res* 28, 513-519.

Visakorpi, T., Kallioniemi, A. H., Syvanen, A. C., Hyytinen, E. R., Karhu, R., Tammela, T., Isola, J. J., and Kallioniemi, O. P. (1995). Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. *Cancer Res* *55*, 342-347.

Vogelstein, B., and Kinzler, K. W. (1992). Carcinogens leave fingerprints. *Nature* *355*, 209-210.

W

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* *58*, 236-244.

Watanabe, T., Wu, T. T., Catalano, P. J., Ueki, T., Satriano, R., Haller, D. G., Benson, A. B., 3rd, and Hamilton, S. R. (2001). Molecular predictors of survival after adjuvant chemotherapy for colon cancer. *N Engl J Med* *344*, 1196-1206.

Welsh, J., Chada, K., Dalal, S. S., Cheng, R., Ralph, D., and McClelland, M. (1992). Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res* *20*, 4965-4970.

Welsh, J., and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* *18*, 7213-7218.

Wielenga, V. J., Smits, R., Korinek, V., Smit, L., Kielman, M., Fodde, R., Clevers, H., and Pals, S. T. (1999). Expression of CD44 in Apc and Tcf mutant mice implies regulation by the WNT pathway. *Am J Pathol* *154*, 515-523.

Wikman, H., Kettunen, E., Seppanen, J. K., Karjalainen, A., Hollmen, J., Anttila, S., and Knuutila, S. (2002). Identification of differentially expressed genes in pulmonary adenocarcinoma by using cDNA array. *Oncogene* *21*, 5804-5813.

Willett, W. C. (2001). Diet and cancer: one view at the start of the millennium. *Cancer Epidemiol Biomarkers Prev* *10*, 3-8.

Williams, A. C., Miller, J. C., Collard, T., Browne, S. J., Newbold, R. F., and Paraskeva, C. (1997). The effect of different TP53 mutations on the chromosomal stability of a human colonic adenoma derived cell line with endogenous wild type TP53 activity, before and after DNA damage. *Genes Chromosomes Cancer* *20*, 44-52.

Y

Yin, Y., Tainsky, M. A., Bischoff, F. Z., Strong, L. C., and Wahl, G. M. (1992). Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles. *Cell* *70*, 937-948.

Z

Zeng, Z. S., Huang, Y., Cohen, A. M., and Guillem, J. G. (1996). Prediction of colorectal cancer relapse and survival via tissue RNA levels of matrix metalloproteinase-9. *J Clin Oncol* *14*, 3133-3140.

Zhao, H., Hastie, T., Whitfield, M. L., Borresen-Dale, A. L., and Jeffrey, S. S. (2002). Optimization and evaluation of T7 based RNA linear amplification protocols for cDNA microarray analysis. *BMC Genomics* *3*, 31.

Zhou, P., Jiang, W., Weghorst, C. M., and Weinstein, I. B. (1996). Overexpression of cyclin D1 enhances gene amplification. *Cancer Res* *56*, 36-39.

Zhou, W., Goodman, S. N., Galizia, G., Lieto, E., Ferraraccio, F., Pignatelli, C., Purdie, C. A., Piris, J., Morris, R., Harrison, D. J., *et al.* (2002). Counting alleles to predict recurrence of early-stage colorectal cancers. *Lancet* 359, 219-225.

Zou, T. T., Selaru, F. M., Xu, Y., Shustova, V., Yin, J., Mori, Y., Shibata, D., Sato, F., Wang, S., Olaru, A., *et al.* (2002). Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. *Oncogene* 21, 4855-4862.

IX. ANNEXOS

Annex I: Informació complementària

Taula 48. Nom i localització dels 128 gens amb una SD major de 0,8 seleccionats per a fer els estudis d'expressió.

acc	símbol del gen	cromosoma	mitjana	nom del gen
H12279	PGM5	9p12-q12	-3.423	<i>phosphoglucomutase 5</i>
W07367	FUBP3	9q34.2	-2.888	<i>far upstream element (FUSE) binding protein 3</i>
W52186	SYNPO2	4	-2.197	<i>synaptopodin 2</i>
N76867	TCF7L2	10q25.3	-1.529	<i>transcription factor 7-like 2 (T-cell specific, HMG-box)</i>
AA429297	RNF103	2p11.2	-1.352	<i>ring finger protein 103</i>
AA402960	RNF5	6p21.3	-1.225	<i>ring finger protein 5</i>
H17504	MAPK6	15q21	-1.105	<i>mitogen-activated protein kinase 6</i>
H94262	EPB41L4B	9q31-q32	-1.090	<i>erythrocyte membrane protein band 4.1 like 4B</i>
N50247	MTMR6	13q12	-1.087	<i>myotubularin related protein 6</i>
AA504656	LTBP1	2p22-p21	-1.077	<i>latent transforming growth factor beta binding protein 1</i>
N73030	C5	9q33-q34	-1.049	<i>complement component 5</i>
N57872	AGXT	2q36-q37	-0.964	<i>alanine-glyoxylate aminotransferase homolog</i>
AA598572	SYK	9	-0.948	<i>spleen tyrosine kinase</i>
R11586	REV1L	2q11.1-q11.2	-0.902	<i>REV1-like (yeast)</i>
W86653	FKBP5	6p31.3-p21.2	-0.895	<i>FK506-binding protein 5</i>
R80041	FPRL1	19q13.3-q13.4	-0.861	<i>formyl peptide receptor-like 1</i>
T77595	TNC	9q33	-0.798	<i>tenascin C (hexabrachion)</i>
AA446108	ENG	9q33-q34.1	-0.719	<i>endoglin (Osler-Rendu-Weber syndrome 1)</i>

R19889	DAZAP2	2q12	-0.707	<i>DAZ associated protein 2</i>
W21081	LOC286512	Xp11.22	-0.703	<i>similar to Eukaryotic initiation factor 4A-II (eIF4A-II)</i>
AA598868	COPB	11p15.2	-0.688	<i>coatamer protein complex, subunit beta</i>
H62396	FANCC	9q22.3	-0.678	<i>Fanconi anemia, complementation group C</i>
T84382	WDR31	9q32	-0.643	<i>WD repeat domain 31</i>
R83836	LYN	8q13	-0.635	<i>v-yes-1 Yamaguchi sarcoma viral related oncogene homolog</i>
AA454784	IL16	15q26.3	-0.605	<i>interleukin 16 (lymphocyte chemoattractant factor)</i>
AA011215	SAT	Xp22.1	-0.591	<i>spermidine/spermine N1-acetyltransferase</i>
R93715	MARCH-VI	5p15.2	-0.570	<i>membrane-associated RING-CH protein VI</i>
AA495901	CBLN1	16q12.1	-0.565	<i>cerebellin 1 precursor</i>
AA486919	RPL28	19q13.4	-0.562	<i>ribosomal protein L28</i>
AA401111	GPI	19q13.3	-0.560	<i>glucose phosphate isomerase</i>
N52568	NR6A1	9q33-q34.1	-0.550	<i>nuclear receptor subfamily 6, group A, member 1</i>
AA464711	C3AR1	12p13.31	-0.532	<i>complement component 3a receptor 1</i>
H51066	LEPROT	1p31.2	-0.511	<i>leptin receptor overlapping transcript</i>
AA401137	LCN2	9q34	-0.448	<i>lipocalin 2 (oncogene 24p3)</i>
N67039	CDK6	7q21-q22	-0.437	<i>cyclin-dependent kinase 6</i>
T77023	RB1CC1	8p22-q21.13	-0.428	<i>RB1-inducible coiled-coil 1</i>
AA041499	LRBA	4q31.23-q31.3	-0.416	<i>LPS-responsive vesicle trafficking, beach and anchor containing</i>
R11490	TPR	1q25	-0.415	<i>translocated promoter region (to activated MET oncogene)</i>
R26526	BNC1	15q25.2	-0.402	<i>basonuclin 1</i>
T70413	XIST	Xq13.2	-0.398	<i>X (inactive)-specific transcript</i>
H69620	ARHGAP4	Xq28	-0.393	<i>Rho GTPase activating protein 4</i>
AA278840	TBC1D5	3p24.3	-0.360	<i>TBC1 domain family, member 5</i>

N99243	TBX2	17q23	-0.342	<i>T-box 2</i>
H00817	LYPLA1	8q11.23	-0.333	<i>lysophospholipase I</i>
N91290	N91290		-0.324	<i>EST</i>
H64260	PRKAG2	7q35-q36	-0.308	<i>protein kinase, AMP-activated, gamma 2 non-catalytic subunit</i>
AA284495	MESDC2	15q13	-0.262	<i>mesoderm development candidate 2</i>
R88904	R88904		-0.234	<i>EST</i>
R62773	R62773		-0.232	<i>EST</i>
H65042	NPR2	9p21-p12	-0.223	<i>natriuretic peptide receptor B/guanylate cyclase B (atriuretic peptide receptor B)</i>
AA424747	HOXD9	2q31.1	-0.189	<i>homeo box D9</i>
H66542	GHRL	3p26-p25	-0.180	<i>ghrelin precursor</i>
AA025333	MGC15407	2p16.1	-0.179	<i>similar to RIKEN cDNA 4931428D14 gene</i>
AA456271	RBM4	11q13	-0.115	<i>RNA binding motif protein 4</i>
H53658	PAI-RBP1	1p31-p22	-0.113	<i>PAI-1 mRNA-binding protein</i>
R67042	KIAA0265	7q32.2	-0.090	<i>KIAA0265 protein</i>
AA599173	MYST3	8p11	-0.085	<i>MYST histone acetyltransferase (monocytic leukemia) 3</i>
N54932	RPL30	8q22	-0.016	<i>ribosomal protein L30</i>
H50582	CASP8AP2	6q15	0.009	<i>CASP8 associated protein 2</i>
AA136910	ACVR1	2q23-q24	0.012	<i>activin A receptor, type I</i>
AA598561	CD164	6q21	0.049	<i>CD164 antigen, sialomucin</i>
R21785	NUP155	5p13.1	0.060	<i>nucleoporin 155kD</i>
AA196465	SLN	11q22-q23	0.072	<i>sarcolipin</i>
R98436	INADL	1p31.3	0.113	<i>PDZ domain protein (Drosophila inaD-like)</i>
AA437064	THBS4	5q13	0.114	<i>thrombospondin 4</i>
R93875	NAP1L1	12q21.2	0.121	<i>nucleosome assembly protein 1-like 1</i>

H16637	VCAM1	1p32-p31	0.121	<i>vascular cell adhesion molecule 1</i>
H60549	CD59	11p13	0.126	<i>CD59 antigen p18-20 (antigen identified by monoclonal antibodies 16.3A5, EJ16, EJ30, EL32 and G344)</i>
R00395	XTP2	1q23.3	0.130	<i>HBxAg transactivated protein 2</i>
H96241	ALB	4q11-q13	0.133	<i>albumin</i>
AA478959	PARP4	13	0.135	<i>poly (ADP-ribose) polymerase family, member 4</i>
AA443982	PPP1CA	11q13	0.137	<i>protein phosphatase 1, catalytic subunit, alpha isoform</i>
R63548	ZNF287	17p11.2	0.143	<i>zinc finger protein 287</i>
N78103	N78103	3	0.147	<i>ESTs</i>
AA424937	G6PD	Xq28	0.160	<i>glucose-6-phosphate dehydrogenase</i>
R70541	WBSCR14	7q11.23	0.163	<i>Williams-Beuren syndrome chromosome region 14</i>
H37989	TUBB	6p21.33	0.172	<i>tubulin, beta polypeptide</i>
H22652	GMFB	14q22.2	0.178	<i>glia maturation factor, beta</i>
AA428518	DDX1	2p24	0.186	<i>DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 1</i>
N66942	PGRMC1	Xq22-q24	0.187	<i>progesterone receptor membrane component 1</i>
AA446682	GNL2	1p34.3	0.201	<i>guanine nucleotide binding protein-like 2 (nucleolar)</i>
AA054073	CEACAM6	19q13.2	0.206	<i>carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)</i>
T57556	HINT1	5q31.2	0.222	<i>histidine triad nucleotide-binding protein</i>
R71913	PSMA2	7p14.1	0.227	<i>proteasome (prosome, macropain) subunit, alpha type, 2</i>
T95804	C20orf14	20q13.33	0.250	<i>chromosome 20 open reading frame 14</i>
AA180046	KPNA1	3q21	0.266	<i>karyopherin alpha 1 (importin alpha 5)</i>
T62547	IGF2R	6q26	0.281	<i>insulin-like growth factor 2 receptor</i>
H09997	MMP16	8q21	0.303	<i>matrix metalloproteinase 16 (membrane-inserted)</i>
R33154	R33154		0.357	<i>ESTs</i>
H81220	ELF1	13q13	0.361	<i>E74-like factor 1 (ets domain transcription factor)</i>

AA281137	USP6NL	10p13	0.362	<i>USP6 N-terminal like</i>
AA485626	AHCY	20cen-q13.1	0.373	<i>S-adenosylhomocysteine hydrolase</i>
N69204	CSE1L	20q13	0.410	<i>CSE1 chromosome segregation 1-like (yeast)</i>
AA463610	ITGA2	5q23-q31	0.417	<i>integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)</i>
H48420	PTMA	2q35-q36	0.421	<i>prothymosin, alpha (gene sequence 28)</i>
AA504348	TOP2A	17q21-q22	0.429	<i>topoisomerase (DNA) II alpha (170kD)</i>
H60549	CD59	11	0.430	<i>CD59 antigen p18-20 (antigen identified by monoclonal antibodies 16.3A5, EJ16, EJ30, EL32 and G344)</i>
AA412053	CD9	12p13.3	0.436	<i>CD9 antigen (p24)</i>
AA599177	CST3	20p11.21	0.471	<i>cystatin C (amyloid angiopathy and cerebral hemorrhage)</i>
AA453175	BIN1	2q14	0.473	<i>bridging integrator 1</i>
H57850	PPP2R1B	11q23.2	0.478	<i>protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), beta soform</i>
AA488626	SUMO1	2q33	0.487	<i>SMT3 suppressor of mif two 3 homolog 1 (yeast)</i>
R13546	EDG1	1p21	0.495	<i>endothelial differentiation, sphingolipid G-protein-coupled receptor, 1</i>
H61209	HARS	5q31.3	0.505	<i>histidyl-tRNA synthetase</i>
N54596	IGF2	11p15.5	0.510	<i>insulin-like growth factor 2 (somatomedin A)</i>
AA443177	SRP72	4q11	0.554	<i>signal recognition particle 72kD</i>
H25546	SAA1	11p15.1	0.556	<i>serum amyloid A1</i>
N27227	PCOLN3	16q24.3	0.566	<i>procollagen (type III) N-endopeptidase</i>
AA463544	MCP	1q32	0.582	<i>membrane cofactor protein (CD46, trophoblast-lymphocyte cross-reactive antigen)</i>
T47229	ZNF32	10q22-q25	0.594	<i>zinc finger protein 32 (KOX 30)</i>
AA458994	EXOSC9	4q27	0.622	<i>exosome component 9</i>
AA424807	p44S10	3p14.1	0.624	<i>proteasome regulatory particle subunit p44S10</i>
AA459519	LAMA5	20q13.2-q13.3	0.624	<i>laminin, alpha 5</i>
T75041	PLP1	Xq22	0.651	<i>proteolipid protein (Pelizaeus-Merzbacher disease, spastic paraplegia 2, uncomplicated)</i>

H05768	ATP6V1C1	8	0.668	<i>ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C, isoform 1</i>
AA460756	JARID1A	12	0.690	<i>Jumonji, AT rich interactive domain 1A (RBBP2-like)</i>
H96850	H96850		0.717	
W67174	ITGB1	19	0.726	<i>integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)</i>
AA406601	ABLIM1	10q25	0.744	<i>actin binding LIM protein 1</i>
AA258396	PHLDA1	12q15	0.745	<i>pleckstrin homology-like domain, family A, member 1</i>
AA441895	GSTO1	10q25.1	0.789	<i>glutathione S-transferase omega 1</i>
AA488609	NUP88	17p13.2	0.815	<i>nucleoporin 88kD</i>
AA083228	CAPZA2	7q31.2-q31.3	0.869	<i>capping protein (actin filament) muscle Z-line, alpha 2</i>
AA278749	NCBP1	9q34.1	0.892	<i>nuclear cap binding protein subunit 1, 80kD</i>
AA487912	GNB1	1p36.33	0.947	<i>guanine nucleotide binding protein (G protein), beta polypeptide 1</i>
AA434487	NAB2	12q13.3-q14.1	0.969	<i>NGFI-A binding protein 2 (ERG1 binding protein 2)</i>
H20652	ARL6IP	16p12-p11.2	1.033	<i>ADP-ribosylation factor-like 6 interacting protein</i>
H73329	TPX2	20q11.2	1.383	<i>TPX2, microtubule-associated protein homolog (Xenopus laevis)</i>

Taula 49. Gens amb una correlació major de 0,85.

	ZNF32	IGF2R	EDG1	CST3	SUMO1	LYN	EXOSC9	N91290	R62773	PHLDA1	NUP88	LRBA	TBX2	C3AR1	IL16	SRP72	AHCY	CD59	PPP1CA	PSMA2	TNC	ATP6V1C1	PTMA
ZNF32	1.00	0.96	0.91	0.93	0.95	-0.15	-0.07	-0.18	-0.21	-0.16	-0.16	-0.11	-0.09	0.11	0.17	-0.07	-0.08	-0.06	-0.14	-0.09	0.08	-0.15	-0.13
IGF2R	0.96	1.00	0.92	0.92	0.92	-0.21	-0.15	-0.24	-0.23	-0.28	-0.25	-0.19	-0.17	0.10	0.15	-0.02	-0.01	0.01	-0.05	-0.04	0.04	-0.10	-0.07
EDG1	0.91	0.92	1.00	0.96	0.91	-0.06	0.04	-0.13	-0.20	-0.13	-0.14	-0.12	-0.13	0.13	0.11	-0.07	-0.05	-0.04	-0.12	-0.13	0.06	-0.13	-0.08
CST3	0.93	0.92	0.96	1.00	0.92	-0.05	0.00	-0.19	-0.24	-0.13	-0.14	-0.13	-0.14	0.11	0.13	-0.10	-0.08	-0.05	-0.18	-0.15	0.07	-0.17	-0.15
SUMO1	0.95	0.92	0.91	0.92	1.00	-0.02	0.02	-0.14	-0.20	-0.09	-0.12	-0.04	-0.08	0.20	0.21	-0.06	-0.06	-0.07	-0.15	-0.08	0.11	-0.16	-0.15
LYN	-0.15	-0.21	-0.06	-0.05	-0.02	1.00	0.90	0.17	0.20	0.25	0.25	0.12	0.06	0.24	0.07	0.03	0.14	0.05	0.07	0.05	-0.07	0.16	0.15
EXOSC9	-0.07	-0.15	0.04	0.00	0.02	0.90	1.00	0.18	0.29	0.16	0.22	0.03	0.05	0.23	0.11	0.11	0.11	0.03	0.10	0.00	-0.11	0.19	0.18
N91290	-0.18	-0.24	-0.13	-0.19	-0.14	0.17	0.18	1.00	0.91	0.35	0.34	0.35	0.38	0.06	0.00	0.07	0.12	0.08	0.14	0.02	0.00	0.18	0.26
R62773	-0.21	-0.23	-0.20	-0.24	-0.20	0.20	0.29	0.91	1.00	0.20	0.26	0.31	0.40	0.05	0.03	0.21	0.17	0.19	0.27	0.07	-0.08	0.28	0.33
PHLDA1	-0.16	-0.28	-0.13	-0.13	-0.09	0.25	0.16	0.35	0.20	1.00	0.88	0.55	0.46	0.22	0.09	-0.04	-0.01	0.01	-0.06	0.02	-0.02	0.06	0.05
NUP88	-0.16	-0.25	-0.14	-0.14	-0.12	0.25	0.22	0.34	0.26	0.88	1.00	0.52	0.50	0.23	0.13	0.01	0.02	0.10	0.00	0.07	-0.07	0.13	0.09
LRBA	-0.11	-0.19	-0.12	-0.13	-0.04	0.12	0.03	0.35	0.31	0.55	0.52	1.00	0.93	0.26	0.26	0.06	-0.09	0.03	0.02	-0.04	0.16	0.02	-0.01
TBX2	-0.09	-0.17	-0.13	-0.14	-0.08	0.06	0.05	0.38	0.40	0.46	0.50	0.93	1.00	0.17	0.23	0.08	-0.09	0.03	0.00	-0.08	0.15	0.05	-0.01
C3AR1	0.11	0.10	0.13	0.11	0.20	0.24	0.23	0.06	0.05	0.22	0.23	0.26	0.17	1.00	0.88	-0.02	0.06	-0.03	0.08	0.13	0.12	0.03	0.01
IL16	0.17	0.15	0.11	0.13	0.21	0.07	0.11	0.00	0.03	0.09	0.13	0.26	0.23	0.88	1.00	-0.14	-0.07	-0.08	-0.07	0.01	0.20	-0.05	-0.12
SRP72	-0.07	-0.02	-0.07	-0.10	-0.06	0.03	0.11	0.07	0.21	-0.04	0.01	0.06	0.08	-0.02	-0.14	1.00	0.82	0.73	0.88	0.80	-0.82	0.85	0.82
AHCY	-0.08	-0.01	-0.05	-0.08	-0.06	0.14	0.11	0.12	0.17	-0.01	0.02	-0.09	-0.09	0.06	-0.07	0.82	1.00	0.80	0.87	0.91	-0.84	0.90	0.88
CD59	-0.06	0.01	-0.04	-0.05	-0.07	0.05	0.03	0.08	0.19	0.01	0.10	0.03	0.03	-0.03	-0.08	0.73	0.80	1.00	0.80	0.73	-0.75	0.80	0.78
PPP1CA	-0.14	-0.05	-0.12	-0.18	-0.15	0.07	0.10	0.14	0.27	-0.06	0.00	0.02	0.00	0.08	-0.07	0.88	0.87	0.80	1.00	0.88	-0.81	0.90	0.91
PSMA2	-0.09	-0.04	-0.13	-0.15	-0.08	0.05	0.00	0.02	0.07	0.02	0.07	-0.04	-0.08	0.13	0.01	0.80	0.91	0.73	0.88	1.00	-0.80	0.87	0.84
TNC	0.08	0.04	0.06	0.07	0.11	-0.07	-0.11	0.00	-0.08	-0.02	-0.07	0.16	0.15	0.12	0.20	-0.82	-0.84	-0.75	-0.81	-0.80	1.00	-0.80	-0.87
ATP6V1C1	-0.15	-0.10	-0.13	-0.17	-0.16	0.16	0.19	0.18	0.28	0.06	0.13	0.02	0.05	0.03	-0.05	0.85	0.90	0.80	0.90	0.87	-0.80	1.00	0.90
PTMA	-0.13	-0.07	-0.08	-0.15	-0.15	0.15	0.18	0.26	0.33	0.05	0.09	-0.01	-0.01	0.01	-0.12	0.82	0.88	0.78	0.91	0.84	-0.87	0.90	1.00

Llistat dels gens que acumulen un 50% de la contribució total dins dels 10 primers metagens..

metagen 1

	contribució	contribució acum	correlació
HXB	0.07	0.07	-0.26
ATP6C	0.06	0.13	0.25
PTMA	0.04	0.17	0.21
NAB2	0.04	0.22	-0.20
AHCY	0.04	0.26	0.20
SRP72	0.04	0.30	0.20
RBM4	0.04	0.34	-0.20
MAPK6	0.04	0.37	-0.19
CD59	0.04	0.41	0.16
BNC	0.03	0.44	-0.18
TLH6	0.03	0.48	-0.18

metagen 2

	contribució	contribució acum	correlació
ARL6IP	0.23	0.23	0.48

H96850	0.06	0.29	0.24
ITGB1	0.04	0.33	0.20
NCBP1	0.04	0.37	0.20
KIAA0107	0.04	0.41	0.20
LYN	0.03	0.44	-0.18
PMSCL1	0.03	0.47	-0.17

metagen 3

	contribució	contribució acum	correlació
CD164	0.05	0.05	0.21
PAI-RBP1	0.03	0.08	0.17
ARL6IP	0.03	0.10	0.17
MGC15407	0.03	0.13	0.17
N91290	0.03	0.16	0.17
PHLDA1	0.03	0.19	0.16
PPP2R1B	0.03	0.21	0.16
CDC4L	0.02	0.24	0.15
DDX1	0.02	0.26	0.15
ALB	0.02	0.28	0.15
ENG	0.02	0.30	-0.15
RBBP2	0.02	0.33	0.15
KIAA0107	0.02	0.35	0.15
R62773	0.02	0.37	0.14
TCF7L2	0.02	0.39	-0.14
LOC51738	0.02	0.41	-0.14
TOP2A	0.02	0.43	0.14
NUP88	0.02	0.45	0.14
N52568	0.02	0.47	-0.14
N78103	0.02	0.49	0.14

metagen 4

	contribució	contribució acum	correlació
KIAA0265	0.07	0.07	0.26
LYN	0.06	0.12	-0.24
BIN1	0.05	0.17	0.21
CSE1L	0.04	0.22	0.21
PMSCL1	0.03	0.25	-0.19
DAZAP2	0.03	0.28	0.19

ADPRTL1	0.03	0.32	0.19
FUBP3	0.03	0.35	-0.17
HPR6.6	0.03	0.37	-0.16
R88904	0.03	0.40	0.16
CloneFBD3	0.02	0.42	0.16
TBX2	0.02	0.45	0.15
N52568	0.02	0.47	0.15
HOXD9	0.02	0.49	-0.15

metagen 5

	contribució	contribució acum	correlació
MESDC2	0.08	0.08	0.29
KIAA0265	0.06	0.14	-0.24
IGF2	0.05	0.20	-0.23
ARL6IP	0.05	0.25	0.23
WDR31	0.05	0.30	0.21
KIAA0203	0.04	0.34	-0.20
HUMAUANTIG	0.03	0.37	-0.17
NUP155	0.03	0.39	0.16
LAMA5	0.03	0.42	-0.16
SAT	0.02	0.44	-0.15
ZNF220	0.02	0.46	-0.15
ENG	0.02	0.48	0.15

metagen 6

	contribució	contribució acum	correlació
ZNF32	0.13	0.13	-0.36
UBL1	0.12	0.25	-0.35
EDG1	0.11	0.36	-0.34
CST3	0.10	0.46	-0.31

metagen 7

	contribució	contribució acum	correlació
1LBG	0.08	0.08	-0.28
N52568	0.07	0.15	-0.27
REV1L	0.07	0.22	-0.27
LOC51738	0.06	0.28	-0.25
PRKAG2	0.04	0.32	-0.20
IL16	0.03	0.36	0.19
C3AR1	0.03	0.39	0.18
R88904	0.03	0.42	-0.17
KIAA0210	0.03	0.44	-0.16

HINT	0.02	0.47	0.15
NUP155	0.02	0.49	0.15

metagen 8

	contribució	contribució acum	correlació
FKBP5	0.06	0.06	-0.25
PHLDA1	0.06	0.13	0.25
TPR	0.06	0.19	-0.25
CloneFBD3	0.04	0.23	-0.20
GSTTLp28	0.04	0.26	0.19
HINT	0.04	0.30	0.19
BIN1	0.04	0.33	0.19
C5	0.03	0.37	-0.18
NUP155	0.03	0.40	-0.18
R62773	0.03	0.43	-0.17
CAPZA2	0.03	0.46	-0.17
C20ORF1	0.03	0.49	-0.17

metagen 9

	contribució	contribució acum	correlació
CAPZA2	0.09	0.09	-0.30
HARS	0.04	0.13	-0.20
LYN	0.04	0.17	-0.20
HUMAUANTIG	0.04	0.21	0.20
CD9	0.04	0.25	0.19
FKBP5	0.03	0.28	0.17
GMFB	0.03	0.31	0.17
RBBP2	0.03	0.34	0.17
LTBP1	0.03	0.37	-0.17
PMSC1	0.03	0.39	-0.17
LYPLA1	0.03	0.42	0.17
PPP2R1B	0.03	0.45	-0.17
VCAM1	0.03	0.48	0.16

metagen 10

	contribució	contribució acum	correlació
PGM5	0.05	0.05	-0.22
FUBP3	0.05	0.10	-0.22
MTMR6	0.05	0.14	-0.22
KPNA1	0.04	0.19	-0.21
HPR6.6	0.04	0.23	0.21

N91290	0.04	0.27	-0.20
R62773	0.04	0.30	-0.19
EIF4A2	0.03	0.33	0.18
LTBP1	0.03	0.37	-0.18
HUMAUANTIG	0.03	0.39	0.16
C20ORF1	0.02	0.41	0.14
LCN2	0.02	0.43	0.14
ZNF220	0.02	0.45	0.14
TUBB	0.02	0.47	0.14
WBCR14	0.02	0.49	-0.14

Annex II: Article

Grau M, Solé X, Obrador A, Tarafa G, **Vendrell E**, Valls J, Moreno V, Peinado M, Capellà G. *Validation of RNA Arbitrarily-Primed PCR (RAP-PCR) probes hybridized to glass cDNA microarrays. Application to the analysis of limited samples.* Clinical Chemistry 2005, 51: 93-101.

La meva participació en aquest treball va ser principalment la posada a punt de la tècnica en els següents aspectes:

- *generació de les RAP-PCRs més adients per la seva reproductibilitat, i per la representativitat del transcriptoma.*
- *protocol de marcatge de les sondes i hibridació sobre els xips HU4.6K.*
- *protocol de recopilació de les dades i anàlisi.*

Validation of RNA Arbitrarily Primed PCR Probes Hybridized to Glass cDNA Microarrays: Application to the Analysis of Limited Samples

MÒNICA GRAU,^{1†} XAVIER SOLÉ,^{1†} ANTÒNIA OBRADOR,¹ GEMMA TARAFÀ,¹
ELISENDA VENDRELL,² JOAN VALLS,¹ VÍCTOR MORENO,¹ MIQUEL A. PEINADO,² and
GABRIEL CAPELLÀ^{1*}

Background: The applicability of microarray-based transcriptome massive analysis is often limited by the need for large amounts of high-quality RNA. RNA arbitrarily primed PCR (RAP-PCR) is an unbiased fingerprinting PCR technique that reduces both the amount of initial material needed and the complexity of the transcriptome. The aim of this study was to evaluate the feasibility of using hybridization of RAP-PCR products as transcriptome representations to analyze differential gene expression in a microarray platform.

Methods: RAP-PCR products obtained from samples with limited availability of biological material, such as experimental metastases, were hybridized to conventional cDNA microarrays. We performed replicates of self-self hybridizations of RAP-PCR products and mathematical modeling to assess reproducibility and sources of variation.

Results: Gene/slide interaction (47.3%) and the PCR reaction (33.8%) accounted for the majority of the variability. From these observations, we designed a protocol using two pools of three independent RAP-PCR reactions coming from two independent reverse transcription reactions hybridized in duplicate and evaluated them in the analyses of paired xenograft-metastases samples. Using this approach, we found that *HER2* and

MMP7 may be down-regulated during distal dissemination of colorectal tumors.

Conclusion: RAP-PCR glass array hybridization can be used for transcriptome analysis of small samples.

© 2005 American Association for Clinical Chemistry

The advent of techniques for the massive analysis of cell transcriptomes by use of microarrays has allowed the description of molecular portraits of biological specimens, including tumor biopsies. Tumor gene expression profiles appear to be useful tools for tumor classification and may reveal individual markers with diagnostic or prognostic applications. Nevertheless, routine application of gene expression profiles is often precluded by the demanding conditions of this type of assay, including the need for large amounts of RNA and the difficulties in performing global validation studies.

Several strategies have been developed to allow the analysis of small samples in which the amount of available RNA (<1 µg of total RNA) is insufficient for massive gene expression studies (1–4). Trenkle et al. (1) proposed the use of nonstoichiometric reduced-complexity probes for hybridization to cDNA arrays and noted the fitness of the RNA arbitrarily primed PCR (RAP-PCR)³ method. RAP-PCR is an unbiased fingerprinting PCR that samples a reproducible subset of message population based on the best matches with arbitrary primers (5). It allows the construction of a probe with reduced complexity, which increases the representation of rare messages, and uses small amounts of total RNA (10–100 ng) or mRNA (0.1–1 ng).

Although different RAP-PCR fingerprints give hybrid-

Institut d'Investigació Biomèdica de Bellvitge (IDIBELL)-Institut Català d'Oncologia, ¹Translational Research Laboratory, Unit of Biostatistics and Bioinformatics, Cancer Epidemiology Department, and ²IDIBELL-Institut de Recerca Oncològica Molecular Oncology Center, L'Hospitalet de Llobregat, Barcelona, Spain.

†These authors contributed equally to this work.

*Address correspondence to this author at: Institut Català d'Oncologia, Laboratori de Recerca Translacional, Av. Gran Via s/n, Km 2.7, 08907 L'Hospitalet de Llobregat, Barcelona, Spain. Fax 34-93-2607466; e-mail gcapella@ico.scs.es.

Received April 29, 2004; accepted September 24, 2004.

Previously published online at DOI: 10.1373/clinchem.2004.036236

³Nonstandard abbreviations: RAP-PCR, RNA arbitrarily primed PCR; SSC, standard saline citrate; and qRT-PCR, quantitative reverse transcription-PCR.

ization patterns with overlapping representations of expressed genes, the concordance is partial, and with several probes, a greater fraction of the message population can be screened than with total cDNA probes (1). Moreover, with reduced-complexity probes, the number of differentially expressed genes detected is 10 times higher than when total cDNA is used (3). This approach has been applied to the identification of genes differentially expressed in laser-microdissected colonic crypts (6, 7) and rheumatoid arthritis (8, 9). In all of these studies, including the original report (1), RAP-PCR products were hybridized to cDNA array membranes.

To gain insight into the molecular basis of the metastatic process, we used orthotopic implantation of human primary tumors in nude mice (10–12). Implanted xenografts resemble, in their early stages, primary tumors and reproduce, in part, their dissemination pattern, allowing completion of the metastatic process in a short time. Nevertheless, transcriptome analyses of experimental metastases may sometimes be hampered by their small size. In this study we evaluated the feasibility of hybridization of RAP-PCR product to slide arrays to allow use of small samples such as experimental metastases. By combining experimental designs and mathematical models, we investigated the distinct factors contributing to variability of hybridization of RAP-PCR probes on cDNA microarrays. From these observations, we designed a simpler procedure that minimizes the number of replicates and laboratory effort. This protocol was validated by the analysis of paired xenograft-metastases samples.

Materials and Methods

SAMPLES AND RNA EXTRACTION

We used 17 samples in evaluating our microarray hybridization protocol: 4 healthy colonic mucosas obtained from patients with colorectal cancer and undergoing curative surgery, and three sets of samples (C1, C4, and HM1)

obtained from colorectal tumors. Each set included a primary colorectal tumor, the corresponding orthotopic xenograft perpetuated in the nude mice, and a variable number (1–3) of experimental metastases (Table 1). Mean passage time from mice to mice of xenografts was 4 months. Fresh samples were embedded in OCT and kept frozen at –80 °C until their analysis. Manual microdissection was performed to ensure that 80% of all nuclei analyzed were from tumoral cells in tumor samples and from healthy epithelial cells in nontumor samples. Reproducibility experiments were performed with a single healthy colonic mucosa that was collected under standard conditions in our center, likely to represent the usual setting in which this technique will be used.

Total RNA was extracted by standard procedures (13). RNA quality and concentration were determined by spectrophotometry and gel electrophoresis. For selected experiments, mRNA was purified from total RNA by use of the PolyAtract® mRNA Isolation System III reagent set (Promega) according to the manufacturer’s instructions.

RAP-PCR

Reverse transcription was performed with 50 ng of total RNA, 200 U of M-MLV reverse transcriptase (Gibco BRL, Life Technologies), 15.3 units of RNAGuard RNase Inhibitor Porcine (Amersham Biosciences UK Limited), 0.5 mM deoxynucleotide triphosphates, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 10 mM dithiothreitol, and 0.5 μM primer in a final volume of 20 μL. The reaction was performed for 1 h at 37 °C, followed by 5 min at 95 °C. Primers were chosen arbitrarily from those available in our stock (none of them was specifically designed to do RAP-PCR) and were evaluated for quality of the fingerprint generated when run in sequencing gel electrophoresis (complexity of the band pattern and reproducibility) as described previously (14). The sequences of the selected primers were as follows: pU6, 5'-GCTTCTGACTTAT-

Table 1. Description of samples and their experimental use.

Sample	Experimental use
Nonneoplastic colonic mucosa	
Colonic mucosa 1 (reference sample)	Reproducibility Sources of variation Pooling Coverage of RAP-PCR
Colonic mucosa 2	Validation of pooling protocol
Colonic mucosa 3	
Colonic mucosa 4	
Tumor tissues	
Case C1	Validation of changes observed in glass hybridization with qRT-PCR
Primary tumor/xenograft/xenograft metastasis 1	
Case C4	
Primary tumor/xenograft/xenograft metastasis 1/ xenograft metastasis 2/xenograft metastasis 3	
Case HM1	
Primary tumor/xenograft/xenograft metastasis 1/ xenograft metastasis 2/xenograft metastasis 3	

TCTTGCTCTTAG-3'; H12B, 5'-CGCCAGGCTCACCTC-TATA-3'; TP53, 5'-AGGATACTATTCAGCCCCGAGGTG-3'; DB2-122dw, 5'-GCCGACGTGCTCCTGATGTT-3'; M11 8up, 5'-GTTTCAGTCTCAGCCATGAG-3'; DB2, 5'-ACAGATCTGAAGGGTGAAATATTCTCC-3'; and p53, 5'-ATGGAGGAGCCGCAGTCA-3'.

PCRs were performed with 5 μ L of the reverse transcription product in a final volume of 50 μ L. The reagents added were 2.5 U of *Taq* DNA polymerase (Boehringer Mannheim), 20 mM Tris-HCl (pH 8), 50 mM KCl, 2.5 mM MgCl₂, 0.1 mM deoxynucleotide triphosphates, and 2 μ M primer, which was the same primer that was used in the reverse transcription. Before the PCR, tubes were heated 4 min at 94 °C. PCR amplification consisted of five low-stringency cycles (1 min at 94 °C; 45 s at 40–55 °C, depending on the primer; and 75 s at 72 °C) followed by 35 high-stringency cycles (1 min at 94 °C; 45 s at 55–60 °C, depending on the primer; 75 s at 72 °C). Finally, tubes were kept for 5 min at 72 °C. We diluted 3 μ L of the RAP-PCR products in 9 μ L of denaturing loading buffer, heated the mixture for 3 min at 95 °C, and electrophoresed it in denaturing sequencing gels (6% polyacrylamide and 8 mol/L urea for 3–4 h at 55 W). Resolved bands were visualized by silver nitrate staining.

PROBE LABELING

RAP-PCR products were labeled with dCTP-Cy3 or dCTP-Cy5 (Amersham Biosciences UK Limited) by use of the Bioprime DNA Labeling System (Gibco BRL, Life Technologies). Probes were purified using ConcertTM Rapid Purification System (Gibco BRL, Life Technologies). DNA concentrations and rates of dye incorporation were quantified spectrophotometrically.

cDNA LABELING METHODS FOR TOTAL AMOUNT OF RNA

Reverse transcription and labeling reactions were performed at the same time to get the cDNA for hybridizing to cDNA microarrays. We used 1 μ g of total RNA, which was diluted in 2.4 μ L. RNA was mixed with 6 μ g of primer random hexamer in a final volume of 11 μ L. This mixture was incubated at 70 °C for 10 min and then was mixed with 600 U of SuperScript II reverse transcriptase (Invitrogen BV/NOVEX); 30 units of RNase inhibitor; 50 mM Tris-HCl; 75 mM KCl; 3 mM MgCl₂; 10 mM dithiothreitol; 0.5 mM each of dATP, dTTP, and dGTP; 0.2 mM dCTP; and 1 mM Cy dye in a final volume of 30 μ L. The mixture was incubated at 42 °C for 2 h, and after the first hour, 1 μ L of reverse transcriptase was added. Finally the reaction was stopped by addition of EDTA and NaOH to a final volume of 47.5 μ L (final concentrations, 0.63 and 31 mM, respectively). The last incubation was at 70 °C for 10 min followed by addition of 24 mM HCl to a final volume of 62.5 μ L. Probes were purified by use of the Concert Rapid Purification System. The cDNA concentration and rate of dye incorporation were quantified spectrophotometrically.

HYBRIDIZATION

For hybridization, we used poly-L-lysine cDNA microarrays [Hu 4.6K; W.M. Keck Foundation Biotechnology Resource Laboratory (<http://keck.med.yale.edu>)] containing 4608 human cDNAs spotted in duplicate. A list of the genes covered can be obtained at http://keck.med.yale.edu/dna_arrays.htm. The hybridization area was 22 \times 22 mm. We dried 1–1.5 μ g of the labeled DNA probe (usually corresponding to 20–50 pmol of dye) and resuspended it in a solution containing 1.6 g/L yeast tRNA, 1.6 g/L poly(dA), 1.6 g/L Cot-1 DNA, 3.5 \times standard saline citrate (SSC), and 3 g/L sodium dodecyl sulfate in a final volume of 14.52 μ L. The labeled DNA was denatured for 1 min at 95 °C and left for 30 min at room temperature to allow it to interact with the blocking agents. The slides were denatured for 2 min in boiling water and kept for 2 min at 76 °C. Finally, the probe was put on the slide, and hybridization was performed in a humidity chamber at 65 °C for 14–16 h. After hybridization, the slides were washed consecutively with 2 \times SSC–1 g/L sodium dodecyl sulfate, 1 \times SSC, and 0.2 \times SSC for 2 min each.

IMAGE ANALYSIS AND DATA PROCESSING

Slides were scanned with a GSI Lumonics ScanArray 4000. Images were analyzed by use of Spot software (CSIRO, Mathematical and Informatics Sciences), and the resulting data were processed to filter out low-quality spots based on spot area and similarity of readings between the two replicates of each gene. Replicate intensities were averaged. Subsequently, the two channels were normalized by use of the lowest nonparametric smoother method of a set of invariant genes because correlation of Cy5 and Cy3 values varies with intensity (15).

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS

To assess the sources of variability in measured log-intensities, we performed multiple analyses of the same sample. The main factors considered to introduce random variation were (a) slide/hybridization, which we considered as the same effect, because in one slide only one hybridization can be performed, and (b) the probe synthesis process, which includes the reverse transcription reaction, PCR reaction, and labeling. Starting from the same RNA sample, different products were obtained and hybridized in multiple combinations, including labeling of individual RAP-PCR reactions or pools of them and hybridization of the same PCR labeled with distinct fluorochromes, as illustrated in Fig. 1.

We used the restricted maximum likelihood method to estimate the components of variance, allowing assessment of the relative magnitude of each source of variability. Initially, we developed a mathematical model for single-slide hybridization that included both the random variation (attributable to the sequence of a given gene and the amount of a given RNA in our target sample) and the variables introducing artifactual variability: the fluoro-

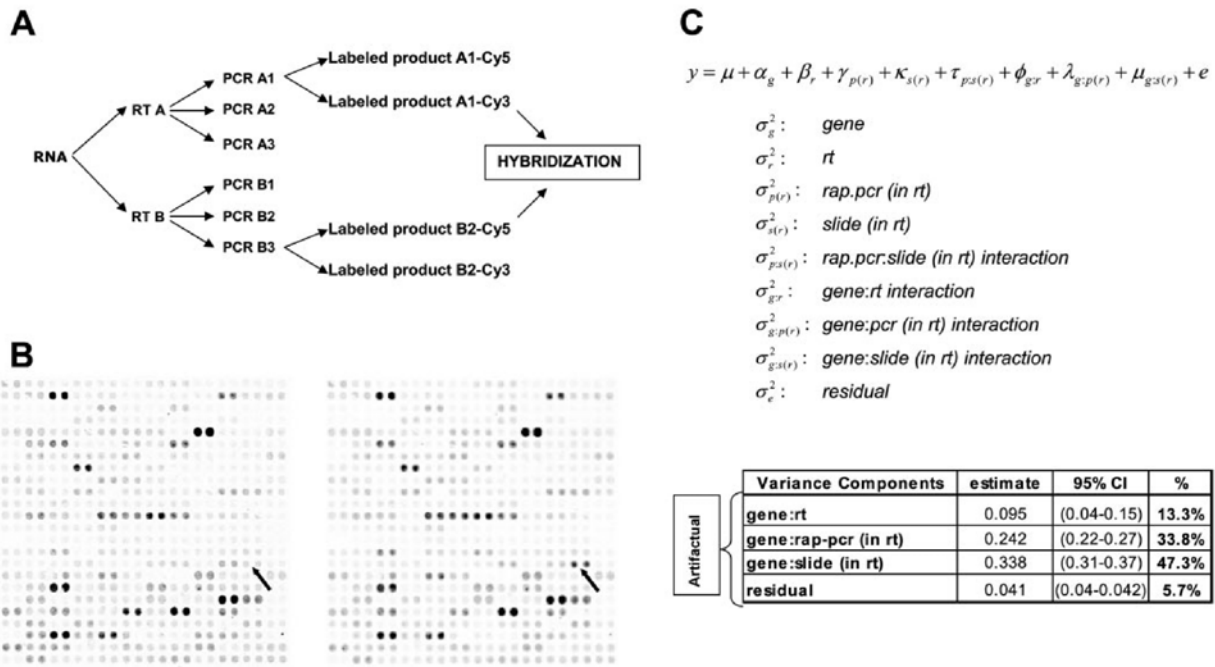


Fig. 1. Reproducibility analysis of RAP-PCR hybridization on glass slides.

Analysis of components of variance. (A), experimental design for reproducibility analysis. The reverse transcription (RT) reaction was performed twice in parallel with the same RNA sample. From each reverse transcription product, three PCR amplifications were performed, each labeled with a different dye. Hybridizations of different combinations of the products obtained, such as the one depicted, were performed in duplicate. Actual experiments included the expansion of all collapsed branches and cohybridization of the different products in multiple combinations. (B), results of hybridization of the two products depicted in A on a subarray containing 288 genes (spotted in duplicate). Arrowheads indicate discordant results in self-self hybridizations. (C), algorithm representing each of the steps of the procedure (reverse transcription, RAP-PCR, labeling, and slide) used to analyze components of variance. Data shown are based on modeling of single-channel signal intensities. CI, confidence interval.

chrome, the reverse transcription reaction, and the PCR reaction. Nevertheless, because of the nature of the experiments to be performed, comparison between intensities obtained with distinct slides is mandatory and may be particularly critical when comparing serial samples (i.e., normal-adenoma-carcinoma-metastasis). Thus, we further developed the model to allow for accurate sample-to-sample comparisons. Once the origin of variability was ascertained, replicates and/or pooling in the more critical steps were performed to produce a final protocol.

VALIDATION OF CHANGES OBSERVED IN PAIRED TUMOR-METASTASES SAMPLES

We compared changes in three specific genes, *MMP7*, *HER2*, and *PCAF*, that showed recurrent differences in the RAP-PCR/microarray system, with data obtained by quantitative reverse transcription-PCR (qRT-PCR) to validate the nature and magnitude of the change. The PCR reactions were performed in a LightCycler apparatus (Roche Diagnostics S.L. Applied Science), using the LC-FastStart DNA Master SYBR Green I reagent set (Roche Diagnostics S.L. Applied Science) according to the manufacturer's instructions. Selection of adequate housekeeping control genes is critical. β_2 -Microglobulin and cyclophilin were chosen based on our previous experience. The

ratio between them was constant at 1, thus (16) one of them, β_2 -microglobulin, was finally used to control for input RNA. A relative calibration curve was constructed for each gene with use of five serial dilutions starting with 100 ng of RNA. Relative quantification between samples was done by use of a mathematical model (17). Primer sequences and PCR conditions may be obtained from the authors on request.

Results

REPRODUCIBILITY AND SOURCES OF VARIATION

For this set of experiments, we used total cellular RNA from a single colonic mucosa. In a preliminary experiment, we observed that hybridization of products coming from total RNA and from mRNA yielded comparable results (data not shown). Although these two representations are not fully equivalent, for methodologic simplicity and feasibility in routine analyses, we used total RNA.

To analyze reproducibility and sources of variation, we performed several self-self hybridizations of RAP-PCR products, as described above, labeled with distinct fluorochromes and hybridized to conventional cDNA microarrays (Fig. 1, A and B). We used the variance of log ratios as a measure of variability. Comparison of independent hybridizations of RAP-PCR probes obtained from

the same sample showed a typical log ratio variance of 1.2^2 (range, 0.82^2 – 1.3^2) when no replicates or pooling was performed. To explore the sources of this variability, we focused on variation of signal intensity, not on variance of log ratios. Single-channel intensities of self-self hybridizations, previously normalized to allow for comparison between distinct fluorochromes, were analyzed for variance components. As expected, differences in the representativeness of the 4608 different genes of the sample used accounted for a major fraction of the variability. This effect was so large [73% of the total variability (95% confidence interval, 63–82%)] that it was taken into account and subsequently fixed in the model, allowing the analysis of artifactual variability. It has previously been suggested that fluorochromes may introduce artifactual variability (16). In our hands, hybridization of the same RAP-PCR product labeled with the two dyes did not vary, suggesting that dye/gene interaction was negligible. This allowed us to exclude a dye effect from the model and simplify it.

Modeling of the variability associated with this type of analysis indicated that the major source of variability (47.3%) was gene/slide interaction, i.e., the effect of performing hybridization in a different slide. The PCR reaction accounted for one third of the variability (33.8%), and the reverse transcription reaction for an additional 13.3% (Fig. 1). Residual variance was very low (5.7%), as expected.

RAP-PCR VARIABILITY CAN BE EFFECTIVELY REDUCED THROUGH POOLING

On the basis of these observations, we attempted to generate an experimental design that could offer an acceptable artifactual variability. To determine whether several hybridizations of replicates performed better than hybridization of pooled replicates, we performed two

distinct reverse transcription reactions with our sample template and three independent RAP-PCR reactions for each reverse transcription, yielding six PCR probes (Fig. 2). Variance calculated from the mean values of three hybridizations of PCRs coming from independent reverse transcriptions (0.81^2 ; range, 0.55^2 – 0.93^2 ; Fig. 2A) was significantly lower than the basal log ratio variance of 1.2^2 from a single hybridization. Because of the high cost of repeated hybridizations, we did not consider hybridization in duplicate of the same PCR product. Rather, we evaluated the effect on reproducibility of pooled RAP-PCR products of the same RNA (Fig. 2B). Interestingly, variances of the hybridization of two pools of independent RAP-PCR reactions coming from two independent reverse transcription reactions were comparable (0.91^2 ; range, 0.63^2 – 1.10^2). On the basis of these results, we decided to establish the protocol described in Fig. 2C, which consists of hybridizing, in duplicate, two pools of three independent RAP-PCR reactions coming from two independent reverse transcription reactions. We validated this protocol by analyzing three additional mucosae. For sample 1, the variance for a single hybridization was 0.71^2 , and the variance for the pooling hybridization was 0.59^2 (range, 0.53^2 – 0.64^2). The corresponding results for sample 2 were 0.92^2 and 0.80^2 (0.78^2 – 0.81^2), and for sample 3 were 1.10^2 and 0.83^2 (0.78^2 – 0.88^2).

COVERAGE OF RAP-PCR VS cDNA

We next wanted to determine whether the transcriptome coverage of the RAP-PCR products compared with that of whole cDNA. Each RAP-PCR product as well as cDNA obtained from the same healthy colonic sample was labeled exclusively with Cy5 to avoid variations in intensity from competition. All products were independently hybridized in distinct slides. With this design, no normalization of the data was performed because distribution of

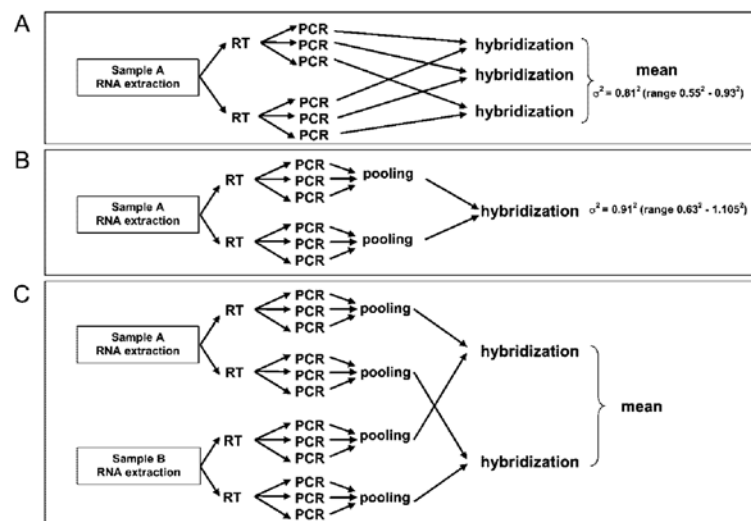


Fig. 2. Design of the protocol for small biopsy analysis.

Panels A and B depict the distinct approaches used to reduce variability. (A), three independently obtained RAP-PCR probes are also independently analyzed. (B), pooling of independent RAP-PCR products in a single hybridization partially reduces the variance (see text). (C), protocol used in the analysis of metastases based on the variance observed in experiments A and B and also taking into consideration costs. RT, reverse transcription.

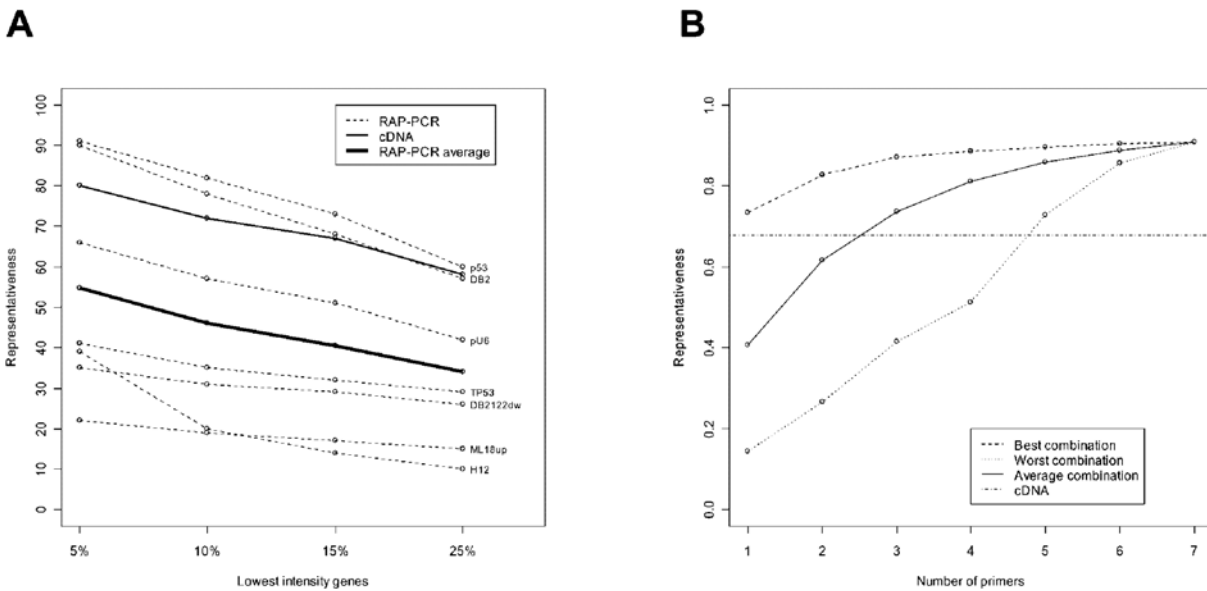


Fig. 3. Representativeness of RAP-PCR products hybridized to cDNA microarrays. (A), percentage of represented genes for seven different RAP-PCR products compared with cDNA. Values shown exclude a variable percentage of the spots showing the lowest intensities. Although no objective threshold has been established, the relationship between different products was maintained. (B), theoretical calculation of the percentage of represented genes after distinct combinations of independent RAP-PCR product hybridizations. The best combination of primers clearly outperforms cDNA. On average, the combination of any three RAP-PCR products can be compared with cDNA hybridization.

intensities for each primer was a priori unknown, and thus distributions could not be assumed to be equal.

A threshold had to be defined to classify the genes as represented (18). The distribution of intensities of RAP-PCR and cDNA hybridizations precluded the definition of an objective threshold. Initially, spots with signal intensities less than twofold higher than the background (<1% of the 4608 genes in the slide) were assumed to have no signal. The representativeness (number of spots with specific hybridization) of the different RAP-PCR products and the cDNA was affected equally when we used distinct arbitrary thresholds based on exclusion of spots depicting the lowest intensity (Fig. 3). When we fixed as a threshold twice the mean intensity of the 15% lowest genes, cDNA hybridization showed relevant signals in 68% of the genes spotted on the slide, in the same range as the sum of the three RAP-PCR products mean of all 3-to-3 combinations; Fig. 3B). The addition of the three best primers accounted for most of this representativeness (Fig. 3) and clearly outperformed cDNA. Finally, the cumulated information provided by all seven RAP-PCR products analyzed reached 91% for the genes, 23% more than was obtained when we used cDNA (Fig. 3). The overrepresentation of RAP-PCR products compared with cDNA is likely to reflect the advantage of RAP-PCR to amplify low-abundance mRNAs that are poorly detected in whole cDNA hybridizations.

An additional theoretical advantage of using RAP-PCR products is the possibility of having multiple controls if the tag gene is represented in independent RAP-PCRs.

When four or more different RAP-PCR products are hybridized, up to 70% of the genes are scored in two or more hybridizations, which indicates the powerful autovalidating feature of this approach even if no replicates are performed (Fig. 4).

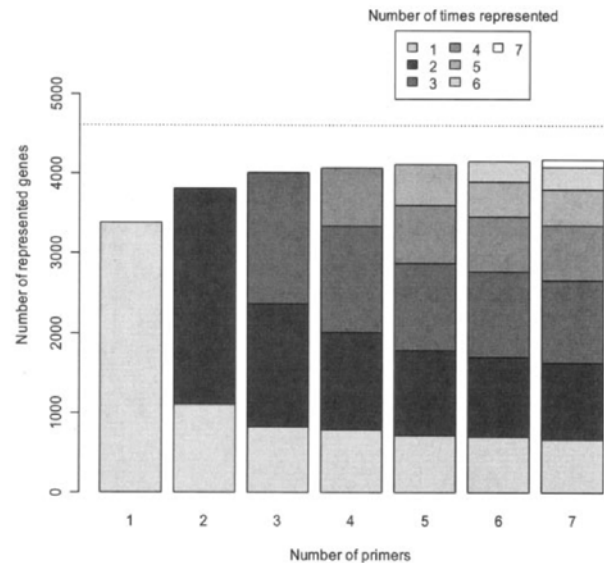


Fig. 4. Overlapping coverage of RAP-PCR probes. The combination of the three best RAP-PCR probes achieves an extremely high representativeness. Furthermore, up to 70% of the genes would have been analyzed at least twice.

CONFIRMATION OF DIFFERENTIAL EXPRESSION BETWEEN DIFFERENT SAMPLES

With this protocol, we studied three sets of related samples including a primary tumor, corresponding orthotopic xenograft, and one to three experimental metastases. Three arbitrarily selected genes previously associated with cancer development (*MMP7*, *HER2*, and *PCAF*) were used to ascertain the usefulness of our approach. According to microarray hybridizations, *MMP7* expression changed in a nonconsistent manner in xenografts compared with the corresponding primary tumors. Unexpectedly, in five of seven metastases analyzed, *MMP7* expression was further diminished. Similar results were obtained for *HER2* and *PCAF*.

All changes with log ratios higher than 1.5, as assessed by RAP-PCR hybridization, were confirmed by qRT-PCR (Fig. 5). The magnitude of ratios from microarray experiments tended to be lower than those from qRT-PCR, suggesting that kinetics of hybridization in the slide setting tend to smooth actual changes. On the other hand, discordant results were obtained for a portion of *HER2* ($n = 3$) and *PCAF* ($n = 5$) analyses, where log ratios of 7–9 were not detected by slide hybridizations, a fact that may be also attributed to the kinetics of hybridization. For the analyzed genes, no false-positive changes were detected.

Discussion

A prominent example of complex disease with heterogeneous components is cancer. Genetic profiling of cancer

cells may contribute to the diagnosis, prognostic assessment, and ultimately, the design of specific therapies. Recent studies have already demonstrated the potential of the technique. However, DNA microarray technology presents critical limitations, including deficient scoring of low-abundance messages, poor adaptability to routine procedures, lack of feasible alternatives to the global validation of the data, and the need for large amounts of RNA.

The hybridization of RAP-PCR to filter arrays has been shown to solve, in part, these limitations (1). The reduction of the complexity of the transcriptome together with its expansion by RAP-PCR was critical (a) to increase the sensitivity of the technique, which is seriously limited by the overwhelming complexity of the sample's transcriptome; and (b) to overcome the small sample sizes routinely available in the clinical setting. Because very low amounts of biological material (<50 ng of total RNA/experiment or 10 pg of mRNA/experiment) are required, the use of microdissected samples is allowed. Additionally, the use of lower-complexity representations is likely to allow the unbiased representation of rare and common mRNAs. This technique has already been applied (6–8) to a filter array platform that requires the use of radioactivity and is difficult to automate. We therefore decided to develop and evaluate a glass microarray system that requires no radioactive handling, facilitates automation, and offers the possibility of cohybridization of two or more samples (if distinct fluorochromes are to be used).

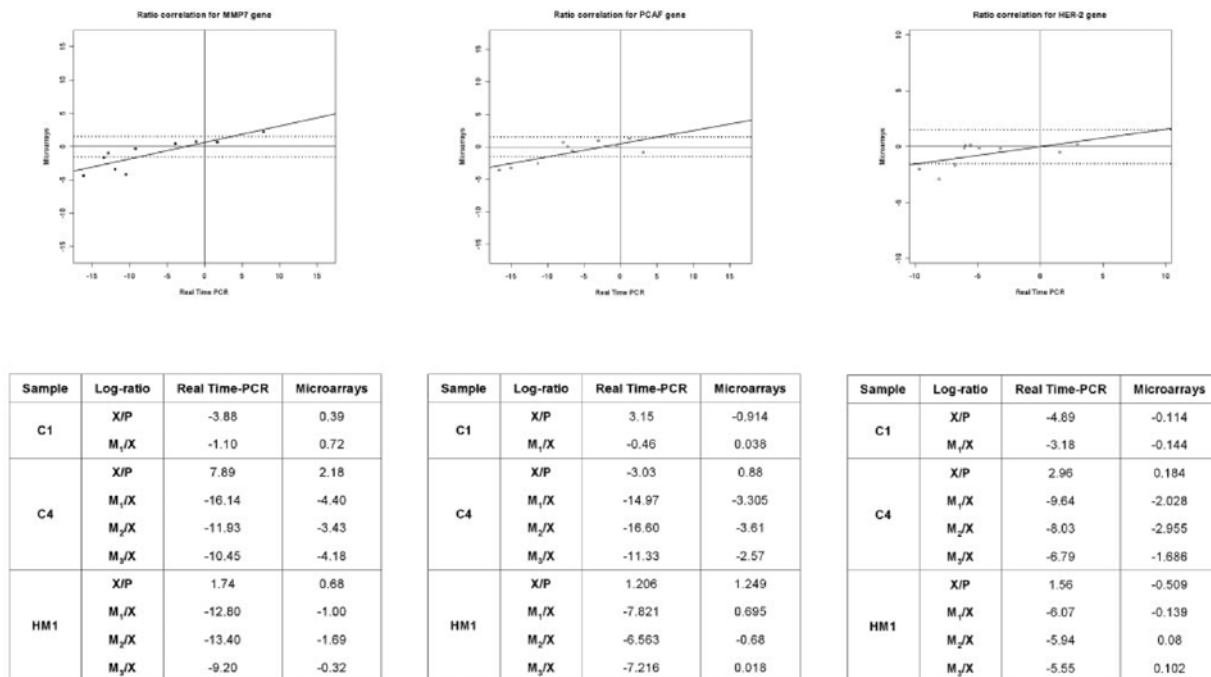


Fig. 5. Validation of RAP-PCR hybridization with qRT-PCR for *MMP7* (left), *PCAF* (middle), and *HER2* (right) expression.

(Bottom of each graph), actual ratios expressed as log₂ changes. (Top of each graph), graphical representation of results where both axes are in the same scale to better visualize the smoothing of the ratios obtained in the RAP-PCR microarray experiments. Thick lines indicate regression equations.

The use of a PCR step in hybridizing a glass array is a major source of variation in the procedure that might well preclude its routine use. In spite of the importance of quantification of artifactual variability, few studies have addressed this issue in a systematic manner in array hybridizations. In the present study, we performed several self-self hybridizations, using distinct conditions, and created a mathematical model to assess the sources of artifactual variability in RAP-PCR hybridization to glass arrays. In agreement with previous studies showing that the array itself is the most common source of error in array-based analyses (16), gene/slide interaction was shown to contribute to 50% of the artifactual variability observed. This can be minimized only by the use of replicates, which will, on the other hand, substantially increase costs. As expected, the RAP-PCR procedure and, to a lesser extent, the reverse transcription reaction also contribute to artifactual variability. Fortunately, pooling of reactions can minimize this variability in an affordable manner, as shown for other DNA amplification procedures such as whole genome amplification (19).

We do not know whether the magnitude of artifactual variability resulting from use of RAP-PCR products as a probe compares with that resulting from use of cDNA probes. Nevertheless, we point out that estimates of variability—based on experimental data—are needed to establish the need for either replicate hybridizations or pooling of reactions. Estimation of variability may be also of help in choosing the strategy of statistical analyses to be performed.

Only the combined use of sample pooling—because it may reduce random amplification events—and replicate hybridizations can reduce the artifactual variability observed. Thus, based on our observations, we finally propose performing two hybridizations of distinct pools of RAP-PCR reactions, which is readily amenable for analysis of small samples, either experimental or clinical, increasing the feasibility of using RAP-PCR in the routine clinical setting. Although this strategy offers a good balance between cost and feasibility, there is still room for improvement. We must take into account that, in our approach, true replicates (hybridization of the same probe on distinct slides) are not used. Some artifactual variability attributable to true replicates may be expected (data not shown). However, although our protocol does not correct for this type of variability, the use of two hybridizations is likely to make these results sufficiently robust.

Another potential advantage of hybridization of RAP-PCR probes is the generation of data that may not be easily obtained by conventional cDNA or oligonucleotide-based microarrays because of nonstoichiometric selection of the represented transcripts. We have shown that, in the stringent hybridization conditions (65 °C) used in our setting, coverage of the best RAP-PCR product is comparable to that of whole cDNA, suggesting that a less-complex probe offers a better relative fraction. Moreover, coverage may be improved by the addition of indepen-

dent hybridizations of different RAP-PCRs, which would also add confidence to the results because of the overlapping of coverages observed between distinct RAP-PCR products. The concomitant hybridization of distinct RAP-PCR products could theoretically be a good option to obtain, in a single hybridization, more coverage. However, experimental data obtained in nylon filters suggest that simultaneous hybridization of distinct RAP-PCR probes may lead to reduced target detection, limiting its potential benefits (20).

In the experimental samples used to validate our protocol, all changes in expression detected by RAP-PCR array hybridization with log ratios >1.5 were confirmed by qRT-PCR. In agreement with previous reports, qRT-PCR confirmed the trend of poor concordance regarding log-ratio values (17). Most of the studies with cDNA hybridizations have used arbitrary cutoff values such as two- or threefold changes to define a true change, but few of them have provided the experimental data to support the criteria used. Thus, in our experimental setting and with use as a reference a threshold value based on the variance of self-self experiments (18), threefold changes should be considered as true changes. However, our results also show the limitations of the sensitivity of our protocol. In some paired samples, qRT-PCR changes showing log ratios of 7–9 were not detected by RAP-PCR slide hybridization. It is likely that the compressed range for detecting intensity changes in slide hybridizations compared with qRT-PCR leads to loss of sensitivity. Finally, the systematic bias of the technique (21), which is not always improved by the use of replicates, may account in part for a loss of accuracy. This is not apparently the case because this bias, if present, tended to improve the performance of our methodology when we used our protocol.

In summary, we have systematically evaluated the sources of artifactual variability for RAP-PCR products hybridized to glass arrays by combining replicates of self-self hybridizations of RAP-PCR products and performing mathematical modeling. Gene/slide interaction and the PCR reaction are the major sources of artifactual variability, which can be reduced by combining hybridization replicates and pooling of samples. We have proposed a protocol that has been essentially validated by qRT-PCR analyses. Using this approach, we found that *HER2*, *MMP7*, and *PCAF* may be down-regulated during distal dissemination of colorectal tumors. We therefore consider that we have developed a consistent hybridization protocol that allows the use of the RAP-PCR/microarray approach for large-scale study of transcriptomes in small samples from experimental protocols or obtained in the routine clinical setting.

This study was supported by grants from the Ministerio de Ciencia y Tecnología (SAF00/81 and SAF03/5821) and

the Fondo de Investigaciones Sanitaria (FIS 01.1264 and FIS 00/0021). M.G. is a recipient of a FPI MCYT fellowship. The research team belongs to the Network of Cooperative Research on Cancer (C03/10) and Epidemiology and Public Health (C03/09), funded by the Instituto Carlos III, Ministerio de Sanidad y Consumo of Spain.

References

1. Trenkle T, Welsh J, Jung B, Mathieu-Daude F, McClelland M. Non-stoichiometric reduced complexity probes for cDNA arrays. *Nucleic Acids Res* 1998;26:3883-91.
2. Welford SM, Gregg J, Chen E, Garrison D, Sorensen PH, Denny CT, et al. Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization. *Nucleic Acids Res* 1998;26:3059-65.
3. Sakai K, Higuchi H, Matsubara K, Kato K. Microarray hybridization with fractionated cDNA: enhanced identification of differentially expressed genes. *Anal Biochem* 2000;287:32-7.
4. Vernon SD, Unger ER, Rajeevan M, Dimulescu IM, Nisenbaum R, Campbell CE. Reproducibility of alternative probe synthesis approaches for gene expression profiling with arrays. *J Mol Diagn* 2000;2:124-7.
5. Welsh J, Chada K, Dalal SS, Cheng R, Ralph D, McClelland M. Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res* 1992;20:4965-70.
6. Lechner S, Muller-Ladner U, Neumann E, Dietmaier W, Welsh J, Scholmerich J, et al. Use of simplified transcripts for the analysis of gene expression profiles in laser-microdissected cell populations. *Lab Invest* 2001;81:1233-42.
7. Lechner S, Muller-Ladner U, Renke B, Scholmerich J, Ruschoff J, Kullmann F. Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia. *Gut* 2003;52:1148-53.
8. Neumann E, Kullmann F, Judex M, Justen HP, Wessinghage D, Gay S, et al. Identification of differentially expressed genes in rheumatoid arthritis by a combination of complementary DNA array and RNA arbitrarily primed-polymerase chain reaction. *Arthritis Rheum* 2002;46:52-63.
9. Judex M, Neumann E, Lechner S, Dietmaier W, Ballhorn W, Grifka J, et al. Laser-mediated microdissection facilitates analysis of area-specific gene expression in rheumatoid synovium. *Arthritis Rheum* 2003;48:97-102.
10. Reyes G, Villanueva A, Garcia C, Sancho FJ, Piulats J, Lluís F, et al. Orthotopic xenografts of human pancreatic carcinomas acquire genetic aberrations during dissemination in nude mice. *Cancer Res* 1996;56:5713-9.
11. Villanueva A, Garcia C, Paules AB, Vicente M, Megias M, Reyes G, et al. Disruption of the antiproliferative TGF- β signaling pathways in human pancreatic cancer cells. *Oncogene* 1998;17:1969-78.
12. Tarafa G, Villanueva A, Farre L, Rodriguez J, Musulen E, Reyes G, et al. DCC and SMAD4 alterations in human colorectal and pancreatic tumor dissemination. *Oncogene* 2000;19:546-55.
13. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987;162:156-9.
14. Tortola S, Capella G, Marcuello E, Gunther K, Aiza G, Masramon L, et al. Analysis of differential gene expression in human colorectal tumor tissues by RNA arbitrarily primed-PCR: a technical assessment. *Lab Invest* 1998;78:309-17.
15. Schadt EE, Li C, Ellis B, Wong WH. Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J Cell Biochem Suppl* 2001;37:120-5.
16. Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 2002;32(Suppl):490-5.
17. Jenson SD, Robetorye RS, Bohling SD, Schumacher JA, Morgan JW, Lim MS, et al. Validation of cDNA microarray gene expression data obtained from linearly amplified RNA. *Mol Pathol* 2003;56:307-12.
18. Sabatti C, Karsten SL, Geschwind DH. Thresholding rules for recovering a sparse signal from microarray experiments. *Math Biosci* 2002;176:17-34.
19. Rook MS, Delach SM, Deyneko G, Worlock A, Wolfe JL. Whole genome amplification of DNA from laser capture-microdissected tissue for high-throughput single nucleotide polymorphism and short tandem repeat genotyping. *Am J Pathol* 2004;164:23-33.
20. Sloan DD, Nicholson B, Urquidi V, Goodison S. Detection of differentially expressed genes in an isogenic breast metastasis model using RNA arbitrarily primed-polymerase chain reaction coupled with array hybridization (RAP-array). *Am J Pathol* 2004;164:315-23.
21. Stirewalt DL, Pogossova-Agadjanyan EL, Khalid N, Hare DR, Ladne PA, Sala-Torra O, et al. Single-stranded linear amplification protocol results in reproducible and reliable microarray data from nanogram amounts of starting RNA. *Genomics* 2004;83:321-31.