



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

La disomia uniparental adquirida en el càncer colorectal

Memòria presentada per

Keyvan Torabi Asensio

Per optar al títol de

Doctor en Biologia Cel·lular

Tesi realitzada sota la direcció de

Dr. Jordi Camps

Dra. Rosa Miró

Dra. Immaculada Ponsa

Departament de Biologia Cel·lular, Fisiologia i Immunologia
Facultat de Medicina de la Universitat Autònoma de Barcelona

Dr. Jordi Camps Polo

Dra. Rosa Miró Ametller

Dra. Immaculada Ponsa Arjona

Keyvan Torabi Asensio

Bellaterra (Cerdanyola del Vallès), Juny 2017

Contingut

1	Introducció	1
1.1	El càncer colorectal	1
1.1.1	Epidemiologia	1
1.1.2	Anatomia i histologia de l'intestí gros.....	2
1.1.3	Classificació del CRC per estadiatges.....	3
1.1.4	Etiologia del CRC	5
1.2	Classificació molecular del CRC	7
1.2.1	Via del fenotip metilador d'illes CpG (CIMP).....	9
1.2.2	Inestabilitat de microsatèl·lits (MSI).....	9
1.2.3	Inestabilitat cromosòmica	10
1.3	Disomia uniparental (UPD).....	16
1.3.1	UPD constitucional	16
1.3.2	Mecanismes de formació de UPD constitucional.....	18
1.3.3	UPD i càncer.....	20
1.3.4	Mecanismes de formació de UPD somàtica	24
1.4	Tècniques per a l'estudi d'alteracions cromosòmiques	26
1.4.1	Arrays de CGH (aCGH)	26
1.4.2	SNP arrays	27
1.4.3	Hibridació <i>in situ</i> fluorescent (FISH)	28
1.5	Arquitectura nuclear	29
2	Premissa i objectius.....	33
3	Resultats.....	34
3.1	Abstract.....	34
3.2	Capítol I.....	36
3.3	Capítol II.....	46
3.4	Capítol III.....	68
4	Discussió	86
4.1	Eines d'estudi de la CIN al CRC	86
4.2	Perfils genòmics de CNAs al CRC	87
4.3	Disomia uniparental: alteració cromosòmica comuna al càncer.....	91
4.4	Perfils genòmics de UPD al CRC.....	92
4.5	UPD com a mecanisme alternatiu d'inactivació de TSGs	94

4.6	La UPD és més freqüent en casos altament aneuploides.....	98
4.7	La UPD telomèrica és el tipus d'esdeveniment més present, afectant principalment a <i>APC</i>	99
4.8	La importància de l'arquitectura nuclear en la formació de UPDs segmentals.....	101
4.9	L'expressió gènica és un factor clau per a la localització intracromosòmica dels gens	106
5	Conclusions	110
6	Bibliografia	111
7	Annex.....	129
8	Agraïments	130

1 Introducció

1.1 El càncer colorectal

1.1.1 Epidemiologia

El càncer colorectal (CRC, de l'anglès *Colorectal Cancer*) és un problema d'abast mundial, amb 746.000 nous casos cada any se situa en la tercera posició dels càncers més freqüents en homes (10%), darrere pulmó i pròstata. En dones, 614.000 nous casos cada any el situen com a segon càncer més freqüent (9,2%), tan sols darrere del de mama segons dades del Globocan de 2012 (Torre et al. 2015; Ferlay et al. 2015). Existeix un patró geogràfic on els països reconeguts com a més desenvolupats mostren incidències més elevades per a aquesta malaltia (Figura 1). La mortalitat de la malaltia és una mica més baixa, 697.000 casos per any, amb més morts en els països reconeguts com a menys desenvolupats (52%), reflectint una pitjor supervivència de la malaltia en aquests països (Torre et al. 2015) (Figura 2).

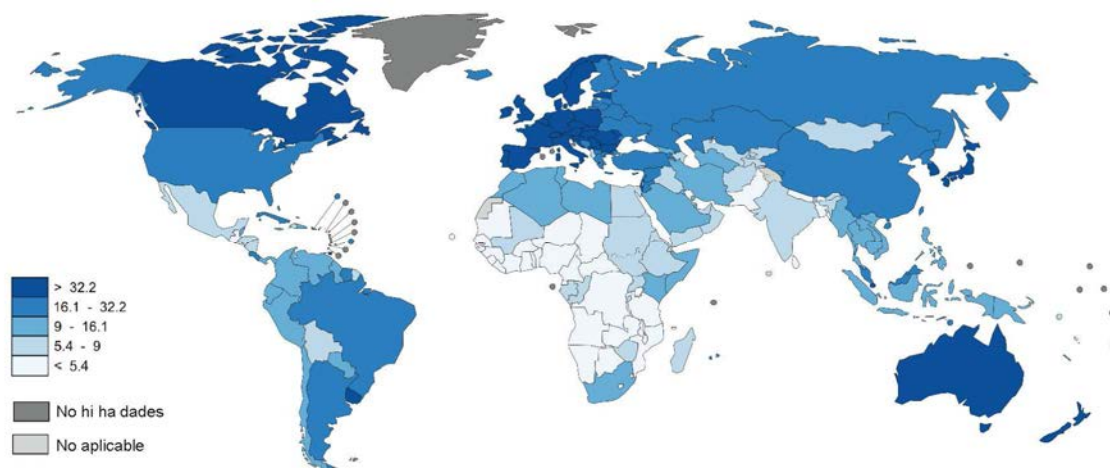


Figura 1. Incidència del CRC a nivell mundial. Figura adaptada de Globocan 2012 (<http://globocan.iarc.fr>).

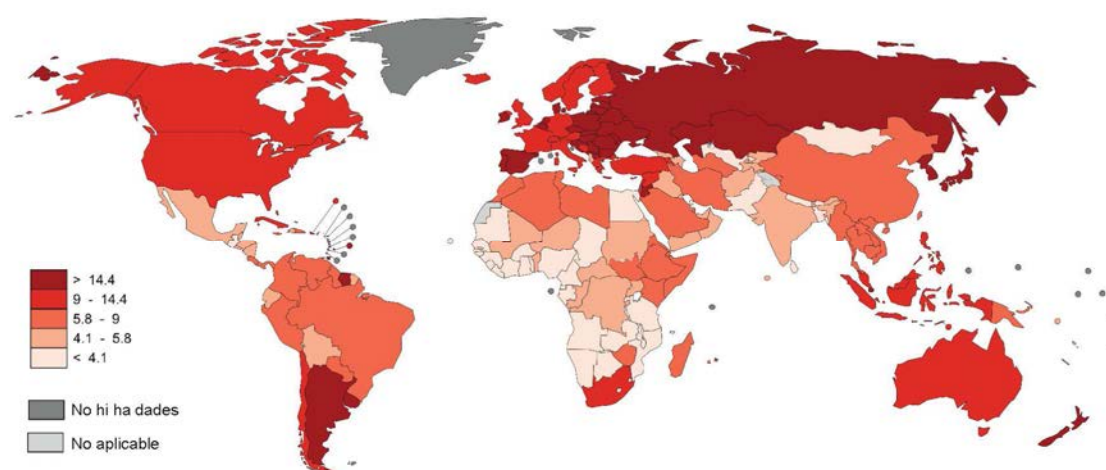


Figura 2. Mortalitat del CRC a nivell mundial. Figura adaptada de Globocan 2012 (<http://globocan.iarc.fr>).

Als Estats Units d'Amèrica, el CRC té una alta incidència, situant-se tercer en nombre total de casos darrere de pròstata i pulmó en homes, i de mama i pulmó en dones (Siegel et al. 2016). No obstant, segons dades provinents del National Institutes of Health (NIH), aquest càncer mostra una tendència a la baixa de nous casos per any, amb una mitjana al voltant d'un 3% menys de casos per any durant el període 2003-2012 (Siegel et al. 2016). L'increment en l'ús de la colonoscòpia, amb la finalitat d'eliminar lesions pre-canceroses, sembla ser el responsable de la baixada de la taxa d'incidència (Siegel et al. 2012). Contràriament, la incidència ha pujat de mitjana un 1,8% per any entre el 1992 i el 2012 al subgrup de pacients menors de 50 anys, on l'*screening* per a la detecció de la malaltia no està implementat (Siegel et al. 2016).

1.1.2 Anatomia i histologia de l'intestí gros

L'intestí gros es divideix en cinc seccions: el cec més el còlon ascendent, el còlon transversal, el descendent, el sigmoide i el recte. Paral·lelament és comú adoptar la divisió de l'intestí gros en dues seccions: el còlon proximal o dret que engloba el cec, el còlon ascendent i el transversal; i el còlon distal o esquerre que engloba el còlon descendent, sigmoide i recte (Figura 3). Aquesta classificació pren importància a nivell patològic ja que, com es descriurà més endavant, els tumors associats a inestabilitat de microsatèl·lits tendeixen a estar localitzats al còlon proximal, mentre que els que mostren inestabilitat cromosòmica tendeixen a estar localitzats al còlon distal (Lothe et al. 1993). A nivell histològic, podem diferenciar les següents capes de l'intestí gros, des del

lumen cap a l'exterior: mucosa, submucosa, muscularis propria i serosa (Figura 3). La mucosa la forma principalment una monocapa de cèl·lules epitelials que presenten invaginacions conegudes com a criptes de Lieberkühn. Entre els diversos tipus cel·lulars, a la base de les criptes es localitzen les cèl·lules mare. Aquestes, són cèl·lules indiferenciades totipotents que proliferen i donen lloc a la resta de tipus cel·lulars. Es pensa que mutacions en aquestes cèl·lules amb la posterior invasió de tota la cripta mitjançant expansió clonal podria ser l'inici del CRC (Brittan and Wright 2004).

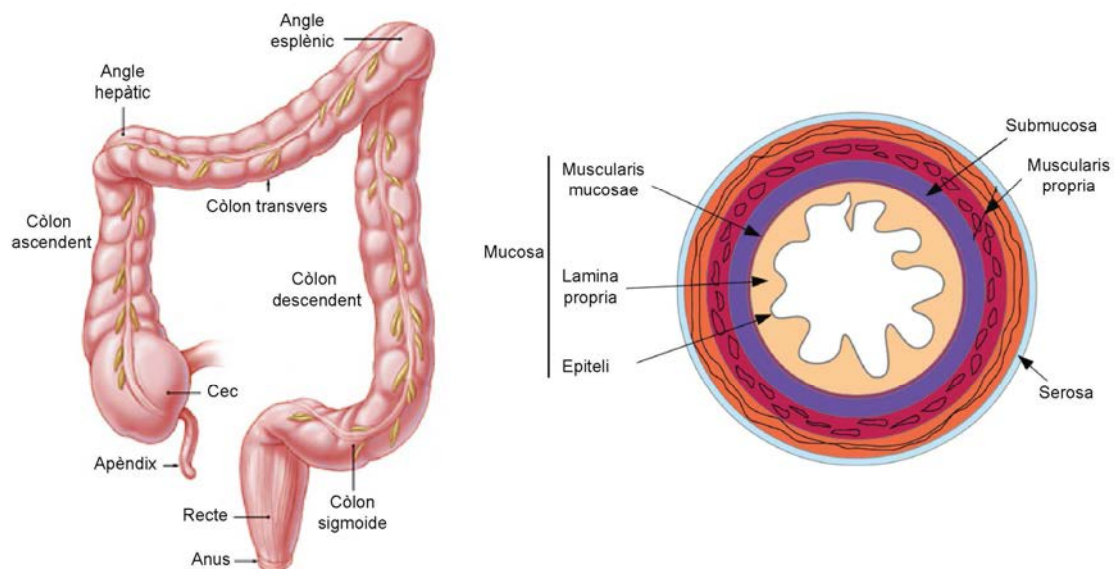


Figura 3. Representacions esquemàtiques de l'anatomia i de la secció transversal de l'intestí gros.

1.1.3 Classificació del CRC per estadiatges

La classificació del CRC per estadiatges és un dels factors més importants a l'hora d'avaluar les diverses opcions terapèutiques i poder determinar com d'efectiu ha estat el mateix tractament. Històricament, s'han usat diversos criteris per a classificar el CRC, com el de Dukes o el posteriorment modificat d'Astler-Coller (Astler and Coller 1954). Actualment, la classificació més usada és el sistema TNM, establert per l'*American Joint Committee on Cancer* (Fleming et al. 1997) (Taula 1). Aquest sistema té en compte tres factors per a classificar el CRC:

1. Tumor (T): S'estableix la progressió del càncer mitjançant l'avaluació del creixement a través de les diferents capes de l'intestí (T1-T4).

2. Nòduls limfàtics (N): Es valora la progressió del càncer mitjançant el comptatge de nòduls afectats (N0-N2).
3. Metàstasi (M): S'examina la progressió del càncer mitjançant el comptatge d'òrgans llunyans envaïts (M0-M1).

Taula 1. Classificació del CRC segons la classificació TNM.

Categories T	Grau d'infiltració a través de les capes de l'intestí
Tx	No hi ha informació disponible sobre la infiltració del tumor
T0	No hi ha evidència de tumor primari
Tis	Carcinoma <i>in situ</i> . El càncer es manté a la mucosa
T1	El tumor ha crescut a través de la muscularis mucosa i envaeix la submucosa
T2	El tumor ha crescut a través de la submucosa i envaeix la muscularis propria
T3	El tumor ha crescut a través de la muscularis propria i arriba a les capes més externes de l'intestí però sense travessar-les. No envaeix òrgans o teixits veïns
T4a	El tumor ha crescut a través de la serosa
T4b	El tumor ha crescut a través de la capa més externa del còlon o recte i s'adhereix o envaeix teixits i/o òrgans veïns
Categories N	Grau d'invasió a ganglis limfàtics veïns
Nx	No hi ha informació disponible d'invasió a ganglis limfàtics
N0	No hi ha invasió a ganglis limfàtics
N1a	Invasió a un gangli
N1b	Entre 2 i 3 ganglis afectats
N1c	El tumor es manté a la subserosa, mesenteri o a les capes més externes del còlon o recte sense ganglis regionals afectats
N2a	Entre 4 i 6 ganglis afectats
N2b	Més de 7 ganglis afectats
Categories M	Grau d'invasió a òrgans distants
M0	No hi ha metàstasi
M1a	Presència de metàstasi distant confinat a un òrgan o teixit
M1b	Presència de metàstasi distant a més d'un òrgan

Un cop establerta la classificació mitjançant el sistema TNM, i normalment posterior a la cirurgia, aquesta informació s'usa per a la classificació d'estadiatge global descrita a la Taula 2. Així el tractament del CRC varia en funció de l'estadiatge diagnosticat. En estadiatges primerencs (I i II), la cirurgia, per tal de extirpar el tumor, és el tractament més comú en còlon (98%) i recte (88%). Per pacient d'estadiatges III i alguns casos d'estadiatges II de còlon, la cirurgia es combina amb uns sis mesos aproximadament de quimioteràpia per tal de reduir el risc de recurrència. No obstant, el tractament amb quimioteràpia

en estadiatges II està força discutit. D'altra banda, els estadiatges II i III de recte es tracten amb quimioteràpia neoadjuvant combinat amb radioteràpia. Pel que fa a la supervivència als 5 anys, aquesta se situa al voltant del 90% en pacients on el tumor està localitzat. Un cop el càncer envaeix nòduls o òrgans propers la supervivència als 5 anys baixa al 70,4%. Finalment, quan la malaltia envaeix òrgans llunyans la supervivència als 5 anys baixa fins al 12,5% (DeSantis et al. 2014).

Taula 2. Estadiatges del CRC

Estadiatges	T	N	M
0	Tis	N0	M0
I	T1	N0	M0
	T2	N0	M0
IIA	T3	N0	M0
IIB	T4a	N0	M0
IIC	T4b	N0	M0
IIIA	T1-T2	N1/N1c	M0
	T1	N2a	M0
IIIB	T3-T4a	N1/N1c	M0
	T2-T3	N2a	M0
	T1-T2	N2b	M0
IIIC	T4a	N2a	M0
	T3-T4a	N2b	M0
	T4b	N1-N2	M0
IVA	T1-T4	N1-N2	M1a
IVB	T1-T4	N1-N2	M1b

1.1.4 Etiologia del CRC

El 95% dels carcinomes de còlon són adenocarcinomes, i per tant, aquests s'originen a les cèl·lules epitelials del còlon. La major part d'aquests tipus de càncer, al voltant d'un 70-80%, són esporàdics. D'altra banda, fins a un 25% dels pacients s'agrupen al tipus de CRC familiar, on en aquestes famílies la freqüència de desenvolupar la malaltia és superior a l'esporàdic, però sense seguir un patró característic del càncer hereditari. Dins d'aquest grup, tan sols entre un 5-6% dels casos estan associats a mutacions en gens d'alta penetrància que presenten una herència mendeliana dominant o recessiva (Figura 4). És en aquest grup de casos on s'agrupen les síndromes

polipòsiques, polipòsiques hemartomatoses i no polipòsiques que s'enumeren a la Taula 3.

Taula 3. Síndromes hereditàries de predisposició al CRC.

Síndrome	Incidència	Gen	Característiques
FAP	1:35000 – 1:50000	<i>APC</i>	Herència autosòmica dominant. Aparició de múltiples pòlips adenomatosos (>100) a còlon i recte al voltant dels 20-30 anys d'edat. Risc a patir CRC proper al 100%. Existeix la versió atenuada (AFAP), la qual presenta menys pòlips (entre 40-50) (Lynch et al. 2003).
Lynch/ HNPCC	1:500	<i>MLH1</i>	Herència autosòmica dominant. És el més freqüent de tots els CRC hereditaris, representant entre un 2-3% dels casos totals de CRC. Associat a mutacions en el gens de reparació de bases mal aparellades. S'han proposat múltiples criteris i revisions d'aquests per tal d'establir una normativa diagnòstica per aquesta síndrome i per l'anàlisi de MSI (Amsterdam I, II i Bethesda) (Vasen et al. 1991, 1999; Rodriguez-Bigas et al. 1997)
		<i>MSH2</i>	
		<i>MSH6</i>	
		<i>PMS2</i>	
		<i>EPCAM</i>	
MAP	1:2000 – 1:4000	<i>MUTYH</i>	Tipus de síndrome adenomatososa similar fenotípicament a la FAP/AFAP però amb herència autosòmica recessiva del gen <i>MUTYH</i> (Al-Tassan et al. 2002).
Poliposi juvenil	1:16000 – 1:100000	<i>SMAD4</i>	Herència autosòmica dominant. És la síndrome de poliposi hemartomatososa més freqüent. Aparició de pòlips en tot el tracte intestinal amb elevat risc de desenvolupar CRC (Desai et al. 1995)
		<i>BMPR1A</i>	
Peutz- Jeghers	1:50000 – 1:200000	<i>STK11</i>	Herència autosòmica dominant. Es caracteritza per l'aparició de múltiples pòlips hemartomatosos al llarg de tot el tracte gastrointestinal. Pigmentació mucocutània (Dunlop et al. 2002).
Cowden	1:200000	<i>PTEN</i>	Herència autosòmica dominant. Els pòlips hemartomatosos d'aquesta síndrome acostumen a presentar-se a la mama, tiroides, pell i sistema nerviós central, a part del tracte gastrointestinal (Schreibman et al. 2005).
Bannayan- Ruvalcaba- Riley	ND	<i>PTEN</i>	Herència autosòmica dominant. Els pòlips es localitzen principalment a la llengua i al còlon. Associació amb manifestacions extracolòniques, com macrocefàlia o retard mental. No s'ha descrit un risc de lesions neoplàsiques associades (Schreibman et al. 2005).
PPAP	ND	<i>POLE</i>	Recentment descrita. Herència autosòmica dominant. Polipòsica. Mutacions germinals al domini exonucleasa de dues polimerases. Fenotip hipermutador (Palles et al. 2012).
		<i>POLD1</i>	

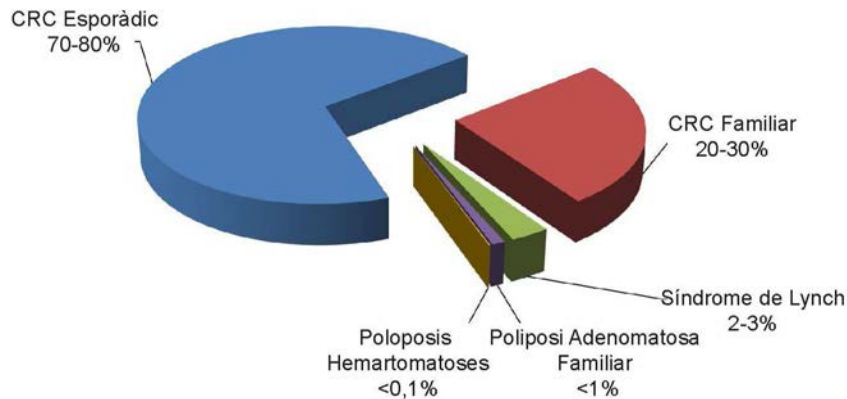


Figura 4. Classificació del CRC. Adaptada de Randall WB. Gastroenterology 2000.

1.2 Classificació molecular del CRC

Quan es resumeixen les característiques comunes que presenten les cèl·lules tumorals de qualsevol tipus de càncer s'acostuma a recórrer a les característiques descrites per Hanahan i Weinberg (Hanahan and Weinberg 2000). Aquest model, que inicialment englobava 6 característiques, s'ha revisat posteriorment donant lloc a l'actual model on es proposen 10 característiques comunes presents a la cèl·lula tumoral (Hanahan and Weinberg 2011). (Figura 5).

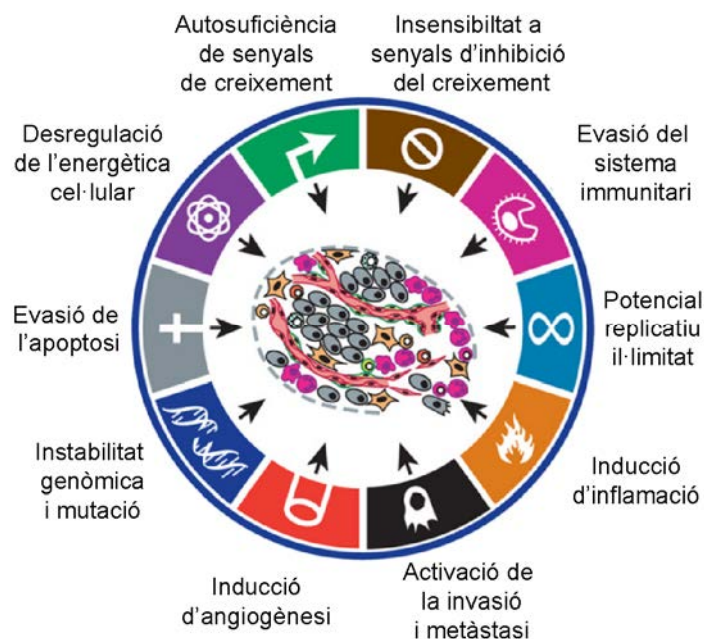


Figura 5. Característiques del càncer. Adaptada de Hanahan and Weinberg. Cell 2011.

L'adquisició de la inestabilitat genòmica és una de les principals característiques que explica el model de progressió del CRC, la seqüència adenoma-carcinoma descrita a principis dels anys 90 (Fearon and Vogelstein 1990). Les constants revisions del model, proposen que el CRC esdevé d'una progressiva transformació de l'epiteli colònic normal fins a l'adenocarcinoma, a través d'una orquestrada sèrie d'esdeveniments genètics i/o epigenètics, entre d'altres anomalies (Colussi et al. 2013) (Fig. 6). Els tipus d'alteracions moleculars que succeeixen durant la progressió són els que han definit, tradicionalment, les tres vies de progressió que encara s'usen a l'hora de classificar el CRC; la via del fenotip metilador d'illes CpG (CIMP), la inestabilitat de microsatèl·lits (MSI, de l'anglès *Microsatellite Instability*) i la inestabilitat cromosòmica (CIN, de l'anglès *Chromosome Instability*). Recentment, s'ha descrit una nova classificació del CRC, dividida en quatre subtipus moleculars (CMS, de l'anglès *Consensus Molecular Subtypes*). Les característiques que engloba cadascun d'aquests subtipus es descriuen a la Taula 4.

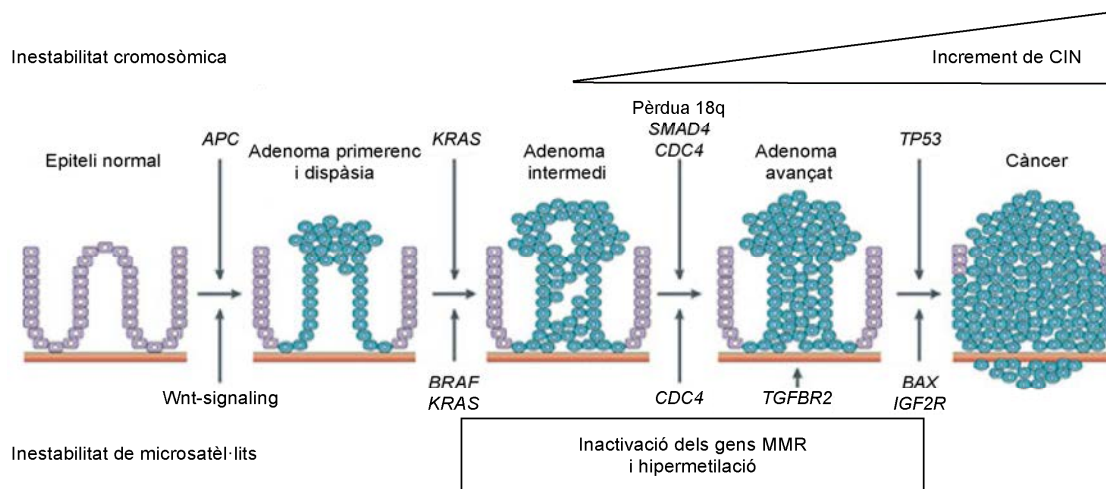


Figura 6. Seqüència adenoma-carcinoma. Adaptada de Walther, et al. Nature Reviews Cancer 2009.

Taula 4. Classificació CMS del CRC

CMS1 MSI Immune	CMS2 Canònica	CMS3 Metabòlica	CMS4 Mesenquimal
14%	37%	13%	23%
MSI, CIMP elevada, Fenotip hipermutador	Presència de CNAs	MSI intermitja, CIN baixa, CIMP baixa	Presència de CNAs
Mutacions a <i>BRAF</i>		Mutacions a <i>KRAS</i>	
Infiltració i activació immunitària	Activació de WNT i <i>MYC</i>	Desregulació metabòlica	Infiltració de l'estroma, activació de TGF β i angiogènesi
Pitjor supervivència després de recurrència			Elevada recurrència i pitjor supervivència

1.2.1 Via del fenotip metilador d'illes CpG (CIMP)

Les illes CpG són regions del DNA riques en dinucleòtids citosina-guanina. Aquestes regions s'han vist associades a gran part de les seqüències promotores del genoma humà. La metilació de les illes CpG al promotor d'un gen en concret pot comportar el silenciament d'aquest gen. Així, els tumors que segueixen la via CIMP del CRC es caracteritzen per presentar hipermetilació en illes CpG dels promotors de certs gens supressors de tumors (TSGs). Exemples clàssics d'aquesta hipermetilació són els que es donen als gens: *CDKN2A*, *MGMT* o *MLH1*.

1.2.2 Inestabilitat de microsatèl·lits (MSI)

Els microsatèl·lits són un tipus d'element repetitiu del DNA, també coneguts com a repeticions curtes en tàndem (STR, de l'anglès *short tandem repeats*) els quals són seqüències curtes, aproximadament entre 1-5 parells de bases i repetides des de 5 fins a 50 vegades. Aquestes seqüències, juntament amb altres (minisatèl·lits, satèl·lits), són repeticions en tàndem disposades al llarg de tot el genoma i que representen el 3% del total del genoma humà. A causa de la seva estructura repetitiva, aquestes seqüències són propenses a patir mutacions arrel d'errors en la replicació del DNA. Específicament la polimerasa llisca en aquestes regions repetitives i consegüentment es generen alteracions

de tipus insercions o delecions. Aquestes alteracions poden comportar canvis en el patró de lectura del DNA (de l'anglès, *frameshift mutacions*). Els tumors que segueixen la via MSI del CRC presenten inactivació a algun dels gens associats a la via de reparació de bases mal aparellades (MMR, de l'anglès *Mismatch Repair*). Les set proteïnes que defineixen el sistema DNA MMR són: MLH1, MLH3, MSH2, MSH3, MSH6, PMS1 i PMS2 (Boland and Goel 2010). Aquests tumors no reparen els errors de replicació produïts als microsatèl·lits, fet que comporta l'acumulació de mutacions *frameshift* en aquestes seqüències repetitives. Al voltant d'un 15% de tots els CRC mostren la via MSI (Boland and Goel 2010). La gran majoria d'aquests tumors són esporàdics (80-85%) amb presència d'hipermetilació somàtica al gen *MLH1* (Cunningham et al. 1998). La resta de casos (15-20%) corresponen a casos hereditaris amb síndrome de Lynch (Boland and Goel 2010). Els tumors que segueixen aquesta via tendeixen a presentar cariotips propers a ser diploides, fenotips hipermutadors i manifestar-se en el còlon proximal.

1.2.3 Inestabilitat cromosòmica

La inestabilitat cromosòmica (CIN) es defineix com l'increment de la taxa de guanys i pèrdues de cromosomes sencers o regions cromosòmiques. (Lengauer et al. 1998). El terme CIN numèrica fa referència als guanys i pèrdues de cromosomes sencers, mentre que la CIN estructural engloba les alteracions que resulten en guanys i pèrdues de regions subcromosòmiques. La conseqüència més evident d'ambdós tipus de CIN al càncer és la de presentar cariotips molt variables entre cèl·lules del mateix tumor.

1.2.3.1 CIN numèrica: Aneuploïdia

L'aneuploïdia és l'estat en que un cariotip presenta un o més d'un guany i/o pèrdua d'un cromosoma sencer i, per tant, difereix de contenir el nombre de cromosomes que presenta una cèl·lula diploide normal ($N=46$). Tot i que l'aneuploïdia s'ha associat al càncer des de fa més d'un segle, si aquest estat cel·lular és *per se* una causa del càncer o una conseqüència del fenotip CIN que presenta la malaltia encara està en discussió (Hansermann 1890; Boveri 1902; Holland and Cleveland 2009). Les causes de l'aneuploïdia són múltiples, tot i que, una fracció força important es deu a errors de segregació en mitosi,

especialment en aquells tumors que mostren genomes propers a ser diploides (Holland and Cleveland 2009). S'ha estimat que aquest tipus de CIN resulta en un guany o una pèrdua cada cinc divisions mitòtiques, més o menys (Lengauer et al. 1998). A continuació es descriuen els diversos mecanismes causants de l'aneuploïdia englobats a la Figura 7.

Alteracions al punt de control de la mitosi (a): Les proteïnes que formen part d'aquest punt de control s'encarreguen d'assegurar que existeix una unió correcte entre els cinetocors i els microtúbuls del fus acromàtic, en la transició de metafase a anafase. Alteracions en aquestes proteïnes són la causa que la mitosi progressi normalment en casos d'unions incorrectes (Thompson et al. 2010).

Defectes en la cohesió de cromàtides germanes (b): Quan la cohesió entre cromàtides germanes persisteix en l'anafase, ambdues cromàtides poden segregar al mateix pol (Holland and Cleveland 2009). Aquest esdeveniment es coneix amb el terme de no-disjunció.

Defectes en la unió cinetocor-microtúbul (c): En condicions normals els cinetocors de cada cromàtide s'uneixen als microtúbuls del fus d'un únic pol. Les unions merotèliques es descriuen quan els cinetocors s'uneixen a microtúbuls d'ambdós pols. En aquestes ocasions s'observa un enderrament de la cromàtide implicada a l'anafase, donant lloc a la possible pèrdua de la mateixa (Nicholson and Cimini 2011). Aquest esdeveniment es coneix amb el terme de pèrdua anafàsica.

Presència de centrosomes supernumeraris (d): Aquest últim mecanisme és comú després d'una tetraploidització, on la formació de divisions multipolars transitòries, amb posteriors agrupacions de centrosomes formant divisions pseudo-bipolars, augmenta la freqüència d'unions merotèliques (Ganem et al. 2009; Silkworth et al. 2009).

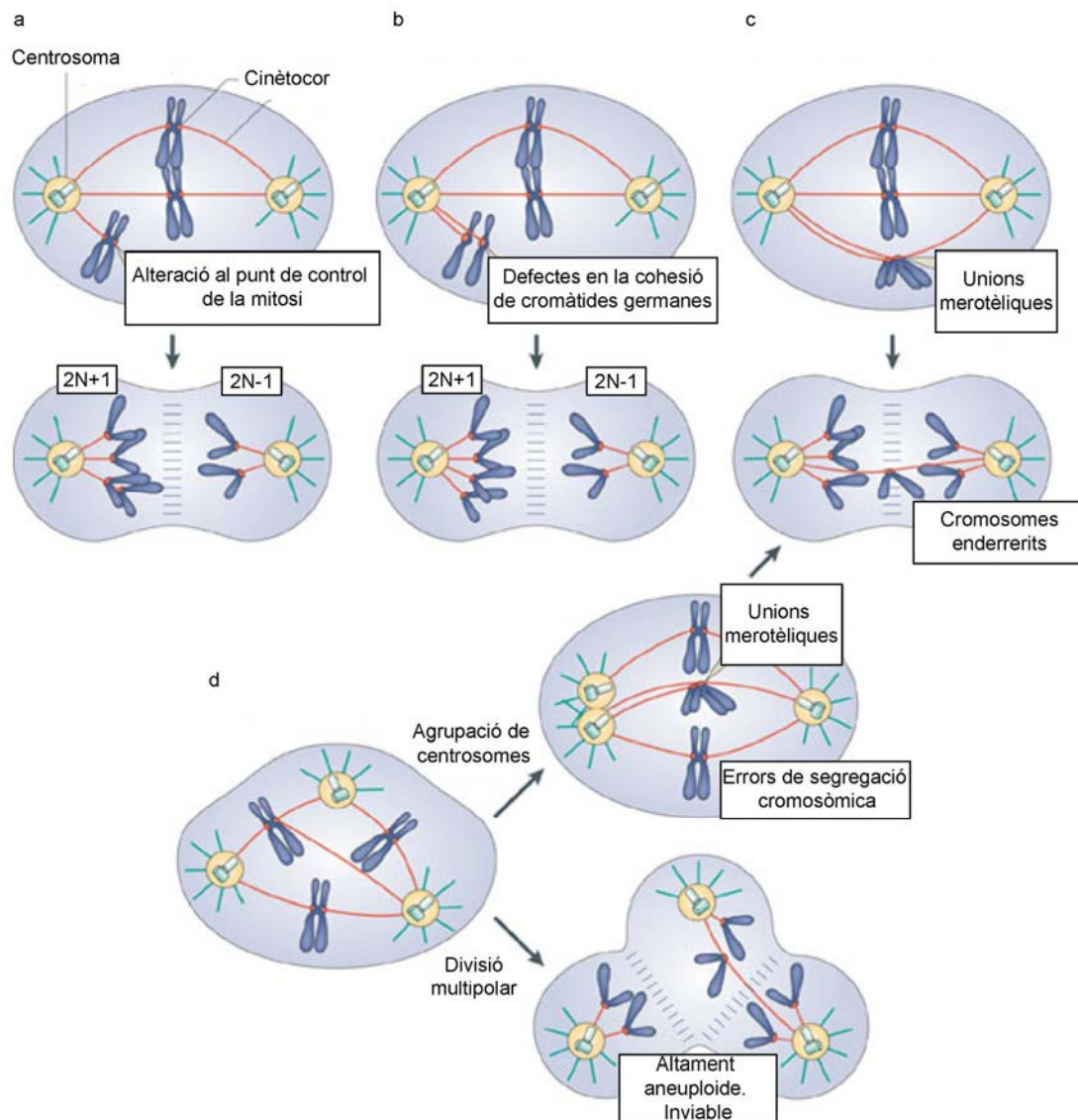


Figura 7. Mecanismes de formació d'aneuploidia. Adaptada de Holland and Cleveland. Nature reviews. Molecular cell biology 2009.

1.2.3.2 Tetraploidia

S'ha discutit àmpliament a la literatura el possible paper funcional que tindria la tetraploidització del genoma amb la posterior i consecutiva pèrdua de diferents cromosomes generant genomes altament aneuploides (Shackney et al. 1989; Storchova and Pellman 2004). Diferents evidències reforcen la hipòtesi de la tetraploidització del genoma tumoral com a un dels esdeveniments principals pels quals la cèl·lula tumoral esdevé inestable. Entre aquestes evidències es troba la presència d'una distribució bimodal en el nombre de cromosomes, on molts tipus de càncers mostren tumors propers a ser diploides i d'altres altament aneuploides, propers a ser triploides o tetraploides (Storchova and

Kuffer 2008). El CRC, entre d'altres tipus de càncers, mostra freqüències molt elevades d'aquests tumors altament aneuploides (Zack et al. 2013a). Una altra evidència és la detecció de centrosomes supernumeraris en línies cel·lulars i tumors primaris, que com prèviament s'ha descrit, resulten en múltiples aneuploidies a causa de divisions pseudo-bipolars i unions merotèliques (Ganem et al. 2009; Silkworth et al. 2009; del Rey et al. 2010). A més, s'han descrit cèl·lules properes a un genoma tetraploide en estadiatges primerencs de càncer de cèrvix i en la progressió de la lesió pre-maligne de Barrett cap a l'adenocarcinoma d'esòfag, on finalment s'observen cariotips altament aneuploides (Galipeau et al. 1996; Olaharski et al. 2006). Aquests cariotips altament aneuploides que mostren certs tipus tumorals són difícilment explicables pel simple guany de cromosomes en mitosi. Diversos autors reforcen aquesta hipòtesi, suggerint que l'estat intermedi del genoma tetraploide, al qual s'hi arriba per una duplicació de tot el genoma, és altament inestable, i mitjançant una elevada taxa de pèrdues cromosòmiques, el tumor assoleix un estat més estable confinat a un genoma altament aneuploide (Baia et al. 2008; Zack et al. 2013a). La tetraploidització de la cèl·lula tumoral es pot formar per un d'aquests tres mecanismes: fusió cel·lular, manca de citocinesi o de la mitosi (Davoli and de Lange 2011) (Figura 8).

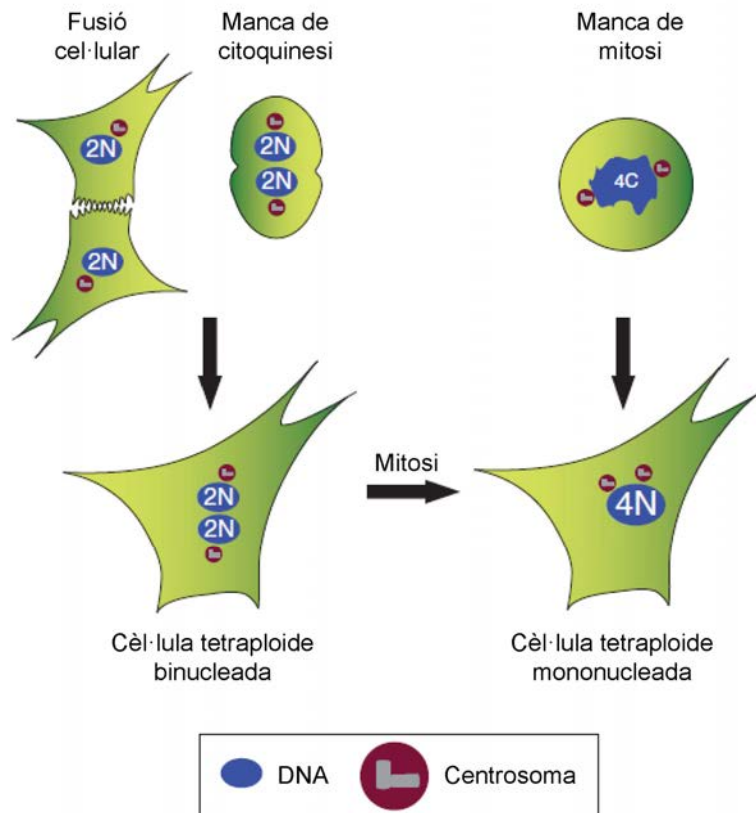


Figura 8. Mecanismes de tetraploidització. Adaptada de Storchova and Kuffer. Journal of Cell Science 2008. En les vies de fusió cel·lular i manca de citoquinesi es generen cèl·lules binucleades que contenen dos centrosomes. Aquestes cèl·lules poden formar cèl·lules tetraploides mononucleades després de dividir-se en la següent mitosi. L'evasió de la mitosi és una adaptació cel·lular després d'una arrest en mitosi permanent. Les cèl·lules passen per anafase, telofase i citoquinesi i progressen en el següent cicle a G1 sense corregir l'error mitòtic que ha produït l'aturada a mitosi. Les cèl·lules que deriven d'aquesta via presenten un nucli tetraploide acompanyat de dos centrosomes. 2N, nucli diploide; 4N, nucli tetraploide; 4C, nucli diploide amb cromosomes replicats.

Mutacions en diversos gens, entre els que destaca *APC*, s'han associat a tetraploidització en càncer (Dikovskaya et al. 2007; Caldwell et al. 2007). La conseqüència més directa d'aquesta tetraploidització és permetre al tumor tenir un fenotip mutador més elevat, on les delecions, pèrdues cromosòmiques, mutacions inactivadores, entre d'altres alteracions, seran permeses més fàcilment que en tumors diploides on la pèrdua d'un cromosoma sencer pot ser letal. Aquesta hipòtesi ha estat validada en llevats on s'han comparat genomes diploides contra haploides (Thompson et al. 2006). D'aquesta manera, mitjançant la duplicació del genoma s'espera que el tumor augmenti la probabilitat d'obtenir canvis favorables des d'un punt de vista selectiu

(Shackney et al. 1989; Storchova and Pellman 2004; Davoli and de Lange 2011). De fet, la tetraploidia i, en conseqüència, l'elevat grau d'aneuploidia s'han relacionat amb l'agressivitat del tumor, pitjor prognosi i generació de metàstasi (Camps et al. 2004; Gerlinger et al. 2012).

1.2.3.3 Inestabilitat cromosòmica estructural

Un altre tipus d'inestabilitat cromosòmica, característica dels cariotips tumorals, és la que dona lloc a alteracions conegudes com estructurals (SV, de l'anglès *Structural Variation*). Aquestes variants estructurals es divideixen en alteracions equilibrades, que no comporten ni guany ni pèrdua de material genètic, i alteracions desequilibrades (Lupski 2015). Entre les alteracions equilibrades es troben les translocacions recíproques, inversions, insercions. Tot i que gran part d'aquestes alteracions no acostumen a tenir un paper funcional per al desenvolupament del tumor, algunes fusions gèniques resultants de translocacions recíproques s'han descrit per ser claus en la progressió de certs tumors hematològics. Un exemple molt conegut és la formació del cromosoma Philadelphia, resultat de la translocació recíproca entre els cromosomes 9 i 22 en leucèmia mieloide crònica (CML, de l'anglès *Chronic Myelogenous Leukemia*). D'altra banda en CRC, entre d'altres tumors sòlids, s'observa una alta freqüència d'alteracions desequilibrades, el que condueix a aquest genomes a presentar diverses delecions i amplificacions de regions cromosòmiques (Cancer Genome Atlas Network et al. 2012). Juntament amb l'aneuploidia, descrita prèviament, aquestes anomalies es coneixen com alteracions en nombre de còpies (CNAs, de l'anglès *Copy Number Alterations*).

Els estudis de càncer, amb l'ús de grans cohorts, han portat a l'observació de diferents perfils de CNAs específics per tipus tumoral (Zack et al. 2013a). El gran repte que existeix actualment referent a l'estudi de CNAs en càncer és el de distingir entre aquells esdeveniments que contribueixen en el desenvolupament, manteniment o progressió del tumor de les CNAs que podríem definir com eventuais i que no contribueixen en el procés de la neoplàsia (Beroukhim et al. 2010a). En molts casos, regions focals de CNAs, per un tipus tumoral concret, s'han associat a l'afectació de gens importants per al desenvolupament del mateix càncer. Així, regions d'amplificació recurrents

acostumen a afectar oncogens, gens que mitjançant la seva sobre-activació poden comportar una proliferació descontrolada. Contràriament, en regions cromosòmiques de pèrdua s'hi localitzen, generalment, gens supressors de tumors (TSGs, de l'anglès *Tumor Suppressor Genes*). Aquests gens, a diferència dels oncogens, tenen un caràcter recessiu i per tant requereixen de la inactivació dels seus dos al·lells per tal que aportin un paper funcional al desenvolupament del càncer. Basat en la teoria dels *two-hit* de Knudson, un TSG pot estar inactivat en un dels seus dos al·lells, a causa d'una mutació inactivant germinal o somàtica (*first hit*). Posteriorment, podria adquirir una segona mutació somàtica inactivant a l'al·lel normal, el que suposaria el *second hit* del model (Knudson 1971). Altres alteracions que s'inclouen en aquest *second hit* podrien ser: inactivació per metilació o una pèrdua del cromosoma sencer o regió cromosòmica que contingui l'al·lel normal. D'aquesta manera, l'estudi de regions recurrents de pèrdua al càncer ha estat clau per la identificació de nous TSGs específics per tipus tumoral.

No obstant, durant els últims anys la disomia uniparental (UPD, de l'anglès *Uniparental Disomy*), un esdeveniment prèviament descrit a nivell constitucional, ha esdevingut clau en la pèrdua de funció de TSGs (Tuna et al. 2009). Seguint el model de Knudson, la UPD actua com a mecanisme alternatiu a la monosomia o deleció per inactivar un TSG. Mitjançant una pèrdua del cromosoma que conté l'al·lel normal i la posterior duplicació del cromosoma que conté l'al·lel mutat, la cèl·lula esdevé disòmica però amb la presència d'una mutació inactivant i en homozigosi en un TSG.

1.3 Disomia uniparental (UPD)

1.3.1 UPD constitucional

La disomia uniparental es va descriure per primera vegada l'any 1980 pel Dr. Eric Engel, el qual va definir que la UPD succeïa a l'heretar dues còpies cromosòmiques del mateix progenitor, principalment a causa d'errors en meiosi (Engel 1980b). Aquest esdeveniment és una de les causes de diverses manifestacions clíniques associades a set grans síndromes: Prader-Willi, Angelman, Bechwith-Wiedeman, Silver-Russell, diabetis neonatal transitòria, UPD materna del cromosoma 14 (síndrome de Temple) i UPD paterna del

cromosoma 14 (Liehr 2010). Així, el concepte de UPD va quedar molt relacionat a un esdeveniment patològic associat a nivell constitucional. D'entrada, cal diferenciar entre la UPD en heterodisomia (UPHD, de l'anglès *Uniparental Heterodisomy*), resultat de l'herència dels dos cromosomes homòlegs provinents de la mateixa via parental, i la isodisomia (UPID, de l'anglès *Uniparental Isodisomy*), resultat de l'herència del mateix cromosoma homòleg com a nova parella d'homòlegs (Tuna et al. 2009; Makishima and Maciejewski 2011; Lapunzina and Monk 2011) (Figura 9).

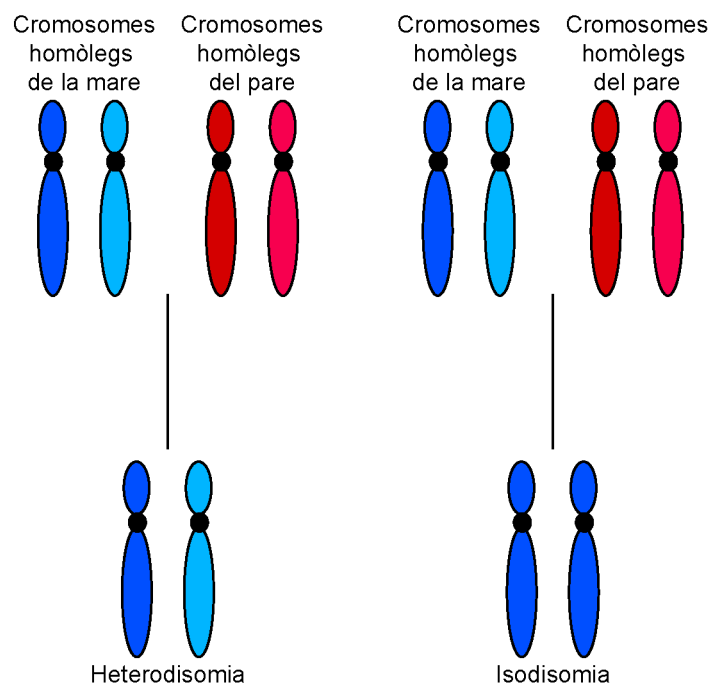


Figura 9. Heterodisomia i isodisomia. En l'heterodisomia s'hereten dos cromosomes homòlegs provinents de la mateixa via parental. En l'isodisomia s'hereta el mateix cromosoma.

Els dos tipus de UPD poden donar lloc a malalties associades a regions d'empremta genòmica amb pèrdua o guany d'empremta (LOI o GOI, de l'anglès *Loss Of Imprinting* o *Gain Of Imprinting*). A més, la UPID té com a conseqüència una pèrdua d'heterozigotat (LOH, de l'anglès *Loss Of Heterozygosity*), el que pot comportar el desenvolupament de malalties associades a gens recessius (Fu et al. 2014). En els casos on la patogenicitat de la UPD succeeix en regions on es localitzen gens d'empremta, s'ha de tenir en compte que la seva expressió és monoal·lèlica. Per exemple, la síndrome de Prader-Willi està associada, entre d'altres anomalies, a la UPD del

cromosoma matern implicant la regió cromosòmica 15q11.13, mentre que la síndrome d'Angelman està associada a la mateixa regió però amb el cromosoma patern duplicat (Cassidy and Schwartz 1998). Un altre exemple és el de la UPD paterna a la regió 11p15, associada a la síndrome de Beckwith-Wiedemann. Aquests pacients pateixen d'hemihiperplàsia localitzada. En aquesta regió es troba un dels gens empremtats més conegut com és *IGF2*. Diversos estudis han descrit que els pacients que mostren aquesta alteració pateixen un risc elevat d'adquirir càncer infantil com el tumor de Wilm's, hepatoblastoma, carcinoma adrenocortical i neuroblastoma, entre d'altres (Shuman et al. 2006; Bertoin et al. 2014).

Recentment en població normal, tres meta-anàlisis han identificat freqüències properes a l'1% de presentar UPDs en mosaic, entre d'altres anomalies cromosòmiques (Rodríguez-Santiago et al. 2010; Jacobs et al. 2012; Machiela et al. 2015). El possible paper funcional d'aquests mosaicismes constitucionals és desconegut, tot i que es suggereix que apareixen com a conseqüència d'un deteriorament del DNA causat per l'edat (Jacobs et al. 2012; Machiela et al. 2015).

1.3.2 Mecanismes de formació de UPD constitucional

Els mecanismes de formació de la UPD són diferents en funció de si l'anomalia afecta cromosomes sencers o segments cromosòmics (Tuna et al. 2009; Makishima and Maciejewski 2011; Lapunzina and Monk 2011). Els mecanismes de formació de la UPD de cromosomes sencers inclouen (Figura 10):

Complementació gamètica (a): Teòricament és l'opció més simple per crear una UPD de cromosoma sencer, no obstant és la menys probable, ja que seria necessari l'error del mateix cromosoma en els gàmetes dels dos pares. Així, la UPD esdevé de la fecundació d'un gàmeta disòmic i un altre nul·lisòmic,

Rescat trisòmic (b): És el mecanisme més comú de formació de UPDs. El rescat trisòmic tindria lloc en estadis postzigòtics molt primerencs, on es perdria un cromosoma mitjançant pèrdua anafàsica.

Rescat monosòmic (c): En aquest cas, el rescat és d'un zigot monosòmic que prové de la fecundació d'un gàmeta monosòmic i d'un altre nul·lisòmic generats

a meiosi. Un esdeveniment de no-disjunció postzigòtic acabaria formant la UPD.

Errors en mitosi (d): Esdeveniments consecutius de no-disjunció i pèrdua anafàsica en casos monosòmics o trisòmics, respectivament.

A més dels quatre mecanismes prèviament descrits, existeixen diferents reordenaments complexos que podrien, potencialment, formar UPDs (Liehr 2010).

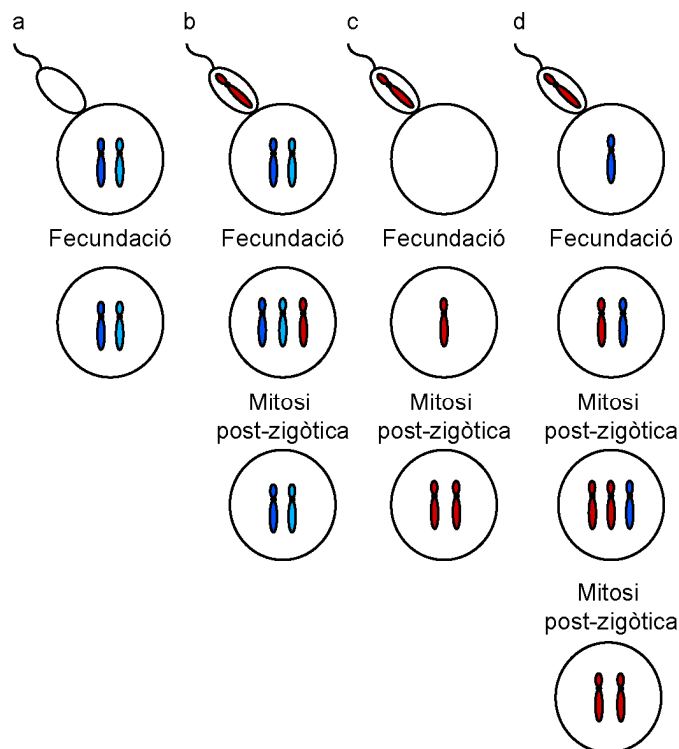


Figura 10. Mecanismes de formació de UPD de cromosoma sencer. (a) Complementació gamètica; (b) rescat trisòmic; (c) rescat monosòmic; (d) errors en mitosi.

D'altra banda, la UPD segmental podria sorgir com a conseqüència d'una recombinació homòloga postzigòtica, entre l'homòleg matern i patern, fet que comportaria l'adquisició d'una regió cromosòmica en isodisomia (Kotzot 2008) (Figura 11).

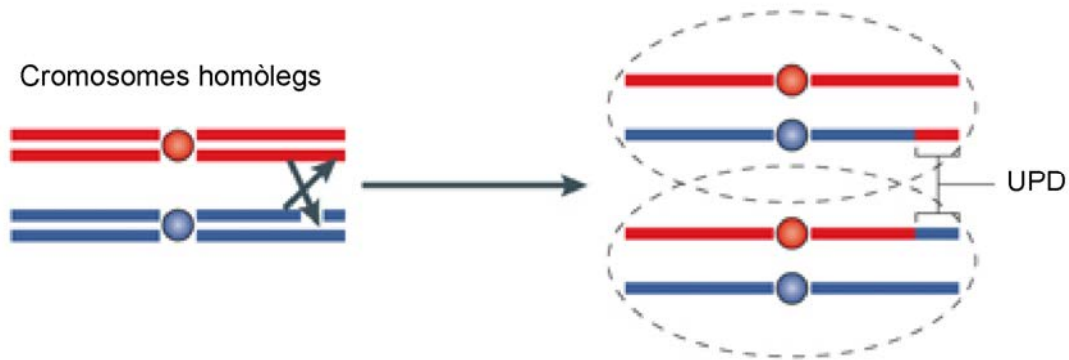


Figura 11. Formació de UPD segmental mitjançant recombinació homòloga. Figura adaptada de Moynahan and Jasin. *Nature Reviews* 2010. La recombinació homòloga pot donar lloc a una UPD telomèrica quan existeix un entrecreuament i posteriorment les cromàtides recombinants segreguen a pols oposats.

1.3.3 UPD i càncer

Arrel de la millora tecnològica de genotipatge (SNP arrays) i la subseqüent anàlisi a gran escala de diferents tipus de tumors, s'ha identificat la UPD com a esdeveniment recurrent en càncer (Tuna et al. 2009; Makishima and Maciejewski 2011). Davant l'increment de publicacions en els darrers anys, cal fer èmfasi en la terminologia usada a la literatura. D'una banda, diversos grups opten per usar el terme *copy-neutral LOH* (cnLOH) o simplement, LOH, per tal de definir aquesta alteració. Altres grups opten per mantenir la terminologia de UPD, sovint amb l'addició de termes que defineixen la naturalesa somàtica de l'esdeveniment com: *acquired UPD* o *somatic UPD*. És important tenir en compte que la LOH pot estar associada també a un guany en nombre de còpies, terme definit com a *copy-gain LOH* (cgLOH) o *uniparental polysomy* (UPP) (Zarzour et al. 2015). A més, en els casos on el genoma tumoral és poliploide (triploide o ploïdies superiors), una cnLOH pot correspondre a casos amb tres cromosomes o més, i per tant, tècnicament no es poden definir com a UPDs. És en aquests casos que també s'adopta la terminologia UPP. A mode de simplificació en aquesta tesi s'utilitzarà el terme UPD.

Múltiples estudis han identificat patrons específics de UPD per tipus tumoral. Alhora, aquestes regions específiques de UPD s'han associat amb la inactivació de TSGs o l'activació d'oncogens. Per exemple, s'ha descrit que un 20% dels pacients de leucèmia mieloide aguda (AML, de l'anglès *Acute Myeloid Leukemia*) presenten UPD de manera freqüent (Raghavan et al. 2005;

Tyybäkinoja et al. 2008). A més, en les regions recurrents de UPD s'han identificat mutacions en quatre gens diferents, inclosos *WT1*, *FLT3*, *CEBPA* i *RUNX1* (Fitzgibbon et al. 2005). En estudis de grans cohorts de pacients amb diversitat de mielodisplàsies s'han trobat freqüències de UPD d'entre el 20 i el 35% (Gondek et al. 2008). Un altre estudi en síndrome mielodisplàsica (MDS, de l'anglès *Myelodysplastic Syndrome*), mostra que dos dels subgrups que conformen aquesta neoplàsia, els que presenten la deleció 5q i els que no la tenen, mostren altes freqüències de UPD (Wang et al. 2008). Alguns estudis descriuen que els pacients amb tumors mieloides amb presència de UPDs a les regions cromosòmiques 7q, 11q o 17p presenten pitjor pronòstic (Gondek et al. 2008; Makishima et al. 2009; Jasek et al. 2010). En leucèmia limfoblàstica aguda (ALL, de l'anglès *Acute Lymphoblastic Leukemia*) també s'han identificat regions freqüents de UPD, especialment a la regió cromosòmica 9p, on s'observa recurrentment afectat el gen *CDKN2A* (Mullighan et al. 2007; Kuiper et al. 2007; Kawamata et al. 2008b). No obstant, un altre estudi proposa la presència potencial d'altres gens importants en la regió, ja que la mutació de *CDKN2A* no s'observa en cap dels pacients amb UPD a 9p (Sulong et al. 2009). Inicialment, la UPD al braç cromosòmic 9p es va detectar en un 33% dels pacients amb policitèmia vera. Posteriorment, es va associar aquesta alteració amb la mutació en homozigosi amb guany de funció de *JAK2V167F* (Kralovics et al. 2002, 2005). Des de llavors, diversos estudis de genotipatge en diferents síndromes mieloproliferatives han identificat UPDs recurrents en aquesta regió, constituint-se d'aquesta manera en una alteració característica d'aquests tumors (Jones et al. 2005; Yamamoto et al. 2007; Kawamata et al. 2008a; Wang et al. 2016). Un altre exemple de UPD característica de tipus tumoral és la que es troba en leucèmia limfoblàstica crònica (CLL, de l'anglès *Chronic Lymphocytic Leukemia*) on gran part del cromosoma 13 mostra LOH, amb una petita deleció en homozigosi a 13q14, on s'hi localitzen els microRNAs miR-15a/miR-16-1 (Pfeifer et al. 2007; Lehmann et al. 2008). En tricoleucèmia, un subtipus de CLL, s'han identificat UPDs afectant la mateixa regió del cromosoma 13 (Forconi et al. 2008). En leucèmies associades a neurofibromatosi de tipus I, s'han descrit UPD segmentals i intersticials a 17q afectant *NF1* (Stephens et al. 2006; Flotho et al. 2007). A més d'aquesta alteració, també s'ha detectat una UPD recurrent a la regió cromosòmica del

11q, associada amb mutacions a *CBL* en casos de leucèmia mielomonocítica juvenil (Loh et al. 2009; Steinemann et al. 2010). Un cas interessant és el que presenta el mieloma múltiple, on s'han descrit UPDs freqüents afectant diverses regions petites al llarg del genoma, només en un subgrup de pacients concrets, els quals no presenten la translocació t(11;14) (Walker et al. 2006). Aquest treball suggereix que la freqüència diferencial de UPDs que existeix entre els dos subgrups pot influir en la prognosi dels pacients (Walker et al. 2006). D'altra banda, en limfoma de cèl·lules del mantell s'ha descrit una associació entre la presència de la translocació t(11;14)(q13;q32) amb diverses alteracions secundàries, inclosa la UPD a la regió del braç curt del cromosoma 17 amb afectació a *TP53* (Bea et al. 2009). En limfoma fol·licular i fol·licular transformat, també es descriuen UPDs afectant, principalment regions cromosòmiques 9p (*CDKN2A*) i 17p (*TP53*) (Fitzgibbon et al. 2007).

Pel què fa a tumors sòlids, el retinoblastoma va ser el primer tipus tumoral en el qual se li van descriure regions de UPD, en aquest cas involucrant el cromosoma 13 on es localitza el TSG *RB1* (Hagstrom and Dryja 1999). En càncer de mama s'han identificat UPDs, amb freqüències properes al 10%, en regions on es localitzen dos TSGs, *RB1* i *TP53* (Murthy et al. 2002). A més, s'han reportat freqüències més elevades de UPD en tumors de mama i ovari associats a mutacions germinals de *BRCA1* comparat amb casos esporàdics (Johnson et al. 2008; Walsh et al. 2008). En adenocarcinoma d'esòfag s'ha descrit que la UPD és un esdeveniment molt freqüent i que afecta principalment als braços cromosòmics 2p, 9q i 17p (Nancarrow et al. 2008; Frankel et al. 2014). D'altra banda, al càncer d'esòfag escamós s'observen regions de UPD diferents a les descrites per l'adenocarcinoma, mostrant els braços cromosòmics 2p, 5p, 19p, i 20q com els més freqüents (Hu et al. 2010). En carcinoma escamós de boca, un tipus de càncer de cap i coll, s'ha observat una freqüència elevada, 90% dels casos, de UPDs a la regió 16p11.2 (Morita et al. 2016). En càncer renal de cèl·lules clares la freqüència de UPDs és més baixa amb algun cas concret a la regió del 3q i cromosomes 6 i 7 (Toma et al. 2008). Recentment, en càncer gàstric s'han identificat regions recurrents de UPD als braços cromosòmics 5q, 11q, i 12q (Arakawa et al. 2017). Els autors destaquen múltiples gens candidats en aquestes regions. En tumors

pancreàtics s'ha identificat una freqüència de UPDs propera al 15%, afectant molt específicament les regions 4q22.3-q23, 4q31.21-q31.23 i 18q21.1 (Harada et al. 2008). En carcinoma de cèl·lules basals les UPDs afecten les regions del 6q i 9q (Teh et al. 2005). En hepatoblastoma la UPD afecta principalment la regió 11p15 on es localitza *IGF2* (Suzuki et al. 2008). En neuroblastoma s'observa la regió del 9p, afectant el TSG *CDKN2A* (Carén et al. 2008). Finalment, en glioblastoma també s'observen UPDs recurrents a 17p, afectant *TP53* (Lo et al. 2008).

1.3.3.1 UPD i càncer colorectal

Referent al CRC, diversos estudis han descrit la UPD com a un esdeveniment freqüent en tipus esporàdics, mentre que la mateixa alteració és molt menys present en CRC amb MSI (van Puijenbroek et al. 2008; Melcher et al. 2011). El mateix grup identifica les regions cromosòmiques 5q21.1-q22.3 i 17p13.2-p13.1 com a les més freqüents, 31% i 25% respectivament, de presentar UPD en CRC esporàdic (Melcher et al. 2011). Les dues regions contenen dos TSGs molt importants en les vies de progressió de CRC, com són *APC* i *TP53*. A més d'aquestes dues regions, un altre estudi que usa una cohort força gran de tumors primaris descriu com a UPDs freqüents els cromosomes 11, 14 i 15 i els braços cromosòmics 8p i 18q (Kurashina et al. 2008). Els autors d'aquest article destaquen *SMAD4*, a més de *APC* i *TP53*, com a gens involucrats en esdeveniments de UPD. De fet, aquests tres TSGs s'han descrit per estar recurrentment afectats per esdeveniment de UPD en dos estudis realitzats en múltiples línies cel·lulars de CRC (Gaasenbeek et al. 2006; Melcher et al. 2007). Tanmateix, altres estudis observen perfils de UPD en CRC diferents als descrits fins ara. Per exemple, Andersen i col·laboradors observen regions recurrents de UPD als braços cromosòmics 3p, 13q, 15q, 20p i 22q i descriuen com a perdudes regions definides com a UPDs en els estudis descrits anteriorment (Andersen et al. 2007). Un altre exemple és l'estudi d'Eldai i col·laboradors, els quals detecten regions curtes de UPD repartides per diferents cromosomes (Eldai et al. 2013). Cal destacar que aquests dos últims estudis analitzen tan sols 15 casos cadascun. Finalment, l'existència de treballs on es comparen adenomes i carcinomes aporten nous coneixements sobre la UPD com a esdeveniment important en la progressió d'adenoma cap a

carcinoma. Lips i col·laboradors van estudiar una cohort de 78 mostres corresponents a adenomes i carcinomes de recte i van observar que la UPD al braç cromosòmic 5q era un dels esdeveniments més comuns en ambdós grups (Lips et al. 2007). Recentment, un estudi similar confirma l'alta freqüència de UPD al braç cromosòmic 5q en adenomes i carcinomes, suggerint que aquesta alteració exhibeix un paper important en la progressió adenoma-carcinoma mitjançant la inactivació d'APC (Zarzour et al. 2015). A més d'aquesta regió, també observen freqüències elevades de UPD al braç cromosòmic 17p en adenomes i carcinomes, juntament amb diferents regions involucrades en empremta genòmica (11p, 14q, 15q i 22q). Finalment, un estudi on es comparen els perfils de tumors i els de mucosa normal confirmen que la regió del 5q és la UPD més freqüent al CRC (Conconi et al. 2016). Mentre que algunes CNAs i la UPD del cromosoma X ja s'observen en la mucosa, la UPD a 5q no s'observa en teixit aparentment normal.

1.3.4 Mecanismes de formació de UPD somàtica

La UPD somàtica de cromosoma sencer es forma principalment per errors de segregació en mitosi (Tuna et al. 2009; Makishima and Maciejewski 2011). Com ja s'ha descrit amb anterioritat, aquests tipus d'errors són freqüents en càncer. Així, esdeveniments consecutius de no-disjunció i/o pèrdua anàfàsica poden conduir a la cèl·lula a perdre un cromosoma en una divisió i duplicar l'homòleg en divisions posteriors. D'altra banda, similar al cas de la UPD constitucional, s'ha descrit que el mecanisme de formació de la UPD somàtica que involucra un segment cromosòmic és la recombinació homòloga en mitosi (Tuna et al. 2009).

La recombinació homòloga és un mecanisme especialitzat en la reparació de DSBs (Moynahan and Jasin 2010). Aquests DSBs s'originen a causa d'agents externs, com la radiació ionitzant, o també espontàniament arrel d'errors en la replicació del DNA. La recombinació homòloga, com bé expressa el terme, usa una molècula homòloga de DNA per tal de reparar el DSB creat en l'altra molècula. D'aquesta manera, la recombinació homòloga succeeix principalment en la fase G2 del cicle cel·lular, quan la cèl·lula disposa d'una altra cromàtide idèntica, la cromàtide germana, per usar com a motlle. Tanmateix, seqüències

homòlogues corresponents al cromosoma homòleg també són motlles potencials per a la reparació (Moynahan and Jasin 1997; Neuwirth et al. 2007). La regió afectada per UPD dependrà dels punts de recombinació que prenguin part en l'esdeveniment de reparació. Finalment, la UPD esdevindrà si les cromàtides recombinants segreguen a pols oposats. Alhora, la conversió gènica també s'ha descrit com a potencial creadora de UPDs segmentals (Tyybäkinoja et al. 2008). No obstant, aquest mecanisme de la via de l'HR tan sols és capaç de crear petites regions intersticials de UPD.

Finalment, s'ha descrit un altre mecanisme de formació de UPDs segmentals, de la via de l'HR, anomenat *Break Induced Replication* (BIR). En aquest cas el DSB a reparar apareix quan el DNA està replicant. Així, la via BIR inicia una aturada de la replicació causada per l'estancament de la polimerasa en el DSB. El complex que conté la polimerasa estancada, salta de la cromàtide on es localitza i usa la cromàtide del cromosoma homòleg com a motlle per continuar la replicació i reparar el DSB, formant una estructura coneguda com a *D-Loop*. És mitjançant aquest mecanisme pel qual s'ha definit que es poden crear llargues regions amb pèrdua d'heterozigositat (Llorente et al. 2008) (Figura 12).

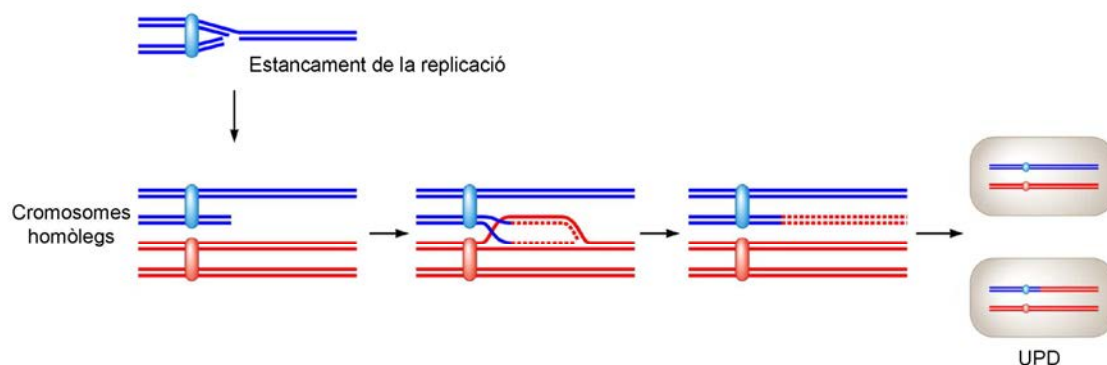


Figura 12. Disomia uniparental causada per *Break Induced Replication*. Adaptada de Rosenberg. Mbio 2011. L'aturada de la replicació està causada per un DSB. En conseqüència, la cromàtide amb el DSB utilitza una cromàtide corresponent al cromosoma homòleg per continuar la replicació.

Diversos estudis suggereixen que els mecanismes de formació de UPDs segmentals descrits en aquest apartat esdevenen en regions molt concretes del genoma, afectant principalment regions repetitives o llocs fràgils (Stephens et al. 2006). Per exemple, s'ha descrit que en pacients amb síndrome

mielodisplàsica el 43% de les UPDs estan localitzades en llocs fràgils (Mohamedali et al. 2007). Els llocs fràgils són regions del genoma on s'observa una incidència de trencament del DNA més elevada. A causa d'aquesta naturalesa, els llocs fràgils s'han descrit com a punts preferencials de trencament per a la creació de diverses alteracions cromosòmiques (Gümüş et al. 2002). Un altre grup identifica que les regions recurrents de UPD en limfomes de cèl·lules del mantell es troben significativament associades amb regions cromosòmiques enriquides per CNVs i duplicacions segmentàries (Bea et al. 2009). En pacients amb FAP s'han identificat punts recurrents de recombinació meiótica en l'inici del segment que dona lloc a la UPD al braç cromosòmic 5q. Els autors suggereixen que la recombinació homòloga en aquests punts del cromosoma 5 és un mecanisme específic per inactivar el TSG *APC* (Howarth et al. 2009). Recentment, un estudi en AML ha reportat un increment en la formació de UPDs en casos d'AML amb mutacions de *FLT3* i *JAK2V617F* (Gaymes et al. 2017). El seu model proposa que els casos amb aquestes mutacions adquireixen un increment d'espècies reactives de l'oxigen que provoca un augment de DSBs al DNA (Gaymes et al. 2017). Aquests DSBs es reparen mitjançant recombinació homòloga arrel d'una sobre-regulació de RAD51, una de les proteïnes més importants d'aquest mecanisme de reparació.

1.4 Tècniques per a l'estudi d'alteracions cromosòmiques

1.4.1 Arrays de CGH (aCGH)

Els aCGH usats en l'actualitat incorporen sondes de seqüències d'oligonucleòtids, tot i que conceptualment es basen en la tècnica de CGH convencional que va sorgir a principis dels anys 90 i que va revolucionar la manera d'estudiar el càncer (Kallioniemi et al. 1992). El principi bàsic de la tècnica és la comparació de dos genomes, un problema i l'altre control o de referència, barrejats a proporcions iguals (1:1) i marcats amb fluorocroms diferents. (Figura 13).

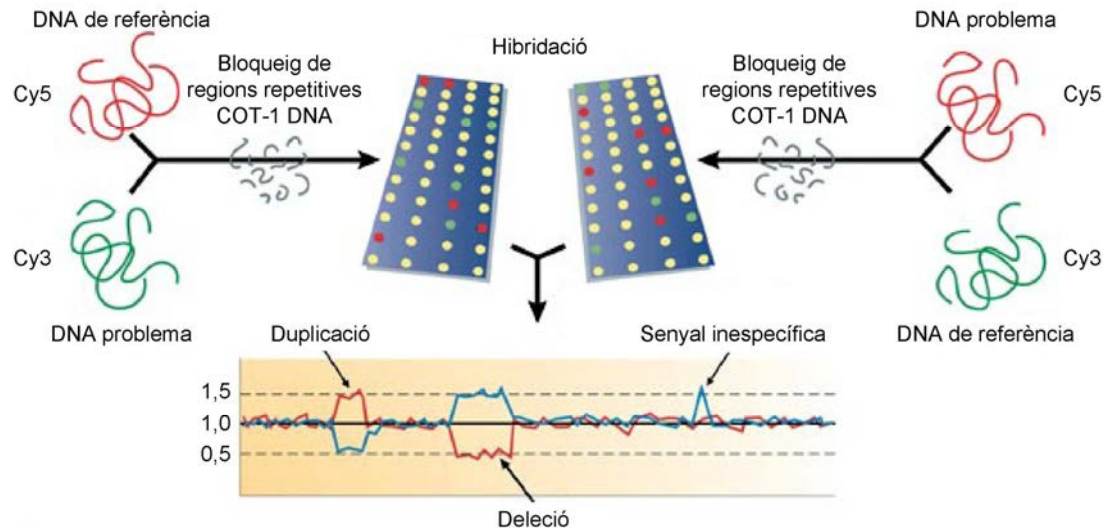


Figura 13. Metodologia de l'array de CGH. El DNA problema i el de referència es marquen amb fluorocroms diferents. Després es desnaturalitzen i es bloquegen les regions repetitives mitjançant l'addició de DNA COT-1. Posteriorment, es procedeix a la hibridació competitiva d'ambdós DNAs amb les sondes disposades en l'array. La detecció dels senyals fluorescents produïts després de la hibridació es realitza mitjançant l'ús d'un escàner. La fluorescència detectada per l'escàner es normalitza i es converteix en una escala logarítmica (\log_2ratio o LRR) la qual s'usa per inferir les CNAs (Coe et al. 2007).

1.4.2 SNP arrays

Els SNPs (de l'anglès, *Single Nucleotide Polymorphisms*) són el tipus de variació de seqüència més comuns en el genoma humà. Són variacions d'un únic nucleòtid, repartits en gran part del genoma i presents en més d'un 1% de la població. Entre moltes altres aplicacions, els SNPs han esdevingut clau en estudis d'associació al llarg del genoma (GWAS, de l'anglès *genome-wide association studies*) per determinar associacions entre variants concretes i malaltia (Cheung et al. 2005). A més, l'ús freqüent en l'última dècada d'arrays consistents amb sondes d'SNPs (SNP arrays), ha permès el poder d'avaluar no només alteracions en nombre de còpies, sinó també alteracions que afecten a la ràtio entre els diferents, concepte que rep el nom de freqüència de l'al·lel B (BAF, de l'anglès *B-Allele-Frequency*) (Alkan et al. 2011; González et al. 2011). La informació més rellevant que aporta el BAF és la detecció de regions de UPD, regions amb LOH sense canvi en nombre de còpies associades (Figura 14).

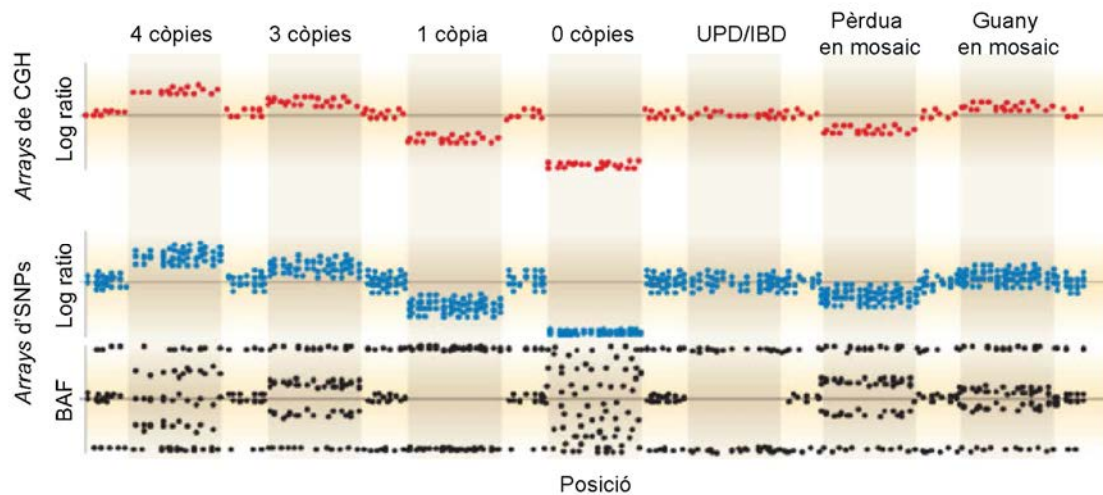


Figura 14. Detecció d'anomalies per aCGH i SNP arrays. Adaptada de Alkan et al. Nature Reviews Genetics 2011. Mentre que l'aCGH utilitza el log ratio per determinar regions de cromosòmiques de guanys i pèrdues, els *arrays* d'SNPs utilitzen paral·lelament el BAF. El BAF permet la quantificació acurada de nombre de còpies (de 4 a 0 còpies en les primeres quatre columnes de la figura). A més, el BAF permet la detecció de regions amb LOH sense canvis en el nombre de còpies (UPD o IBD). Finalment, el BAF pot detectar guanys i pèrdues en mosaic.

La gran avantatge que suposa l'avaluació de CNAs mitjançant els arrays d'SNPs és que no són necessàries mostres totalment pures, ja que gràcies a la informació que aporta el BAF, es pot inferir el mosaïcisme del mateix tumor (Peiffer et al. 2006; Conlin et al. 2010; Rodríguez-Santiago et al. 2010; Jasmine et al. 2012). Paral·lelament, el BAF permet la quantificació en nombre de còpies de cadascun dels al·lells, assignant una ploïdia global del genoma tumoral (Van Loo et al. 2010; Carter et al. 2012).

1.4.3 Hibridació *in situ* fluorescent (FISH)

L'aparició de la hibridació *in situ* fluorescent (FISH, de l'anglès *Fluorescence in situ Hybridization*) a la dècada dels anys 80, va revolucionar la metodologia per avaluar alteracions cromosòmiques (Bauman et al. 1980). La FISH es basa en el mateix principi que qualsevol mètode d'hibridació del DNA, on s'usa una sonda de DNA, la seqüència de la qual és complementària a la regió del genoma que es vol estudiar. La sonda acostuma a estar marcada fluorescentment, fet que permetrà ser avaluada al microscopi de fluorescència

un cop hagi hibridat amb la seqüència complementària de la mostra que es vulgui analitzar (cromosomes metafàsics o nuclis interfàsics). L'avaluació de cèl·lules úniques, com a base metodològica de la tècnica, manté la FISH com a tècnica essencial per a l'estudi del càncer (Heselmeyer-Haddad et al. 2012). Entre les múltiples aplicacions de la tècnica original, cal destacar l'ús de la FISH en protocols que preserven l'estructura tridimensional del nucli, FISH en 3D (3D-FISH, de l'anglès *Three-Dimensional Fluorescence in situ Hybridization*).

1.5 Arquitectura nuclear

La 3D-FISH combinada amb la microscòpia confocal per tal d'obtenir imatges reconstruïdes del nucli ha esdevingut una metodologia clau per a l'estudi de l'organització del genoma al nucli interfàsic. L'ús de pintats cromosòmics en cèl·lules en cultiu ha permès la visualització dels cromosomes en el seu format a interfase, els anomenats territoris cromosòmics (CTs, de l'anglès *Chromosome Territories*). La hipòtesi que els cromosomes ocupen un territori concret al nucli interfàsic ja va ser proposada a finals del segle XIX pel científic austríac Carl Rabl en els seus estudis en cèl·lules d'amfibis (Rabl 1885). No obstant, va ser el biòleg alemany Theodor Boveri el que va introduir per primera vegada el terme territori cromosòmic (Boveri 1909). Aquest concepte va romandre en l'oblit durant diverses dècades del segle XX (Cremer et al. 2014). No va ser fins a finals de la dècada dels 70 i gràcies a experiments amb tinció Giemsa o inducció de dany al DNA per radiacions ultraviolades que el terme va tornar a aparèixer i va acabar sent acceptat per la comunitat científica (Stack et al. 1977; Zorn et al. 1979; Cremer and Cremer 2010). L'ús de sondes de pintat cromosòmic va permetre visualitzar la localització dels cromosomes en interfase (Cremer et al. 1984; Gray et al. 1987). Més tard, amb el desenvolupament de protocols que preserven l'estructura tridimensional del nucli va ser possible avaluar l'organització dels CTs dins el nucli interfàsic (Cremer et al. 2008). Així, s'ha descrit que els CTs mantenen una posició no aleatòria dins el nucli interfàsic, amb la densitat gènica com a factor organitzador principal (Cremer and Cremer 2001). Mentre que els CTs més rics en gens, com el CT19 (CT corresponent al cromosoma 19), es posicionen a prop del centre del nucli, els CTs més pobres en gens, com el CT18 es

posicionen més a prop de la perifèria (Croft et al. 1999; Boyle et al. 2001; Cremer and Cremer 2001). Aquesta organització radial dels CTs dins el nucli interfàsic s'ha observat ben conservada en primats (Tanabe et al. 2002). Tanmateix, s'ha descrit que en fibroblasts és la mida dels cromosomes i no la densitat gènica dels mateixos el principal factor que marca el posicionament radial dins el nucli (Sun et al. 2000; Bolzer et al. 2005). Altres paràmetres com, l'activitat transcripcional, el moment de replicació i el contingut de GC s'han descrit també com a factors determinants per al posicionament dels CTs i de regions subcromosòmiques dins el nucli interfàsic (Mayer et al. 2005; Federico et al. 2006; Goetze et al. 2007; Grasser et al. 2008; Heppenger et al. 2008). La multitud d'estudis que avaluen aspectes fonamentals de l'organització de la cromatina dins el nucli interfàsic ha promogut el desenvolupament de diversos models d'arquitectura nuclear a nivell cromosòmic i subcromosòmic (Figura 15) (Cremer and Cremer 2010).

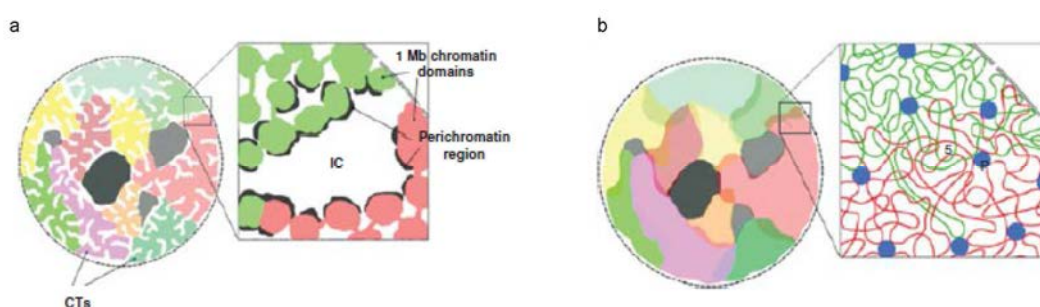


Figura 15. Models d'arquitectura nuclear. Extreta de Cremer and Cremer. Cold Spring Harbour Perspectives in Biology 2010. (a) *Chromosome Territory Interchromatin Compartment* (CT-IC). En aquest model es descriuen uns dominis d'1Mb interconnectats pel compartiment intercromatínic. En la perifèria dels dominis es localitza la regió pericromatínica corresponent a cromatina activa i descondensada. (b) *Interchromatin Network* (ICN). En aquest model es descriuen regions cromatíniques (*loops*) del mateix CT, i d'altres CTs, que mantenen contacte en *cis* i *trans*. Els punts blau representen punts de contacte intra- i inter-cromosòmica.

Un dels models més reconegut és l'anomenat territori cromosòmic i compartiment intercromatínic (CT-IC, de l'anglès *Chromosome Territory Interchromatin Compartment*) (Cremer et al. 2015). En aquest model es defineixen els CTs com a estructures independents similars a esponges, format per dominis d'aproximadament 1 Mb. Connectant aquests dominis cromatínics s'observen regions menys denses, el que definiria el compartiment intercromatínic (Albiez et al. 2006; Rouquette et al. 2009). Els autors del model

proposen que els gens més actius podrien ser fàcilment transcrits al trobar-se a la perifèria dels dominis cromatínics i, per tant, en contacte directe amb l'espai intercromatínic on es localitzarien els sistemes de transcripció. Contràriament al CT-IC, el model anomenat xarxa intercromatínica (ICN, de l'anglès *Interchromatin network*) s'oposa a l'existència d'un compartiment intercromatínic dins els CTs (Branco and Pombo 2006). Els autors proposen que la cromatina que forma els CTs es pot estendre fins CTs veïns en forma de llargues estructures cromatíniques definides com a *loops*. Múltiples autors reforcen la presència d'aquestes extensions cromatíniques, les quals afavoririen una potencial co-regulació de gens localitzats a grans distàncies entre ells, fins i tot en cromosomes diferents. (Chubb and Bickmore 2003; Kosak and Groudine 2004a; Fraser and Bickmore 2007a). Aquests *loops* s'han descrit per localitzar-se en regions del nucli interfàsic amb presència de maquinària de transcripció (Fraser and Bickmore 2007a).

Una de les grans controvèrsies més discutides a la literatura ha estat la relació entre l'activitat transcripcional i la posició de gens a nivell nuclear. A nivell global, a principis dels anys 90 ja es va proposar que els gens més expressats acostumaven a localitzar-se a la perifèria dels CTs (Zirbel et al. 1993). Posteriorment, el mateix grup va estudiar la posició de dos gens respecte el CTX (Dietzel et al. 1999). Per una banda, van analitzar el gen *ANT2*, el qual està involucrat en la inactivació d'un dels dos cromosomes X. D'altra banda, van estudiar la posició del gen *ANT3*, localitzat en la regió pseudoautosòmica del cromosoma X i que, per tant, no pateix la inactivació del mateix cromosoma. Els autors van observar que el gen *ANT3* i l'al·lel actiu del gen *ANT2* es localitzaven a la perifèria del CTX, mentre que l'al·lel inactiu de *ANT2* es localitzava en regions més internes del CTX. Contràriament, dos estudis van observar que gran part dels gens se situen a la perifèria del CT sense mantenir cap mena de relació amb els nivells d'expressió dels mateixos (Kurz et al. 1996; Clemson et al. 2006). No obstant, diversos treballs han identificat regions cromatíniques, riques en gens i altament expressats, que s'estenen a l'exterior del CT corresponent. Aquestes extensions cromatíniques s'han definit com a *loops*. (Dietzel et al. 1999; Volpi et al. 2000; Williams et al. 2002; Mahy et al. 2002; Chambeyron and Bickmore 2004). Volpi i col·laboradors van identificar

nombrosos *loops*, corresponents a regions del cromosoma 6, estenent-se a l'exterior del CT6 (Volpi et al. 2000). La freqüència per la qual aquest esdeveniment es va observar tenia relació amb el nombre de gens del complex major d'histocompatibilitat actius en el tipus cel·lular avaluat. La sobre-activació dels gens d'aquest complex, mitjançant l'addició d'interferó gamma, va augmentar la freqüència d'aquestes projeccions cromatíniques. En un altre estudi, es descriu que el complex de diferenciació epidèrmic (*EDC*) es localitza a l'exterior del CT1 en queratinòcits, on aquest complex és altament actiu (Williams et al. 2002). Contràriament, en limfoblasts *EDC* és inactiu, i s'observa en posicions més internes del mateix CT. En un altre estudi on s'han usat cèl·lules de ratolí s'han identificat extensions cromatíniques recurrents de la regió on s'inclou la família de gens *Hox*, respecte el CT11 (Chambeyron and Bickmore 2004). Aquests *loops* es van observar després de la inducció de l'expressió del gen *HoxB*. Finalment, Mahy i col·laboradors van detectar casos freqüents de *loops* respecte el CT11, de la regió cromosòmica rica en gens 11p15.5 (Mahy et al. 2002). A més, altres regions riques en gens també es van trobar projectades a l'exterior del CT en aquest mateix estudi. Els autors suggerien que la densitat gènica local, més que l'activitat de gens individuals, era el factor més determinant d'organització nuclear intra-cromosòmica. (Mahy et al. 2002). Posteriorment, un altre grup va detectar també *loops* recurrents en la regió cromosòmica rica en gens 11p15.5 en un estudi realitzat en limfòcits i fibroblasts (Küpper et al. 2007). Tanmateix, a part d'aquesta regió, els autors no van observar *loops* de manera freqüent, suggerint que aquestes estructures són poc comunes. Més recentment, l'ús freqüent de les noves tecnologies derivades de la captura de la conformació de la cromatina (3C, 4C, 5C i Hi-C) (3C, de l'anglès *Chromatin Conformation Capture*) han confirmat que dominis cromatínics actius tendeixen a localitzar-se a la perifèria dels CTs (Nagano et al. 2013).

2 Premissa i objectius

La disomia uniparental s'ha identificat recurrentment en diversos tipus tumorals. Específicament, aquest esdeveniment somàtic s'observa en regions freqüentment involucrades en pèrdues cromosòmiques i conseqüentment associades a la presència de gens supressors de tumors. Aquest fet, suggereix que la UPD pot ser un mecanisme rellevant per a la carcinogènesi de determinats càncers. Una singularitat del càncer colorectal és la presència d'una elevada inestabilitat genòmica, caracteritzada per un increment d'errors cromosòmics numèrics i estructurals. D'aquesta manera el CRC esdevé un bon model per a l'estudi de la UPD en el càncer.

D'altra banda, l'aplicació de tècniques de FISH juntament amb protocols que preserven l'estructura tridimensional dels nuclis ha portat a l'estudi de l'organització del genoma dins el nucli interfàsic. Conseqüentment, s'ha descrit que els cromosomes ocupen uns territoris específics anomenats territoris cromosòmics, els quals es posicionen dins el nucli interfàsic en funció de la seva densitat gènica. No obstant, la relació entre l'activitat transcripcional de gens individuals i la localització d'aquests respecte el CT en el qual es localitzen està més discutida.

És arrel d'aquest escenari que en aquesta tesi es pretenen estudiar els següents objectius específics.

Objectius:

1. Establir un perfil genòmic de UPDs i CNAs recurrents en CRC i altres tipus tumorals per tal de determinar l'especificitat d'aquests patrons en el càncer.
2. Estudiar la presència de variants puntuals inactivants en gens supressors de tumors localitzats en regions involucrades en UPD per tal de determinar la funcionalitat biològica d'aquest esdeveniment.
3. Determinar el posicionament tridimensional de gens diferencialment expressats respecte el territori cromosòmic en el qual es localitzen.

3 Resultats

3.1 Abstract

Colorectal cancer (CRC) is characterized by specific patterns of copy number alterations (CNAs), which helped with the identification of driver oncogenes and tumor suppressor genes (TSGs). More recently, the usage of single nucleotide polymorphism arrays (SNP arrays) provided information of copy number neutral loss of heterozygosity (cnLOH), thus suggesting the occurrence of somatic uniparental disomy and uniparental polysomy events (hereafter referred as UPDs). The aim of the study corresponding to the first chapter of this thesis is to establish an integrative profiling of recurrent UPDs and CNAs in sporadic CRC. Our results indicate that regions showing high frequencies of UPD mostly coincide with regions typically involved in genomic losses. Among them, chromosome arms 3p, 5q, 9q, 10q, 14q, 17p, 17q, 20p, 21q and 22q preferentially showed UPDs over genomic losses suggesting that tumor cells must maintain the disomic state of certain genes to favor cellular fitness. A meta-analysis using over 300 samples extracted from The Cancer Genome Atlas data portal (TCGA) confirmed our findings. Finally, sequencing and fluorescence *in situ* hybridization analysis of the gene *APC* underlined that a somatic UPD event might represent the second hit to achieve biallelic inactivation of this TSG in colorectal tumors.

In the second chapter, we aimed at establishing a map of UPDs in a tumor-type specific manner and integrating these data with mutational status of TSGs. To do so, we analyzed SNP array and whole-exome sequencing data extracted from the TCGA of different types of gastrointestinal cancers, including colon (COAD), rectum (READ), gastric (STAD) and esophageal (ESAD) adenocarcinomas and esophageal squamous cell carcinoma (ESCC). Our results showed that UPDs profiles of COAD cohort matched with those identified in READ. On the other hand, UPD profiles of STAD, ESAD and ESCC showed significant differences from the ones identified in COAD and READ, confirming the existence of cancer-specific UPD profiles. Moreover, our analysis indicated that such an event acts as a “second hit”, in the well-known Knudson’s model, to biallelically inactivate TSGs in all cohorts. Specifically, *APC*, *ARID1A*

and *NOTCH1* were recurrently inactivated by UPDs in CRC, STAD and ESCC, respectively. Besides, *TP53* showed inactivations caused by UPD at chromosome arm 17p in all tumor-types, although copy number losses as a “second hit” were much frequent in this case. In addition, by inferring genome ploidy in all samples we demonstrated that a higher number of UPD events, both numerical and structural, are detected in highly aneuploid genomes compared to near-diploid tumors.

Finally, in the third chapter we aimed to assess the nuclear distribution of chromosome territories and gene positioning in CRC cell lines. It is well known that human chromosomes occupy distinct territories in the interphase nucleus. Such chromosome territories (CTs) are positioned according to gene density. Gene-rich CTs are generally located in the center of the nucleus, while gene-poor CTs are positioned more towards the nuclear periphery. However, the association between gene expression levels and the radial positioning of genes within the CT is still under debate. Thus, we performed three-dimensional fluorescence *in situ* hybridization experiments in the colorectal cancer cell lines DLD-1 and LoVo using whole chromosome painting probes for chromosomes 8 and 11 and BAC clones targeting four genes with different expression levels assessed by gene expression arrays and RT-PCR. Our results confirmed that the two over-expressed genes, *MYC* on chromosome 8 and *CCND1* on chromosome 11, are located significantly further away from the center of the CT compared to under-expressed genes on the same chromosomes, i.e., *DLC1* and *SCN3B*. Furthermore, when *CCND1* expression was reduced after silencing the major transcription factor of the WNT/ β -catenin signaling pathway, *TCF7L2*, the gene was repositioned and mostly detected in the interior of the CT.

In summary, our data define specific profiles of somatically acquired UPDs in sporadic CRC and other tumors from the gastrointestinal tract. Moreover, the integration with exome data confirms that TSGs are recurrently mutated in UPD regions. Thus, highlighting the importance of these events as a mechanism to achieve biallelically inactivation of TSGs. Finally, our last study provides insights into how genes are positioned with respect to the CT in which they reside and according to their expression levels.

3.2 Capítol I

Keyvan Torabi, Rosa Miró, Nora Fernández-Jiménez, Isabel Quintanilla, Laia Ramos, Esther Prat, Javier del Rey, Núria Pujol, Jonathan Keith Killian, Paul S. Meltzer, Pedro Luis Fernández, Thomas Ried, Juan José Lozano, Jordi Camps and Immaculada Ponsa.

Patterns of somatic uniparental disomy identify novel tumor suppressor genes in colorectal cancer.

Carcinogenesis (2015) 36:1103-10.

Impact Factor (2015): 4.874

ORIGINAL MANUSCRIPT

Patterns of somatic uniparental disomy identify novel tumor suppressor genes in colorectal cancer

Keyvan Torabi¹, Rosa Miró^{1,2}, Nora Fernández-Jiménez^{1,2,7}, Isabel Quintanilla³, Laia Ramos^{1,2,8}, Esther Prat^{1,2,9}, Javier del Rey^{1,2}, Núria Pujol¹, J. Keith Killian⁴, Paul S. Meltzer⁴, Pedro Luis Fernández⁵, Thomas Ried⁴, Juan José Lozano⁶, Jordi Camps^{1,3,4,†} and Immaculada Ponsa^{1,2,*,†}

¹Unitat de Biologia Cel·lular i Genètica Mèdica, Departament de Biologia Cel·lular, Fisiologia i Immunologia, Facultat de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Catalonia 08193, Spain, ²Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Catalonia 08193, Spain, ³Gastrointestinal and Pancreatic Oncology Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Barcelona, Catalonia 08036, Spain, ⁴Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA, ⁵Department of Pathology, Hospital Clínic/IDIBAPS, Universitat de Barcelona, Barcelona, Catalonia 08036, Spain and ⁶Bioinformatics Unit, CIBERehd, Barcelona, Catalonia 08036, Spain

⁷Present address: Epigenetics Group, International Agency for Research on Cancer 69008, Lyon, France

⁸Present address: Unitat de Genòmica i Bioinformàtica, Institut de Medicina Predictiva i Personalitzada del Càncer (IMPPC), Badalona, Catalonia 08916, Spain

⁹Present address: Laboratori de Genètica Molecular, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Hospitalet de Llobregat, Catalonia 08908, Spain

*To whom correspondence should be addressed. Tel: +34 935811724; Fax: +34 935811025; Email: imma.ponsa@uab.cat

†These authors contributed equally to this work.

Abstract

Colorectal cancer (CRC) is characterized by specific patterns of copy number alterations (CNAs), which helped with the identification of driver oncogenes and tumor suppressor genes (TSGs). More recently, the usage of single nucleotide polymorphism arrays provided information of copy number neutral loss of heterozygosity, thus suggesting the occurrence of somatic uniparental disomy (UPD) and uniparental polysomy (UPP) events. The aim of this study is to establish an integrative profiling of recurrent UPDs/UPPs and CNAs in sporadic CRC. Our results indicate that regions showing high frequencies of UPD/UPP mostly coincide with regions typically involved in genomic losses. Among them, chromosome arms 3p, 5q, 9q, 10q, 14q, 17p, 17q, 20p, 21q and 22q preferentially showed UPDs/UPPs over genomic losses suggesting that tumor cells must maintain the disomic state of certain genes to favor cellular fitness. A meta-analysis using over 300 samples from The Cancer Genome Atlas confirmed our findings. Several regions affected by recurrent UPDs/UPPs contain well-known TSGs, as well as novel candidates such as *ARID1A*, *DLC1*, *TCF7L2* and *DMBT1*. In addition, *VCAN*, *FLT4*, *SFRP1* and *GAS7* were also frequently involved in regions of UPD/UPP and displayed high levels of methylation. Finally, sequencing and fluorescence *in situ* hybridization analysis of the gene *APC* underlined that a somatic UPD event might represent the second hit to achieve biallelic inactivation of this TSG in colorectal tumors. In summary, our data define a profile of somatic UPDs/UPPs in sporadic CRC and highlights the importance of these events as a mechanism to achieve the inactivation of TSGs.

Received: March 31, 2015; Revised: July 6, 2015; Accepted: July 29, 2015

© The Author 2015. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com.

Abbreviations

CNA	copy number alteration
cnLOH	copy number neutral loss of heterozygosity
CRC	colorectal cancer
FISH	fluorescence in situ hybridization
PSCBS	parent-specific circular binary segmentation algorithm
LOH	loss of heterozygosity
SNP	single nucleotide polymorphism
TSG	tumor suppressor gene
UPD	uniparental disomy
UPP	uniparental polysomy

Introduction

Copy number alterations (CNAs) are the defining feature of tumors of epithelial origin, including colorectal, providing a consistent landscape of genome-wide gains and losses in a tumor-type specific manner (1,2). These genomic imbalances might contain oncogenes in areas of amplification and tumor suppressor genes (TSGs) in regions commonly subjected to deletion. More recently, the usage of single nucleotide polymorphism (SNP) arrays allowed the identification of allele specific imbalances thus defining regions of copy number neutral loss of heterozygosity (cnLOH) or uniparental disomy (UPD) (3–5). UPD was firstly described by Engel (6), and arises when an individual inherits two copies of maternal or paternal chromosomes as a result of a meiotic error. While this gives rise to constitutional UPD associated with developmental disorders, this phenomenon has also been described in somatic cells, the so-called somatically acquired UPD (3–5). In fact, UPD has been recently described in several malignancies, including both solid tumors and hematological neoplasias (7–9), and it has been suggested as a mechanism to potentially alter the expression of driver genes involved in carcinogenesis (4). In addition, in many cancer cells of epithelial origin, whole genome duplications occur frequently, thus potentially resulting in trisomies and tetrasomies of chromosomes inherited from the same progenitor, a scenario designated as uniparental polysomy (UPP).

Colorectal cancer (CRC) is one of the most common cancers in Western Europe and North America (10). Continuous efforts are made to comprehensively characterize the genome of CRC cells in order to understand the genetic basis of this devastating disease and to identify biomarkers that can help with early detection and improve prognostication. In particular, genomic profiling of sporadic CRC has revealed consistent gains and losses during the emergence and evolution of these tumors (11). Based on the two-hit Knudson hypothesis, regions with LOH might contain known TSGs, which can be inactivated by either a genomic loss or a second inactivating mutation (12). In this context, UPD arises as an alternative mechanism to reach functional impairment of TSGs in cancer cells, in which one allele holds an inactivating mutation, and due to a duplication of the chromosome containing the mutated allele and a loss of the chromosome with the wild-type allele, the tumor cell becomes disomic with an inactivating mutation in homozygosis. Foremost intriguing is the question whether different profiles of UPD/UPP are observed depending on the tumor's tissue of origin, and with what frequency UPD/UPP does occur in different cancer types. For example, deletion of chromosome 5q is very common on myelodysplastic syndrome and acute myeloid leukemia, but UPD at 5q is not extensively reported in these cancers types. On the other hand, UPD on chromosome 5q containing the TSG APC

has been reported to be a common event in CRC (13–15) and in *in vitro* models (16,17). Moreover, it has been recently suggested that cnLOH affecting APC may play a role in earlier stages of tumorigenesis as this event was already found in adenomas (18).

In the present study, we aimed at establishing a map of UPD/UPP in sporadic CRC, and integrating these data with somatic CNAs and the methylation status of cancer-associated genes. Our results have been cross-compared to The Cancer Genome Atlas (TCGA) data to confirm which specific regions of the CRC genome are prone to develop UPD/UPP. In addition, combining mutation analysis and fluorescence *in situ* hybridization (FISH) of the gene APC provided insights into the mechanism by which UPDs result in biallelic inactivation of TSGs.

Materials and methods

Sample collection

Thirty colorectal adenocarcinomas provided by the Hospital Clinic of Barcelona/IDIBAPS Biobank were included in this study (Table 1). All patients signed the corresponding informed consent and the sample collection was approved by the local Ethics Committees. Fresh tumor and adjacent normal mucosa samples were collected immediately after surgical resection and preserved in DMEM culture medium supplemented with fetal bovine serum (Life Technologies, Carlsbad, CA), antibiotics and antifungals before storage at -80°C . An experienced pathologist macroscopically dissected the fresh tumor samples in order to minimize the inclusion of normal mucosa and necrotic tissue.

DNA extraction

DNA from tumor and matched adjacent normal mucosa was extracted using phenol-chloroform as described previously (19). DNA concentration and purity was assessed using a NanoDrop ND-1000 (Thermo Scientific, Waltham, MA) spectrophotometer.

SNP array

Two hundred and fifty nanograms of DNA were processed for hybridization on the GeneChip Human Mapping 250K Sty arrays (Affymetrix, Santa Clara, CA). The experimental procedure was performed according to recommendations of the manufacturer. After hybridization, the chips were processed using the GeneChip Fluidics Station 450, high-resolution microarray GeneChip Scanner 3000 and GCOS Instrument Control Workstation version 1.2 (Affymetrix). SNP calls were determined by GDAS version 3.0 with 25% level of confidence. Only samples with call rates >90% were included.

Array comparative genomic hybridization

Oligonucleotide-based array comparative genomic hybridization was performed according to the protocol provided by the manufacturer (Agilent Technologies, Santa Clara, CA), with minor modifications. Briefly, 3 μg of DNA from each sample were labeled with Cy3 and combined with DNA from matched normal colonic mucosa labeled with Cy5. Oligonucleotide-based Human Genome CGH Microarray (Agilent) with 244K features was hybridized, washed accordingly and scanned with an Agilent G2565BA scanner. Data were quality controlled and extracted using Agilent Technologies' Feature Extraction (version 9.1). Visualization was performed using the software Nexus Copy Number version 7.5 (BioDiscovery, El Segundo, CA).

Data processing

Genome-wide LOH and CNA information was obtained using the Affymetrix array. The CRMA version 2.0 was used for normalizing allelic estimates of one tumor sample based on estimates from a single matched normal (20). Next, the Paired parent-specific circular binary segmentation algorithm, PSCBS version 0.4, was utilized (21). Different cutoffs, computed by the PSCBS algorithm, were used to consider allelic imbalances and CNAs for each pair of samples. When the distance between two adjacent segments was < 2.5Mb, a single region was considered. cnLOH was

Table 1. Clinical information and experimental setup

Patient ID	Sex	Age ^a	Stage	Microsatellite instability status	aCGH	SNP array	Methylation array	APC sequencing	APC FISH
S1	M	51	IIA	Negative	X	X	X	X	X
S2	F	59	IIIB	Negative	X	X	X	X	X
S5	M	62	I	n.a.	X	X	X	n.a.	X
S6	M	74	IIIB	n.a.	X	X	X	n.a.	X
S7	F	62	IV	n.a.	X	X	X	n.a.	X
S8	F	76	IIA	Negative	X	X	X	X	X
S9	F	59	I	n.a.	X	X	X	n.a.	X
S10	F	74	IIIB	Negative	X	X	X	n.a.	X
S11	M	59	IIIB	n.a.	X	X	X	n.a.	X
S12	M	85	IIB	n.a.	X	X	X	n.a.	X
S13	F	65	IIB	Negative	X	X	X	n.a.	X
S32	M	46	IIIB	Negative	X	X	X	X	X
S33	F	80	I	Negative	X	X	X	X	X
S34	F	80	IIA	Negative	X	X	X	n.a.	n.d.
S35	M	46	IIIB	Negative	X	X	X	n.a.	n.d.
S36	M	61	IIB	Positive	X	X	X	n.a.	n.d.
S37	M	78	IV	n.a.	X	X	X	X	n.d.
S38	M	84	IIIB	n.a.	X	X	X	n.a.	n.d.
S39	F	44	IIIB	Negative	X	X	X	n.a.	n.d.
S40	M	75	IIA	n.a.	X	X	X	n.a.	n.d.
S41	M	68	IV	n.a.	X	X	X	n.a.	n.d.
S42	F	42	IV	n.a.	X	X	X	n.a.	n.d.
S43	F	78	IIA	n.a.	X	X	X	X	n.d.
S44	M	68	IIIB	Negative	X	X	X	n.a.	n.d.
S45	F	48	I	Negative	X	X	X	X	n.d.
S46	M	58	IV	Negative	X	X	X	X	n.d.
S47	F	38	IIIB	Negative	X	X	n.d.	X	n.d.
S48	M	80	I	n.a.	X	X	X	n.a.	n.d.
S49	M	62	IIA	n.a.	X	X	X	n.a.	n.d.
S50	M	67	IIA	n.a.	X	X	X	X	n.d.

n.a., not available; n.d., not determined due to technical issues.

^aAge of surgery.

defined as a region without CNAs but with LOH, including UPD and UPP events. On the other hand, copy number gains with loss of heterozygosity were defined as gained regions with LOH, including solely UPP events. Both, cnLOH and copy number gains with loss of heterozygosity, were added up together when identifying regions of UPD or UPP. Finally, circos plots from different sets integrating both UPD/UPP and CNA regions were drawn using the Rcircos package (22). The Mann-Whitney U test was performed to assess the difference in the segment length and the Student's t-test for the comparison between stages. For the Agilent platform, the CBS algorithm was used to detect CNAs (23). For this purpose, R-functions available in the DNACopy package were used with default parameter values. The segment list was then converted into a matrix format by aligning samples based on chromosome segments condensed by genes (human genome release: hg19). In order to measure the concordance of CNAs between Agilent and Affymetrix platforms, the Pearson correlation was computed using the CNAs condensed log₂ ratio by gene in both platforms. Only CNAs computed from the Affymetrix platform were taken into consideration for further analysis. Finally, Gene Set Enrichment Analysis was performed to determine whether recurrent UPD/UPP contained a statistically significant number of TSGs. For this analysis, we considered the genes obtained from the TSGene database (<http://bioinfo.mc.vanderbilt.edu/TSGene/>) (24).

Data were deposited in the NCBI Gene Expression Omnibus (GSE64114).

The Cancer Genome Atlas data analysis

With the purpose to validate our results, level 1 SNP6 array data from the colon adenocarcinoma dataset and pair-matched normal DNA obtained from the TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>) were systematically curated. Only those samples showing less than 500 segments after segmentation were included in the analysis. Finally, PSCBS algorithm was carried out, and 303 samples were considered for analysis.

Methylation analysis

One microgram of purified DNA from each paired tumor-normal sample was bisulfite converted using the EZ DNA Methylation-Gold Kit (Zymo Research, Orange, CA). Subsequently, methylation status of 1505 CpG sites representing 807 cancer-related genes were evaluated by using the Illumina GoldenGate Methylation Cancer Panel I Array (Illumina, San Diego, CA) as described previously (25). Methylation probes representing a change of beta values >50% and a false discovery rate ≤ 0.001 using the Benjamini-Hochberg procedure were considered differentially expressed (26).

Mutational screening

A subset of tumors and their matched normal mucosa DNA (n = 10) underwent PCR amplification and were screened for sequence variations at the APC mutational cluster region (exon 16: one fragment from nucleotide 3081–3210, and another fragment from nucleotide 3529–4767, NCBI hg19) by Sanger sequencing. All samples were sequenced in either forward or reversed orientation to ensure the highest amount of sequence overlapping among PCR products (Supplementary Table 1, available at Carcinogenesis Online). Only PCR products with a single amplicon went forward for sequencing. Sequences were visualized and analyzed using the dedicated software 4Peaks (Nucleobytes, Aalsmeer, The Netherlands). Any identified mutation was verified against the Catalogue of Somatic Mutations in Cancer database (COSMIC, <http://cancer.sanger.ac.uk/cosmic>).

FISH

FISH analysis was performed on tissue microarrays containing two replicates of the tumor and two replicates of the normal adjacent mucosa per sample. Two BAC clones were used: one covering the APC gene (RP11-3B10), kindly provided by Dr B. Espinet (IMIM-Hospital del Mar, Barcelona, Spain),

and a second one located at 5q31.2 (RP11-461O14), kindly provided by Dr X. Estivill (Center for Genomic Regulation, Barcelona, Spain), that was used as a control. They were labeled with Spectrum Orange and Spectrum Green (Abbott Molecular, Des Plaines, IL), respectively. DNA extraction, nick translation and hybridization were performed as described previously (19). Signal quantification was carried out with a Nikon Eclipse 50i fluorescence microscope in 100 nuclei per sample, and image acquisition was done by using the Isis Fluorescence Imaging System (MetaSystems, Altussheim, Germany).

Results

Minimal regions of CNAs

Thirty primary colorectal adenocarcinomas were analyzed with the GeneChip Human Mapping 250K Sty SNP arrays and oligonucleotide-based Human Genome CGH microarrays. The mean concordance of CNAs between both platforms was 92%. CNAs were found in 28 out of 30 samples (93.3%), with a mean of 22 gains and 7 losses per sample. One of the two samples without CNAs showed microsatellite instability, while the other sample belonged to a patient with familial CRC Type X (27). Genomic profiling of frequent gains and losses based on SNP-arrays and PSCBS analysis is displayed in a circos plot (Figure 1A). Chromosomes 7, 8q, 13 and 20q were gained in more than 40% of the samples. Chromosomes 1p, 4, 5q, 8p, 14, 15, 17p, 18 and 20p were lost in more than 20% of the cases. The most frequently gained region was 20q11.21 (70%), and the most frequently lost region was 18q21.2-q22.3 (40%). Detailed frequencies of gains and losses in a gene-centered manner are presented in Supplementary Table 2, available at Carcinogenesis Online. Furthermore, a comparison between samples with stages I and II, defined as low stage tumors, and samples with stages III and IV, defined as high stage tumors, was performed. Only genomic regions 6p21.31, 6p24.1-p25.3 and 19q13.32-q13.43 were found significantly lost in high stage compared to low stage colorectal adenocarcinomas (false discovery rate < 0.25).

UPD/UPP profiling

Genome-wide UPD/UPP profiling was performed on 30 adenocarcinomas by applying the PSCBS algorithm to the SNP array dataset. Overall, 21 out of 30 samples (70%) showed UPD/UPP

segments, ranging from 1 to 17. A total of 163 genomic segments corresponded to cnLOH, including both events UPD and UPP. In addition, 22 segments corresponded to copy number gains with loss of heterozygosity, including only UPP.

In order to establish an overview of regions simultaneously involved in UPD/UPP and CNAs, data have been displayed in an integrative circos plot (Figure 1A). Interestingly, most of the regions showing high frequency of UPD/UPP matched with those regions of the genome showing high frequency of losses. However, the median length of the segments involved in UPD/UPP was higher than those affected by genomic losses (41.7 versus 34.8Mb) (Figure 2). While copy number losses are more frequently affecting interstitial chromosome segments (28.8 versus 14.8%; $P < 0.05$), UPD/UPP regions displayed a tendency for whole chromosome events (35.2 versus 23.7%). On the other hand, UPD/UPP and copy number losses shared a similar frequency of telomeric events (34.3 versus 33.1%) (Table 2).

Genomic regions showing a frequency of UPD/UPP greater than 10% were located at chromosomes 1, 3, 5, 8, 9, 10, 14, 17, 18, 19, 20, 21 and 22 (Table 3). Remarkably, there are several genomic regions whose frequency of UPD/UPP events is equal or higher than the frequency of copy number losses. This is the case for genomic regions located at chromosome arms 3p, 5q, 9q, 10q, 14q, 17p, 17q, 20p, 21q and 22q, indicating a prevalence of UPD/UPP over genomic loss. Among these regions, TSGs such as APC and TP53 map to 5q (20% of UPD/UPP and 20% of losses) and 17p (23.3% of UPD/UPP and 23.3% of losses), respectively. Here, we report a novel region of UPD/UPP in CRC at 10q11.23-q26.3 containing the TSG PTEN (13.3% of UPD/UPP and 6.7% of copy number losses) as well as other tumor suppressor candidates, such as TCF7L2 and DMBT1 (16.7% of UPD/UPP and 6.7% of copy number losses). On the other hand, genomic regions with a higher prevalence of copy number losses than UPDs/UPPs are located on chromosomes 1p, 8p and 18. The analysis of the minimal regions of interest led to the identification of several candidate TSGs reported in the COSMIC database, such as ARID1A at 1p35.3 (13.3% of UPD/UPP, 20% of copy number losses), DLC1 at 8p22 (16.7% of UPD/UPP, and 33.3% of copy number losses), and SMAD4 at 18q21.1 (20% of UPD/UPP and 36.7% of losses). Furthermore, the Gene Set Enrichment Analysis showed a

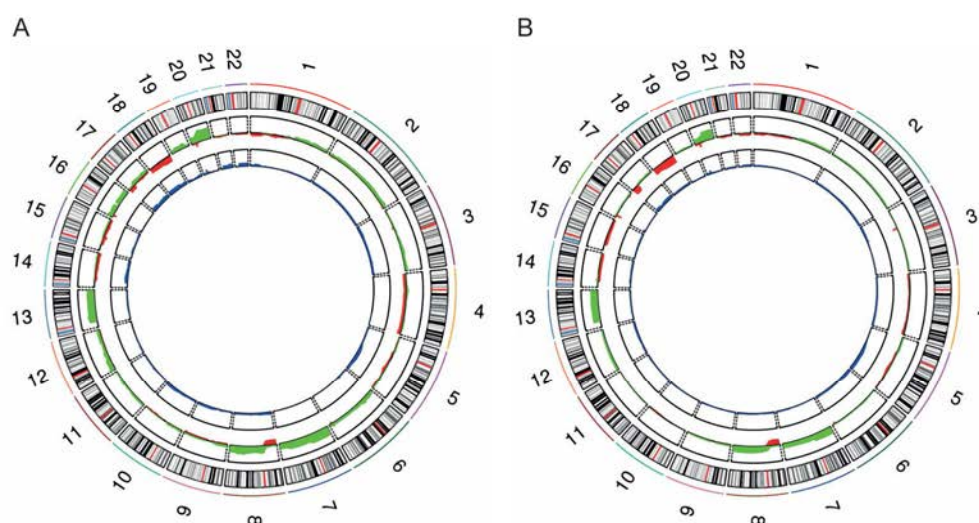


Figure 1. Profile of CNAs and UPDs/UPPs from our study (A) and from the TCGA cohort (B). Chromosomes are represented in the outer ring of the circos plot. In the middle, gained regions are in green and lost regions in red. The inner ring shows frequency of UPDs/UPPs in blue. To note, regions with high frequency of UPD/UPP matched with regions with high frequency of copy number loss.

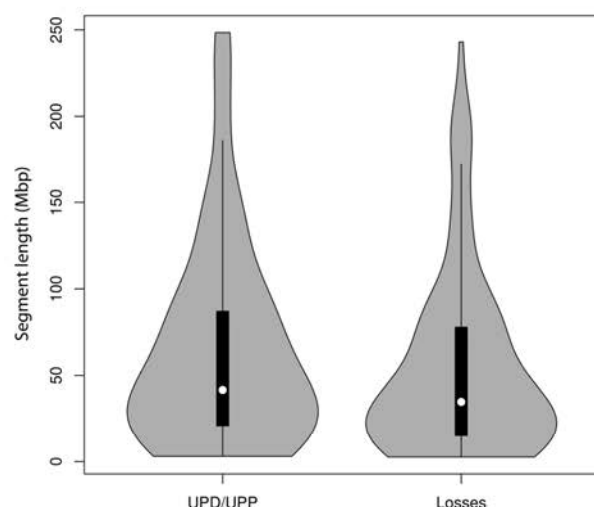


Figure 2. Distribution of the length of UPD/UPP events and genomic losses. The interquartile range is represented by the black box inside the violin plots. The median is represented by the white dot. Events smaller than 2.5Mb have been discarded.

Table 2. Distribution of UPDs/UPPs and genomic losses

Type of event	UPDs/UPPs	Genomic losses	P value ^e
Whole chromosome ^a	38 (35.2%)	28 (23.7%)	0.0785
Span centromere ^b	17 (15.7%)	17 (14.4%)	0.8531
Telomeric p ^c	18 (16.7%)	20 (16.9%)	1
Telomeric q ^c	19 (17.6%)	19 (16.1%)	0.8591
Interstitial ^d	16 (14.8%)	34 (28.8%)	0.0156

^aEvents including the entire chromosome.

^bEvents including the centromere but not the entire chromosome.

^cEvents including one of the two telomeres but not the centromere.

^dSpan comprised between two location within the same chromosome arm not involving the centromere nor the telomeres.

^eFisher's exact test was performed to assess statistical significance.

statistically significant presence of TSGs in the recurrent regions of UPD/UPP ($P < 0.05$) (Table 3).

In order to further narrow down novel TSGs potentially driving the selection of UPD/UPP events, we performed a cancer-gene oriented CpG methylation analysis. Our data show that the genes *VCAN*, *FLT4*, *SFRP1* and *GAS7* were located in regions of recurrent UPD/UPP (>10%) and showed high levels of hypermethylation in CpG islands ($\Delta\beta > 0.5$). A total of 23.3% of the samples showed UPD/UPP events affecting the hypermethylated gene *GAS7*. In addition, 46.7% of the remaining samples showed genomic losses in 17p involving this gene. Moreover, 13.3% of the samples showed that the highly hypermethylated genes *VCAN* and *FLT4* were included in regions with UPD/UPP. Finally, the gene *SFRP1*, which is located in the commonly gained region 8p11.21 in colorectal (and other) cancers, showed 13.3% of UPD/UPP and high levels of hypermethylation of the promoter or exon of the gene in all samples displaying a genomic gain.

Overall, a total of 14873 genes mapped to regions of UPD/UPP present in at least 10% of the samples, whereas only 9092 genes were included in genomic losses. A detailed table showing gene-centralized frequencies for each event (i.e. gains, losses, UPDs/UPPs and LOH) is presented in Supplementary Material (Supplementary Table 2, available at Carcinogenesis Online).

Meta-analysis using TCGA dataset

Next, we cross-compared the results obtained in our analysis using an independent sample set. Genome-wide CNAs and UPD/UPP events were analyzed from 303 normal-matched colon adenocarcinomas from the TCGA Level 1 dataset. An integrative circos plot with CNAs and UPD/UPP data is presented in Figure 1B. A very similar profile of gains and losses was observed when comparing CNAs from both cohorts. Despite the high concordance of the regions affected by UPD/UPP, we detected a decreased frequency of these events in the TCGA cohort. Regions showing a frequency of UPD/UPP higher than 10% in both datasets were located at 3p, 5q, 17, 18, 20, and 22. Of note, the frequency of UPD/UPP at 5q involving the TSG *APC* was very similar in both cohorts, showing a higher frequency of UPD/UPP (10.2–19.8% and 13.3–20% in the TCGA and our cohort, respectively) than copy number losses (6.3–17.2% in TCGA and 6.7–20% in our cohort). Regions with focal UPDs/UPPs were identified at 3p14.2, 16p13.3 and 20p12.1, which also matched with a high incidence of focal deletions. Interestingly, the gene *C20orf133* located at 20p12.1 is the most commonly affected gene by UPD/UPP in both cohorts (26.7% in our cohort and 22.1% in TCGA). A gene-centered table indicating all genomic events and their associated frequencies from the TCGA dataset is presented in Supplementary Table 3, available at Carcinogenesis Online.

Mutational status and FISH analysis of APC

We then explored whether UPDs/UPPs affecting the TSG *APC* act as a “second hit” by analyzing if a mutation, when present, could be homozygous. Sequence variations in the *APC* mutational cluster region were found in three out of six samples with UPD/UPP events at the chromosome arm 5q. All mutations identified were nonsense. In sample 1, the mutation c.3871C>T was found to be heterozygous in the tumor, showing a lower frequency of the inactivating allele. On the other hand, in samples 33 and 50 nonsense mutations c.4360A>T and c.3856G>T, respectively, were also found in heterozygosity in the tumor; however, the frequency of the mutated allele was much higher than the normal one, suggesting that the inactivating allele is present in the majority of the tumor cells (Figure 3A and B). In fact, we cannot discard that partial heterozygosity is due to the presence of non-tumoral infiltrating cells. In addition, fluorescent-labeled DNA probes covering *APC* and a control region at 5q31.2 were used to perform FISH analysis onto two tissue microarrays containing all samples included in our dataset. Interestingly, cases 33 and 1 displayed two signals of *APC* in 84% and 86% of all the cells, respectively, confirming that these samples displayed UPD/UPP in 5q (Figure 3C).

Discussion

Our study generated a profile of recurrent UPD/UPP regions in sporadic CRC to identify novel TSGs important for this disease. By utilizing the PSCBS algorithm in two different sample cohorts, we confirmed that chromosomal regions with the highest frequency of UPD/UPP mapped to chromosome arms 5q, 17p and 18q, containing the TSGs *APC*, *TP53*, and *SMAD4*, respectively (13,14,16,18). In addition, we have also identified regions of UPD/UPP affecting chromosomes 1p, 3p, 8p, 9q, 10q, 14q, 19, 20p, 21q and 22q, several of which have been previously reported as genomic losses in CRC (1). Although UPD/UPP profiles are very similar between our cohort and the TCGA sample set, some regions showed differences, most probably due to the sample size and heterogeneity. Of note is the high frequency of UPD/UPP events affecting chromosomes 6p

Table 3. Minimal regions of frequent UPD/UPP^a

Region	Size (bp)	Loss (%)	UPD/UPP (%)	LOH (%) ^b	Number of genes ^c	Genes of interest ^d
1p35.1-p36.33	15 209 290	16.7–20	13.3	30–33.3	243	ARID1A
1q44	372 585	3.3	13.3	16.7	9	
3p21.31-p26.3	46 624 171	3.3–6.7	13.3	16.7–20	314	VHL; MLH1
3q25.1	442 601	0	13.3	13.3	7	
5q14.1-q31.2	101 633 161	6.7–20	13.3–20	20–40	743	APC
8p12-p23.3	33 954 420	26.7–33.3	13.3–16.7	40–50	277	DLC1
8p11.21-p11.23	2 364 608	10–16.7	13.3	23.3–30	36	
8p11.21-p11.21	9 093 559	0–3.3	13.3	13.3–16.7	41	
8q21.13	333 470	0	13.3	13.3	1	
8q24.3	286 530	0	13.3	13.3	1	
9q21.13-q34.3	64 365 620	10–13.3	13.3	23.3–26.7	599	PTCH1; DAPK1
10q11.23-q26.3	83 250 609	3.3–10	13.3–16.7	16.7–26.7	624	PTEN; TCF7L2; DMBT1; PLCE1
14q11.2-q24.2	52 186 574	10–16.7	13.3–20	30	425	
14q31.3-q32.33	21 331 905	16.7–20	13.3	30–33.3	286	
17p13.3-q25.3	80 999 143	6.7–26.7	13.3–23.3	20–46.7	1434	TP53; NF1
18p11.32-q23	77 842 670	30–40	16.7–26.7	53.3–60	363	SMAD4
19p13.2-p13.3	9 862 965	0	13.3	13.3	307	
19q13.32-q13.43	12 657 480	0–3.3	13.3	13.3	569	PEG3
20p12.1	780 860	10–23.3	13.3–20	23.3–43.3	1	
20p12.1	96 087	10	13.3	23.3	1	
21p11.21-q22.3	37 157 795	0–10	16.7–23.3	20–30	307	
22q11.1-q13.33	34 352 671	10	20	30	558	MYO18B

^aOnly regions showing UPDs/UPPs in more than 10% of the cases are listed.

^bLOH represents frequencies of losses and UPDs/UPPs altogether.

^cOnly RefSeq genes have been taken into account.

^dGenes of recurrent somatic mutation (>8%) based on COSMIC database.

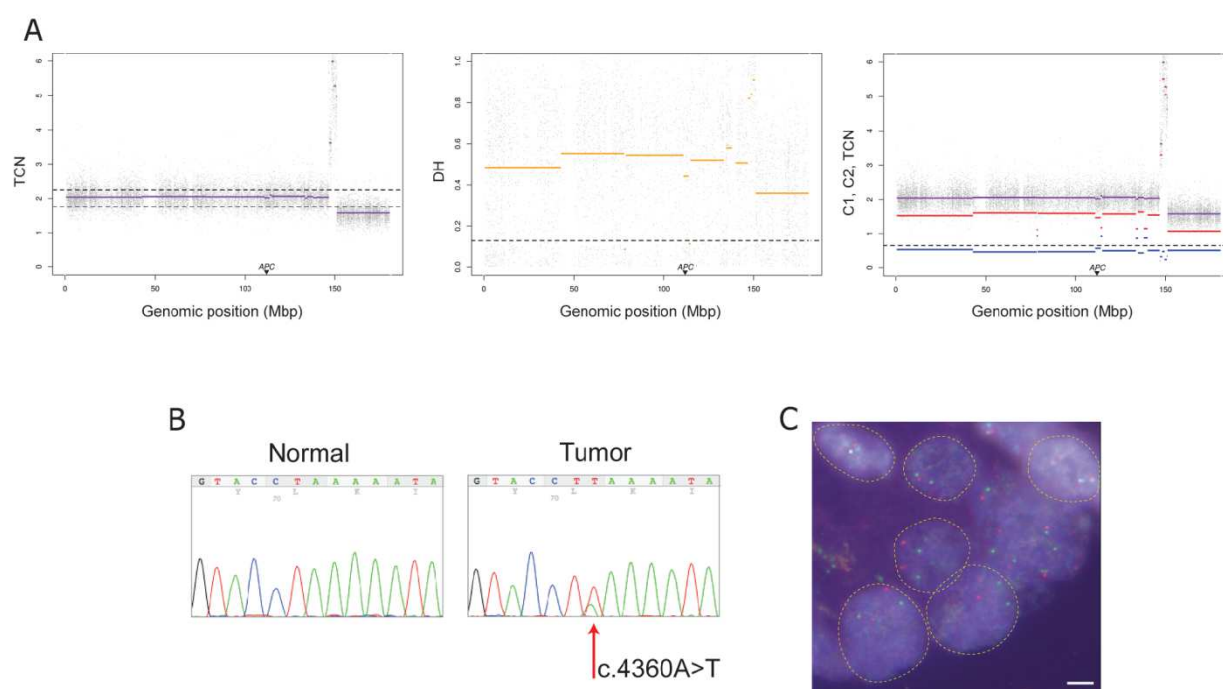


Figure 3. UPD of chromosome 5 accompanied by a homozygous nonsense mutation of sample 33. (A) Tumor-matched normal PSCBS profiling of SNP array data showing total copy number (panel 1), decrease of heterozygosity (panel 2) and allele-specific copy number (panel 3). Dotted horizontal lines represent specific thresholds for each parameter. As illustrated, APC is located in a region with cnLOH. (B) Sanger sequencing of the gene APC showing the existence of a nonsense mutation (c.4360A>T) in the tumor. The minority allele could represent intratumoral heterogeneity and/or infiltration of normal cells. (C) Microscopic assessment of FISH signals using fluorescent probes covering the gene APC (red) and a control probe at 5q31.2 (green). Single isolated nuclei displaying two copies of APC and the control probe confirmed that most tumor population showed a disomic chromosome 5q within this sample. Scale bar = 5 μ m.

and 12 in the TCGA cohort. Frequent allelic losses at these chromosomes have been described previously in mismatch repair deficient carcinomas (28). While only one sample in our cohort showed microsatellite instability, many more microsatellite

instability positive samples are included in the TCGA cohort, thus possibly explaining the frequency of such alterations.

The identification of LOH and high-level amplifications has led to the discovery of driver TSGs and oncogenes, respectively

(29,30). However, fewer studies have focused on regions with recurrent UPD/UPP. Our data show that these regions frequently coincide with genomic losses, suggesting that these events might also be guiding the inactivation of TSGs. In fact, several efforts have attempted to identify TSGs on chromosome arm 8p due to its recurrent loss in CRC (31). Integrative analysis of our data unveiled that the frequency of LOH in this genomic location is between 40 and 50%, including UPD/UPP events (13.3–16.7%) and copy number losses (26.7–33.3%). Among all genes encompassed within this area, *DLC1*, which encodes a RhoGAP protein that catalyzes the conversion of GTP-bound Rho to the inactive GDP-bound form, falls within the minimal region of UPD/UPP. *DLC1*, has been suggested to function as a TSG in several common cancers including hepatocellular carcinoma and CRC (32–34). Furthermore, mutations in this gene have been identified in nearly 10% of all CRC reported in the COSMIC database, thus representing a highly mutated gene and a good candidate for biallelic inactivation via UPD/UPP or genomic loss. Furthermore, here we also report novel regions of UPD/UPP that have not previously been associated with CRC. Among the 624 genes encompassing the 10q11.23–q26.3 region, *PTEN* stands out having already been considered a TSG in CRC (35). Likewise, *TCF7L2*, encoding a transcription factor that plays a key role in the Wnt/ β -catenin pathway, is also located within this region and has been found to be mutated in 9% of CRC (1,36). Moreover, a candidate TSG located at 1p35.3, *ARID1A*, a subunit of the SWI/SNF chromatin remodeler and transcription regulator of *MYC*, showed a high frequency of UPD/UPP (16.7%) and has been also found recurrently mutated (5%) in CRC (1,37).

The genome-wide integration of regions with UPD/UPP and methylation patterns has provided insights into the identification of novel putative TSGs. Of note, the genes *VCAN*, *FLT4*, *SFRP1* and *GAS7* were among the highest hypermethylated genes and were consistently involved in UPDs/UPPs. Foremost interesting is the gene *SFRP1*, which is a soluble modulator of the Wnt/ β -catenin signaling pathway and is located at 8p11.21, a genomic region with a high incidence of copy number gains in CRC (38). Our results revealed that in all cases with copy number gains of *SFRP1*, simultaneous hypermethylation of the promoter or the exon close to the promoter occurred. *SFRP1* has been already identified as being hypermethylated in CRC (39); however, our data suggest the mechanism by which biallelic inactivation of this candidate TSG might have arisen in the presence of copy number gains. Importantly, gene expression data extracted from the Oncomine Portal (<https://www.oncomine.org>) and the Cancer Genomics Browser (<https://genome-cancer.ucsc.edu>) supported the hypothesis that the aforementioned genes might be considered TSGs as three out of four candidates (*SFRP1*, *GAS7* and *FLT4*) showed downregulation in colorectal tumors compared to normal mucosa (data not shown).

Besides whole chromosome or chromosomal arm UPD/UPP events, we have also detected focal regions of UPD/UPP at 3p14.2, 16p13.3 and 20p12.1, containing the genes *FHIT*, *RBFOX1* and *C20orf133*, respectively. Interestingly, these regions have been also described as recurrent focal deletions in several cancer types, and the genes located at these genomic regions are among the largest genes in the genome (2). Whether deletions of these genes have an impact on tumorigenesis or whether they occur as a consequence of the genome plasticity in these regions (e.g. the presence of fragile sites or the fact that they encode structural proteins which carry tandem repeats prone to recombination) remains unknown (40,41). Further analysis of genes with high incidence of UPD/UPP is required in order to understand their functional role as TSGs.

Whole-chromosome UPD/UPP can occur in cancer cells as a consequence of segregation errors in mitosis through anaphase

lagging and/or non-disjunction events (4). On the other hand, homologous recombination, with the purpose of repairing a double-strand break, has been proposed as a possible mechanism of segmental UPD/UPP formation in cancer (42). In fact, mitotic recombination as a result of double strand breaks on the long arm of chromosome 5 has been suggested to be the most frequent mechanism to achieve cnLOH affecting *APC* in patients with familial adenomatous polyposis (43). Moreover, a high frequency of segmental UPD has been described in base excision repair deficient *MUTYH*-associated polyposis carcinomas (44). In our sample set, the frequency of UPD in the genomic region 5q21 is as high as the frequency of genomic loss, suggesting that UPD is indeed an alternative mechanism to acquire biallelic inactivation of *APC*. The higher prevalence of UPD/UPP in chromosome 5q, as well as in other chromosomes, compared to the frequency of genomic loss, suggests that tumor cells need to preserve such chromosomes in two or more copies in order to maintain cellular viability and fitness. Despite the fact that most of the cases showed segmental UPD for 5q, one specific sample displayed cnLOH involving the centromere region of chromosome 5, thus indicating that one possibility for the origin of such rearrangement was through a chromosome segregation error. As FISH analysis proved that 16% of cells in this tumor showed monosomy or trisomy for the chromosome arm 5q, it is plausible to hypothesize that a non-disjunction event occurred early in the tumor development and originated the UPD, which resulted in biallelic representation of a nonsense mutation required for tumor progression. Furthermore, the TCGA dataset showed several samples with whole chromosome 5 UPD events. Altogether, this would indicate that mitotic recombination is not the only mechanism to generate cnLOH affecting *APC*, and a missegregation event resulting in a whole chromosome UPD/UPP might as well represent the second hit to achieve homozygosity for an inactivating mutation.

In summary, sporadic CRC shows a specific pattern of UPD/UPP, which greatly coincides with regions of recurrent genomic losses. Further research efforts are needed to determine whether other cancer types present unique landscapes of UPD/UPP as well. The integrative analysis of regions with UPDs/UPPs, CNAs, methylation patterns, and mutational gene status provided further evidence of novel putative TSGs. In fact, we propose that UPD/UPP events act as the second hit suggested by Knudson. Finally, our data show that cnLOH affecting *APC* can occur by either chromosome segregation errors or mitotic recombination events, thus generating whole chromosome or segmental UPD, respectively.

Supplementary material

Supplementary Tables 1–3 can be found at <http://carcin.oxford-journals.org>

Funding

Intramural program of National Institutes of Health; Ministerio de Educación y Ciencia (SAF2007-64167); Ministerio de Economía y Competitividad (SAF2012-40017-C02-02); RTICC and FEDER (RD06/0020/1020); Generalitat de Catalunya (2009 SGR 1107, 2014 SGR 0135, 2014 SGR 0903); Catalanian Biobank Network; the European Commission (COLONGEVA); the Universitat Autònoma de Barcelona (PIF fellowship to K.T.); Asociación Española Contra el Cáncer (AIO2011 to J.C.); Instituto de Salud Carlos III (CP13/00160 to J.C.); CIBERehd program (J.J.L.); the International Agency for Research on Cancer (Fellowship program to N.F.J.) and Spanish Ministry (FPU program to I.Q.).

Acknowledgements

We thank Drs Jean-Claude Zenklusen, Margaret C. Cam and Yuri Kotliarov from the Neuro-Oncology Branch, NCI/NIH for their technical and bioinformatics support, and Drs Xavier Estivill and Blanca Espinet for supplying BAC clones.

Conflict of Interest Statement: None declared.

References

- The Cancer Genome Atlas Network. (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487, 330–337.
- Beroukhi, R. et al. (2010) The landscape of somatic copy-number alteration across human cancers. *Nature*, 463, 899–905.
- Lapunzina, P. et al. (2011) The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biol. Cell*, 103, 303–317.
- Tuna, M. et al. (2009) Uniparental disomy in cancer. *Trends Mol. Med.*, 15, 120–128.
- Makishima, H. et al. (2011) Pathogenesis and consequences of uniparental disomy in cancer. *Clin. Cancer Res.*, 17, 3913–3923.
- Engel, E. (1980) A new genetic concept: uniparental disomy and its potential effect, isodisomy. *Am. J. Med. Genet.*, 6, 137–143.
- Frankel, A. et al. (2014) Genome-wide analysis of esophageal adenocarcinoma yields specific copy number aberrations that correlate with prognosis. *Genes Chromosomes Cancer*, 53, 324–338.
- Beà, S. et al. (2009) Uniparental disomies, homozygous deletions, amplifications, and target genes in mantle cell lymphoma revealed by integrative high-resolution whole-genome profiling. *Blood*, 113, 3059–3069.
- O'Keefe, C. et al. (2010) Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. *Blood*, 115, 2731–2739.
- Siegel, R. et al. (2012) Cancer statistics. *CA Cancer J. Clin.*, 62, 10–29.
- Ried, T. et al. (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.
- Knudson, A.G. Jr. (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA*, 68, 820–823.
- Andersen, C.L. et al. (2007) Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis*, 28, 38–48.
- Melcher, R. et al. (2011) LOH and copy neutral LOH (cnLOH) act as alternative mechanism in sporadic colorectal cancers with chromosomal and microsatellite instability. *Carcinogenesis*, 32, 636–642.
- Kurashina, K. et al. (2008) Chromosome copy number analysis in screening for prognosis-related genomic regions in colorectal carcinoma. *Cancer Sci.*, 99, 1835–1840.
- Gaasenbeek, M. et al. (2006) Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. *Cancer Res.*, 66, 3471–3479.
- Melcher, R. et al. (2007) SNP-Array genotyping and spectral karyotyping reveal uniparental disomy as early mutational event in MSS- and MSI-colorectal cancer cell lines. *Cytogenet. Genome Res.*, 118, 214–221.
- Zarzour, P. et al. (2015) Single nucleotide polymorphism array profiling identifies distinct chromosomal aberration patterns across colorectal adenomas and carcinomas. *Genes Chromosome Cancer*, 54, 303–314.
- Prat, E. et al. (2008) Genomic imbalances in urothelial cancer: intratumor heterogeneity versus multifocality. *Diagn. Mol. Pathol.*, 17, 134–140.
- Bengtsson, H. et al. (2009) A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6. *Bioinformatics*, 25, 2149–2156.
- Olshen, A.B. et al. (2011) Parent-specific copy number in paired tumor-normal studies using circular binary segmentation. *Bioinformatics*, 27, 2038–2046.
- Zhang, H. et al. (2013) RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics*, 14, 244.
- Olshen, A.B. et al. (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, 5, 557–572.
- Zhao, M. et al. (2013) TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res.*, 41, D970–D976.
- Bibikova, M. et al. (2009) GoldenGate assay for DNA methylation profiling. *Methods Mol. Biol.*, 507, 149–163.
- Benjamini, Y. et al. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*, 57, 289–300.
- Goel, A. et al. (2010) Aberrant DNA methylation in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology*, 138, 1854–1862.
- van Puijenbroek, M. et al. (2008) Genome-wide copy neutral LOH is infrequent in familial and sporadic microsatellite unstable carcinomas. *Fam. Cancer*, 7, 319–330.
- Beroukhi, R. et al. (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl. Acad. Sci. USA*, 104, 20007–20012.
- Dulak, A.M. et al. (2012) Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. *Cancer Res.*, 72, 4383–4393.
- Zimonjic, D.B. et al. (2012) Role of DLC1 tumor suppressor gene and MYC oncogene in pathogenesis of human hepatocellular carcinoma: potential prospects for combined targeted therapeutics (review). *Int. J. Oncol.*, 41, 393–406.
- Xue, W. et al. (2008) DLC1 is a chromosome 8p tumor suppressor whose loss promotes hepatocellular carcinoma. *Genes Dev.*, 22, 1439–1444.
- Low, J.S. et al. (2011) A novel isoform of the 8p22 tumor suppressor gene DLC1 suppresses tumor growth and is frequently silenced in multiple common tumors. *Oncogene*, 30, 1923–1935.
- Peng, H. et al. (2013) Downregulation of DLC-1 gene by promoter methylation during primary colorectal cancer progression. *Biomed. Res. Int.*, 2013, 181384.
- Nassif, N.T. et al. (2004) PTEN mutations are common in sporadic microsatellite stable colorectal cancer. *Oncogene*, 23, 617–628.
- Angus-Hill, M.L. et al. (2011) T-cell factor 4 functions as a tumor suppressor whose disruption modulates colon cell proliferation and tumorigenesis. *Proc. Natl. Acad. Sci. USA*, 108, 4914–4919.
- Wu, R.C. et al. (2014) The emerging roles of ARID1A in tumor suppression. *Cancer Biol. Ther.*, 15, 655–664.
- Camps, J. et al. (2008) Chromosomal breakpoints in primary colon cancer cluster at sites of structural variants in the genome. *Cancer Res.*, 68, 1284–1295.
- Rawson, J.B. et al. (2011) Promoter methylation of Wnt antagonists DKK1 and SFRP1 is associated with opposing tumor subtypes in two large populations of colorectal cancer patients. *Carcinogenesis*, 32, 741–747.
- Bignell, G.R. et al. (2010) Signatures of mutation and selection in the cancer genome. *Nature*, 463, 893–898.
- Rajaram, M. et al. (2013) Two distinct categories of focal deletions in cancer genomes. *PLoS One*, 8, e66264.
- Moynahan, M.E. et al. (2010) Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.*, 11, 196–207.
- Howarth, K. et al. (2009) A mitotic recombination map proximal to the APC locus on chromosome 5q and assessment of influences on colorectal cancer risk. *BMC Med. Genet.*, 10, 54.
- Middeldorp, A. et al. (2008) High frequency of copy-neutral LOH in MUTYH-associated polyposis carcinomas. *J. Pathol.*, 216, 25–31.

Supplementary Table I. APC mutation cluster region PCR primers

Primer	Sequence (5'-3')	Fragment start (bp)^a	Fragment end (bp)^a
<i>APC</i> MCR A_For	GGAAGGCCAAAGTCCTTCACA	3081	3210
<i>APC</i> MCR A_Rev	TGTCCAAAATGTGGTTGGAA		
<i>APC</i> MCR B_For	TTCCTTCATCACAGAAACAGTCA	3529	3776
<i>APC</i> MCR B_Rev	CTTCAGCTGATGACAAAGATGAT		
<i>APC</i> MCR C_For	GCACAGAGTAGAAGTGGTCAGC	3674	3898
<i>APC</i> MCR C_Rev	TGGAACTTCGCTCACAGGAT		
<i>APC</i> MCR D_For	CAGACGACACAGGAAGCAGA	3837	4077
<i>APC</i> MCR D_Rev	AACATGAGTGGGGTCTCCTG		
<i>APC</i> MCR E_For	CCAAAAGTGGTGCTCAGACA	4051	4264
<i>APC</i> MCR E_Rev	AGCTGTTTGAGGAGGTGGTG		
<i>APC</i> MCR F_For	CATGCCACCAAGCAGAAGTA	4256	4448
<i>APC</i> MCR F_Rev	CACTCAGGCTGGATGAACAA		
<i>APC</i> MCR G_For	TTTGCCACGGAAAGTACTCC	4437	4681
<i>APC</i> MCR G_Rev	TGTTGGCATGGCAGAAATAA		
<i>APC</i> MCR H_For	TGAAACAGAATCAGAGCAGCC	4572	4767
<i>APC</i> MCR H_Rev	ACAGGCAGCTGACTTGGTTT		

MCR, mutation cluster region.

^aBased on the alignment of the hg19_refGene_NM_001127511 cDNA.

3.3 Capítol II

Keyvan Torabi, Pau Erola, Marcos Díaz-Gay, Antoni Castells, Juanjo Lozano, Sergi Castellví-Bel, Rosa Miró, Imma Ponsa and Jordi Camps.

Quantitative analysis of somatically acquired uniparental disomy in gastrointestinal cancers.

Pendent d'enviar.

Quantitative analysis of somatically acquired uniparental disomy in gastrointestinal cancers

Keyvan Torabi^{1,2,§}, Pau Erola^{3,4,§}, Marcos Díaz-Gay², Antoni Castells², Juanjo Lozano³, Sergi Castellví-Bel², Rosa Miró^{1,5}, Imma Ponsa^{1,5,†} and Jordi Camps^{1,2,†,*}

¹Unitat de Biologia Cel·lular i Genètica Mèdica, Departament de Biologia Cel·lular, Fisiologia i Immunologia, Facultat de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, 08193, Spain.

²Gastrointestinal and Pancreatic Oncology Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Barcelona, Catalonia, 08036, Spain.

³Bioinformatics Unit, CIBERehd, Barcelona, Catalonia 08036, Spain.

⁴Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, Scotland, EH25 9RG, UK.

⁵Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, 08193, Spain.

§These authors contributed equally to this work.

†These authors contributed equally to this work.

*Corresponding author:

Jordi Camps, Ph.D.

Gastrointestinal and Pancreatic Oncology Group

Institut D'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS)

Rosselló 153, 4th floor, 08036 Barcelona, Spain.

Phone: +34-93-2275400 ext.4560; Fax: +34-93-3129405;

E-mail: jcamp@clinic.cat

Abstract

Somatically acquired uniparental disomies (UPDs) are frequent events in solid tumors and have been recurrently associated with cancer-related genes malfunctioning. Nevertheless, studies assessing their functional consequences across several cancer types are necessary. In the present study, we aimed at integrating UPD profiles with the mutational status of a TSG panel in a tumor-type specific manner by using 548 gastrointestinal (GI cancers) from the TCGA including colon adenocarcinoma (COAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), esophagus adenocarcinoma (ESAD) and esophagus squamous cell carcinoma (ESCC) cohorts. Our results showed that UPD profiles in COAD matched with those identified in READ cohorts. In addition, UPD profiles of STAD, ESAD and ESCC showed significant differences from those identified in COAD and READ cohorts, confirming the existence of a cancer-specific landscapes of UPDs. Moreover, our analysis indicated that such events act as a “second-hit”, in the Knudson’s model in order to achieve biallelic inactivation of TSGs. Specifically, *APC*, *ARID1A* and *NOTCH1* were recurrently inactivated by UPDs in CRC, STAD and ESCC, respectively. Furthermore, while *TP53* showed inactivation caused by UPDs at chromosome arm 17p across all tumor-types, copy number losses were much more frequent for this genomic position. By inferring genome ploidy, we demonstrated that an increased number of UPD events, both affecting the whole chromosome or segmental, were detected in highly aneuploid genomes compared to near-diploid tumors. Finally, our analysis showed that *APC* inactivation in CRC is mainly involving a telomeric UPD at chromosome arm 5q, suggesting a cluster of mitotic recombination breakpoints. In summary, our study defines specific profiles of somatically acquired UPDs in GI cancers and provides further evidences of such event as a mechanism to inactivate TSGs.

Introduction

Copy number alterations (CNAs) are a hallmark for several types of cancer, including gastrointestinal (GI) cancers (GI), providing a consistent landscape of gains and losses in a tumor-type specific manner (Beroukhi et al. 2010b; Zack et al. 2013b). A huge challenge in genome-wide cancer studies is to distinguish driver CNAs, which lead to development, progression and maintenance of tumors, from those considered passenger events, resulting as a consequence of high rates of chromosome instability (Beroukhi et al. 2007a). Therefore, several efforts have been made in order to identify tumor suppressor genes (TSGs) in recurrent deleted regions, based on the “two-hit” hypothesis (Knudson 1971). In this model, one copy of a TSG is inactivated by a nonsense or frameshift mutation, and the other copy is inactivated by a similar mutation or by loss of heterozygosity (LOH) acting as the “second hit”. However, specific regions of the genome, e.g. fragile sites, have often limited the association between deletions and driver TSGs (Rajaram et al. 2013a). Thus, the integration of the gene mutational status with regions frequently involved in genomic imbalances has become essential in order to report functional inactivation of genes (Bignell et al. 2010b). More recently, the application of SNP arrays allowed the possibility to profile copy neutral LOH (cnLOH) or somatically acquired uniparental disomies (hereafter referred as UPDs) in cancer (Tuna et al. 2009; Makishima and Maciejewski 2011; Lapunzina and Monk 2011). UPD was firstly described as a constitutional event, consisting in the inheritance of two copies of chromosomes coming from the same parental origin caused by a meiotic error and leading to development disorders (Engel 1980a). However, such an event has been recurrently observed in several cancer malignancies, including both hematological and solid neoplasms (Tuna et al. 2009; O’Keefe et al. 2010), thus becoming a somatically acquired genomic event. In this context, somatic UPDs have arisen as an alternative mechanism to act as the “second hit” in Knudson’s model. In fact, somatically acquired UPD events have already been described involving several cancer-related driver genes (Melcher et al. 2011; Zarzour et al. 2015). Similar to genomic losses, the mutational status of genes located in recurrent regions of UPD can provide evidence to determine whether a UPD event is indeed a driver alteration in cancer (Torabi et al. 2015). Yet, few studies have focused on profiling such events in a genome-wide and tumor-type specific manner.

Furthermore, whole genome duplication (WGD), giving rise to tetraploid genomes, has been accepted as a common event in several tumor types and has been postulated as a driver event in the progression of cancer (Storchova and Pellman 2004; Zack et al. 2013b). The consequence of a genome tetraploidization includes the acquisition of numerical chromosome instability, which is defined by the increasing rate of mitotic segregation errors. Recently, the usage of bioinformatic tools has allowed the assessment of allele-specific copy number and, consequently, the identification of the tumor ploidy (Van Loo et al. 2010; Carter et al. 2012). Thus, it has been systematically established that highly aneuploid genomes are not rare in cancer, specially in epithelial tumors, but very common (Storchova and Kuffer 2008; Zack et al. 2013b). In such a

polyploid scenario, recurrent losses of the same parental chromosome could originate recurrent UPDs to functionally inactivate of TSGs (Segditsas et al. 2009).

In the present study, we aimed at integrating a map of UPDs and the mutational status of TSGs located in these recurrent regions, thus providing insights into the mechanisms of bi-allelic inactivation in cancer. Moreover, by inferring tumor ploidy, we also determined to which extent highly aneuploid genomes are enriched by UPD events and the distribution of breakpoints of these events along specific chromosomes.

Materials and methods

TCGA samples or cohorts

Data from five tumor-types of the gastrointestinal tract were obtained from TCGA project through the NCI's Genomic Data Commons (GDC) data portal (<https://portal.gdc.cancer.gov/>). The samples for each tumor-type consisted in 434 paired samples of colon (COAD), 155 samples of rectum (READ), 325 of stomach (STAD) and 120 of esophagus (ESCA) carcinomas. Esophagus cohort included two datasets: esophageal adenocarcinoma (ESAD) and esophageal squamous cell carcinoma (ESCC).

SNP array analysis

Affymetrix SNP 6.0 array data was used from the abovementioned cohorts, consisting in tumor-matched normal blood samples. To identify copy number (CN) changes and loss of heterozygosity (LOH) microarray data segmentation was performed using Paired Parent-Specific Circular Binary Segmentation (Paired PSCBS) method implemented in PSCBS package (Olshen et al. 2011). PSCBS uses a parametric bootstrapping technique to estimate the different allelic mean levels. LOH call is tested on the segments that are not in allelic balance and is positive on those cases where the allelic CN is under a certain threshold, derived from data that contemplates the background signal. Unfortunately, the background signal estimation may fail on tumors that do not present LOH, so we discarded those samples with a background signal $\Delta\text{LOH} > 0.75$. After the segmentation, those segments smaller than 2.5 Mbps were eliminated and two segments were considered the same if the gap between them was shorter than 2.5 Mbp. Of all the paired samples examined in this study, only those that presented less than 300 segments were considered in further analyses.

Whole exome sequencing analysis

Whole exome sequencing data were extracted from all cohorts mentioned before from the GDC data portal. In this regard, single nucleotide variants (SNVs) and short insertions and deletions (INDELs) were considered. Annotated variant calling files were used in the form of open-access Mutation Annotation Format (MAF) tab-delimited text files. In order to base the somatic variant calling analysis on GATK Best Practices, MuTect2-derived MAF files were selected (DePristo et al. 2011). In order to allow an accurate annotation of missense variants with Oncotator (<http://portals.broadinstitute.org/oncotator/>), genomic coordinates were switched from genome build GRCh38 (hg38) to GRCh37 (hg19) using the NCBI Genome Remapping Service (<https://www.ncbi.nlm.nih.gov/genome/tools/remap>). In addition, matching with PSCBS results through an in-house automated pipeline was performed in R (R Core Team 2017). Regarding SNVs and INDELs pathogenicity, truncating variants (nonsense, frameshift and affecting splicing variants) were directly considered, whereas in the case of missense variants six different prediction tools were assessed: PhyloP (deleteriousness threshold of phyloP46way_placental score ≥ 1.6), SIFT (prediction of damaging), PolyPhen2 (HumVar prediction of probably damaging or possibly

damaging), MutationTaster (prediction of disease-causing or disease-causing-automatic), LRT (prediction of deleterious) and CADD (Phred score ≥ 15). Only those missense variants predicted as pathogenic by at least 3 of the tools were considered for further analysis. Additionally, genes obtained from the TSGene database (<https://bioinfo.uth.edu/TSGene/>) were considered in order to identify those variants affecting TSGs (Zhao et al. 2016).

FISH on TMAs

Twenty colorectal adenocarcinomas provided by the Hospital Clinic of Barcelona/IDIBAPS Biobank were included in this study. Clinical features of each patient were previously described (Torabi et al. 2015). All patients signed the corresponding informed consent and the sample collection was approved by the local Ethics Committees. FISH analyses were performed in tissue microarrays of 4-5 μ m thickness containing two replicates of both tumor and normal adjacent mucosa for each sample. Centromeric probes for chromosomes 7 and 15 labeled in green, and for chromosomes 18 and 20 labeled in orange were used according the manufacturer's recommendations (Vysis Inc, Downers Grove, IL). Pretreatment included three xylene incubations, increasing concentration of ethanol series, permeabilization with EDTA and treated with pepsin. Next, slides were incubated in 1X PBS with MgCl₂ and fixed with 1% paraformaldehyde. Denaturation was performed in a Thermo Brite (Vysis) at 78°C during 6 minutes for panel 1 probe series and 85°C during 3 minutes for panel 2 probe series. Hybridization was performed at 37°C overnight. Post-hybridization washes were performed in 0.4X SSC/0.3% NP40 at 74°C for 2 minutes and 2X SSC/0.1% NP40 at room temperature during 1 minute. A minimum of 100 cells was imaged with a Nikon Eclipse 50i fluorescence microscope using the Isis Fluorescence Imaging System (MetaSystems, Altussheim, Germany).

Statistical Analysis

The Mann-Whitney sum-rank test was used in order to compare the number of UPD events between tumor-types and also between highly aneuploid and diploid genomes. Correlation analysis was applied when comparing ploidy values extracted from FISH and ASCAT methods. The software GraphPad Prism 6.0 (GraphPad) was used to calculate p-values and to depict the results.

Results and Discussion

UPD profiling in GI cancers

Genome-wide UPD analysis was performed in 265 colon (COAD cohort), 105 rectum (READ cohort), 121 stomach (STAD cohort) and 57 esophagus (ESCA cohort) tumor-matched normal carcinomas extracted from level 1 TCGA data portal by applying the algorithm Paired PSCBS. Most of the samples showed, at least, one genomic region affected by UPD (85.38% in COAD, 88.89% in READ, 91.74% in STAD and 96.49% in ESCA cohorts). In particular, ESCA was the tumor type showing the highest number of regions with UPD, with a median of five events per sample, compared to the other tumor types ($P < 0.05$). In contrast, COAD showed a median of three UPD events per sample being the tumor type with statistically significant lower number of UPDs ($P < 0.05$) (Supplementary Fig. 1a). We then classified UPDs in two types of events: whole chromosome and segmental UPDs, the last including both telomeric and interstitial fragments. Overall, segmental UPDs were the most frequent alterations with a mean of 3.39 events per sample across all cohorts compared to 0.87 UPDs affecting whole chromosome events ($P < 0.0001$). Moreover, we observed a higher amount of segmental UPDs in the upper GI tract compared to colon and rectum (mean of 4,65 for ESCA and STAD vs mean of 2,84 for COAD and READ; $P < 0,0001$) (Supplementary Fig. 1b). Altogether, our analysis suggested that all these cancer types displayed high incidences of chromosomal instability, both numerical and structural. Remarkably, the higher rate of segmental UPDs in stomach and esophagus cancers suggested higher levels of structural CIN in these cancer types compare to colon and rectum adenocarcinomas. This prevalence of structural CIN in gastric and esophageal adenocarcinomas was previously suggested in a study showing more focal amplifications events in upper GI tract compared to COAD and READ (Dulak et al. 2012). Higher rates of double strand breaks formation, caused by the exposure to bile acids, might explain these results in STAD and ESCA (Bernstein et al. 2005).

In order to establish an overview of recurrent regions affected by UPD in a tumor-type manner, circos plots were generated for each individual cohort (Fig. 1a-e). In addition, CNA profiles were also calculated in order to compare all chromosome imbalances, both genomic gains and losses. An overlay between regions of UPD and genomic losses is depicted in the corresponding plots ($P < 0,001$). The most frequently affected regions by UPDs for each cohort are listed in Table 1. Our results showed that UPD profiles for the COAD cohort mirrored those identified in the READ cohort, resembling the observations as far as genomic copy number alterations is concerned. Nevertheless, the UPD profile of the STAD, ESAD and ESCC cohorts showed significant differences compared to the COAD and READ cohorts. Overall, STAD, ESAD and ESCC cohorts displayed similar regions affected by UPD, although some specific differences were also detected (Fig. 1a, b and c). First of all, we identified that chromosome arm 17p13.3-p13.1 showed the highest frequency of UPD events in esophagus and stomach adenocarcinomas (up to 55.6% and 25.7% in ESAD and STAD, respectively), and it was the second most frequent region in esophageal squamous cell carcinoma, rectal and colon adenocarcinomas (up to 53.8%, 26.7% and 17.4% in ESCC, READ and COAD,

respectively). When including copy number alterations in the analysis, chromosome arm 17p also showed high frequencies of copy number losses across all cohorts, thus being the most altered genomic region in the tumor types we studied. The second most prevalent genomic region affected by UPD across all the tumor types was the chromosome arm 9q. Specifically, UPDs affecting 9q21.11-q34.3 were present in 64.1% of ESCC samples, and in 27.8% and 21.5% of ESAD and STAD samples, respectively. Thus, the region 9q21.11-q34.3 in esophagus squamous cell carcinomas was the most frequently affected region by UPD across all cohorts (Fig. 1a). Finally, UPDs at 5q12.1-q35.3 were the most common events in colon and rectum carcinomas, affecting up to 20.4% and 32.4% of the samples, respectively (Fig. 1d and e). While this alteration was present in a maximum frequency of 22.2% in esophagus adenocarcinomas, it was rarely detected in stomach and esophagus squamous cell carcinomas, showing a frequency of 9.1% and 10.7%, respectively. In contrast, these latest cohorts showed high frequencies of copy number losses in this region of chromosome 5q (61.5% in ESCC, 50.0% in ESAD and 38.8% in STAD datasets), suggesting different mechanisms to achieve loss of heterozygosity. In summary, our results demonstrate tumor-type specific UPD profiles. Whereas we identified UPDs affecting chromosome 9q21.11-q34.3 in STAD, ESAD and ESCC cohorts, we rarely detected this alteration in COAD or READ. Furthermore, we confirmed that the UPD at 5q is the most frequent event in COAD and READ (Lips et al. 2007; Melcher et al. 2011; Zarzour et al. 2015; Torabi et al. 2015). These results suggest that the genomic loss of chromosome 5q in colorectal cancer might be more detrimental for the tumor cells compared to stomach and esophagus carcinomas.

Functional consequences of UPDs on tumor suppressor genes

In order to explore whether UPDs contribute to the inactivation of TSGs, we assessed the exomes from all five tumor types extracted from the MAF masked files of the TCGA. Only high-impact variants affecting TSGs (i.e., protein truncating and damaging missense mutations) were considered in this analysis. Our results showed that the STAD cohort is the tumor type with the highest amount of mutations per sample (mean of 19.2 mutations). Means of deleterious variants corresponding to COAD, READ, ESAD and ESCC cohorts were 13.7, 7.7, 7.2 and 5.5, respectively. We then evaluated which of these variants were located in regions of UPD. Our analysis indicated that ESAD showed the highest percentage of mutations in regions affected by UPDs (19.2%), while the rest of the tumor types showed the frequencies of 15.8% in ESCC, 10.4% in STAD, and 8.3% in COAD, 6.6% in READ cohorts.

Next, by integrating the mutational status of TSGs and genomic profiles, we assessed which specific genes were inactivated as a consequence of a UPD. Common genes inactivated by UPDs and CN losses, possibly representing the second hit, are listed in Table 1. Our results identified that *TP53* was simultaneously mutated and affected by a UPD event in 50%, 38.5%, 11.5%, 10% and 9.8% of the cases in ESAD, ESCC, READ, STAD and COAD, respectively. Moreover, a mutation in this TSG accompanied by a copy number loss was detected in 44.9%, 38.8%, 28.9%, 20.5% and 17.5% of the cases in READ, ESAD, COAD, ESCC and STAD, respectively. Therefore, our analysis

shows that tumor samples with *TP53* mutated are recurrently perturbed by second hit events, either UPD or genomic losses, across all tumor types. Likewise, our analysis showed several examples of TSGs inactivated by UPDs in a tumor type dependent manner. In colon and rectum adenocarcinomas, 22.2% and 29.9% of the samples, respectively, showing an inactivating mutation for *APC* also presented UPD events at 5q21, while only 16.2% and 20.9% of samples displayed copy number losses. In esophagus squamous cell carcinoma, the second most mutated gene involved in regions with UPD was *NOTCH1*. Strikingly, in all samples where an inactivating mutation at *NOTCH1* was detected, a UPD including this gene was also identified. Finally, in stomach adenocarcinomas, *ARID1A* was the second most frequently inactivated gene involved in UPDs. However, the chromosome region containing *ARID1A*, 1p35.3, was found in UPDs only in 13.2% of the STAD cases, suggesting that unknown TSGs in other regions might also be affected by UPDs. To conclude this section, it is important to highlight that our results confirmed that UPD events represent an important mechanism to functionally inactivate *APC* in colon and rectum adenocarcinomas (Torabi et al. 2015). Moreover, we also identified *NOTCH1* in esophagus squamous cell carcinomas and *ARID1A* in stomach adenocarcinomas as TSGs involved in UPDs events, suggesting that the second hit to inactivate these TSGs is through these events. Finally, we detected that *TP53* was frequently inactivated by UPD across different tumor types. These results altogether highlight the importance of UPD in cancer as a mechanism to induce the inactivation of TSGs.

UPD is frequently detected in highly aneuploid genomes

Next, we aimed at assessing to which extent the total DNA content might be contributing to the generation of UPD events. First, we sought to validate the ploidy results provided by ASCAT by performing FISH analysis in a cohort of colorectal adenocarcinomas previously studied in our laboratory (Torabi et al. 2015). We quantified the chromosomes 7, 15, 18 and 20 using specific centromeric probes. Our results showed that chromosomes 7 and 20 were recurrently gained with a median of 3.58 copies per chromosome 7 [2.02-5.59] and a median of 2.90 copies for chromosome 20, [1.88-6.01]. On the other hand, chromosome 18 was mainly lost showing a median of 1.49 copies [1.09-3.45]. Chromosome 15, which usually shows copy number neutrality in CRC, showed a median of 1.86 copies [1.18-3.16] (Supplementary Fig. 2a). In order to infer the genome ploidy based on the FISH results, we used the weighted mean copy number of all four chromosomes in each sample. By applying this, we detected that 60% of the samples showed a highly aneuploid genomes, which were defined by ploidy values higher than 2.5 (Table 2). Similarly, ASCAT identified 50% of cases with a highly aneuploid genome, confirming that ASCAT output was experimentally validated (Table 2). In fact, ploidies inferred from both methodologies showed a positive correlation between them ($R = 0.6$; $P < 0.01$) (Supplementary Fig. 2b). Therefore, our results further confirmed that most epithelial cancers contain highly aneuploid genomes (Zack et al. 2013b). These results are compatible with the hypothesis that samples with high aneuploid genomes might have undergone a whole

genome doubling event and a subsequent loss of several chromosomes (Storchova and Pellman 2004).

Next, we explored the ploidy computed by ASCAT from TCGA cohorts. We observed that COAD and STAD cohorts were showing the lowest genome ploidy, with a median of 2.44 and 2.60, respectively. On the other hand, READ showed a median of 2.96, and ESCA showed a median of 3.07. We then aimed at comparing the frequency of UPD events in highly aneuploid genomes versus the near-diploid samples. Overall, our results showed that highly aneuploid tumors displayed a significantly higher amount of UPD events than near-diploid tumors ($P < 0.001$) (Fig. 2a). In addition, when UPD events were classified in whole chromosome and segmental UPDs, we also detected that the number of these events was higher in highly aneuploidy tumors ($P < 0.001$ for whole chromosome and $P < 0.0001$ for segmental UPDs) (Fig. 2b,c). To the best of our knowledge, we report, for the first time, evidence that highly aneuploid tumor genomes show higher prevalence of generating UPDs.

Comparison between UPDs and genomic losses of chromosome 5q in colorectal tumors

As UPDs and genomic losses were the most common events to inactivate *APC* in COAD and READ cohorts, we mapped the breakpoints of any segment involving this gene. As mentioned earlier, UPDs were more frequently observed than genomic losses in these samples, thus suggesting that they represent the main event to generate a complete inactivation of *APC*. Our analysis indicated that 76.74% of the cases showed telomeric UPDs affecting the majority of the chromosome arm 5q (median size was 111 Mb [72.27-136.51Mb]). In addition, 20.93% of samples showed UPDs affecting the whole chromosome 5, and only a 2.33% displayed interstitial UPDs. When considering copy number losses, 42% of samples showed interstitial deletions, (median size was 41.23 Mb [3.79-120.75 Mb]), 37.68% showed arm-level events with a peri-centromeric breakpoint spanning up to the telomere, and the remaining 20.29% corresponded to whole chromosome losses. We then assessed the localization of the breakpoints of each alteration affecting chromosome 5. We identified differentially localized breakpoints involving UPDs or copy number losses events. Telomeric UPDs showed breakpoints extending from a location between the centromere and the *APC* locus on chromosome 5q to the long-arm telomere, with a specific prevalence close to the centromere or in a region close to 70-80 Mbp (Fig. 3). In contrast, breakpoints corresponding to copy number losses were either affecting peri-centromeric loci, resulting in arm-level losses, or a region close to 80 Mbp up and spanning to the *APC* locus, thus generating interstitial deletions. (Fig. 3). Surprisingly, none of the breakpoints associated to copy number losses were localized within the region of 70-80 Mbp, where the most recurrent breakpoints for UPD events detected. Our analysis confirmed that bi-allelic inactivation of *APC* is mainly achieved by telomeric UPDs caused by mitotic recombination events (Lips et al. 2007; Melcher et al. 2011; Zarzour et al. 2015; Torabi et al. 2015). These results confirm previous findings in familial adenomatous polyposis patients, where similar genomic breakpoints were detected and were also associated with mitotic recombination events (Howarth et al. 2009). Altogether, this results further strength the

hypothesis that colorectal tumor cells urge to maintain a disomy for chromosome 5 despite the benefit of inactivating *APC*.

Acknowledgments

The authors would like to thank the Biobank Platform from IDIBAPS for sample set preparation. This work was supported by grants from the European Commission (COLONGEVA to J.C.), the Instituto de Salud Carlos III and co-funded by the European Regional Development Fund (ERDF) (CP13/00160 to J.C.), the CIBEREHD program, the Agència de Gestió d'Ajuts Universitaris i de Recerca, Generalitat de Catalunya (2014 SGR 135 and 2014 SGR 903). K.T. received a PIF-fellowship from Universitat Autònoma de Barcelona.

Conflict of Interest: The authors declare that they have no conflict of interest.

Figure Legends

Fig. 1: Genomic profiles of UPD and CNAs in GI tract cancer. Circos plots showing frequencies of UPDs, losses and gains in Esophageal Squamous Cell Carcinomas **(a)**, Esophageal Adenocarcinomas **(b)**, Gastric Adenocarcinomas **(c)**, Colon Adenocarcinomas **(d)** and Rectal Adenocarcinomas **(e)**. Chromosomes are represented in the outer ring of the circos plots. Regions of UPD are represented in green in the outer ring, and copy number losses and gains are depicted in blue and red, respectively, in the following rings.

Fig. 2: Quantification of UPDs in highly aneuploid and near-diploid tumors. Box-plots showing significant differences between highly aneuploid and near-diploid genomes of all UPDs events **(a)** whole-chromosome UPDs **(b)** and segmental UPDs **(c)**. The Mann-Whitney sum-rank test was used to compare number of UPDs events between the two groups of genome ploidies. *** $P < 0.001$, **** $P < 0.0001$.

Fig. 3: Location of telomeric UPDs and copy number losses breakpoints in chromosome arm 5q. Histogram showing the breakpoints of telomeric UPDs and copy number losses along chromosome 5q. Chromosome positions were binned in groups of 5 Mbp.

Supplementary Figure Legends

Fig. S1: Characterization of UPDs event across GI datasets. Box-plot showing significant differences of **(a)** UPD events between tumor-types. **(b)** UPD segmental events between upper- and lower-gastrointestinal tract.

Fig. S2: FISH and ASCAT ploidy values comparison. **(a)** Graph depicting weighed mean FISH signals of centromeric probes 7, 15, 18 and 20 in normal and tumor samples. **(b)** Correlation analysis between FISH and ASCAT ploidy values.

Table 1. UPD frequent regions

Cohort	Chromosome Region	Size (Mbp)	UPD (%)	Mutated gene	1st hit Mutation (n)	2nd hit UPD (n)	2nd hit CN Loss (n)
COAD	5q13.2-q35.3	110.1	10.2-20.4	<i>APC</i>	185	41	30
	6p25.3-21.1	42.2	10.6-15.1				
	17p13.3-q25.3	81.1	10.6-17.4	<i>TP53</i>	137	23	68
READ	5q12.1-q35.3	119.2	15.2-32.4	<i>APC</i>	67	20	14
	14q11.2-q32.33	88.2	11.4-15.2				
	17p13.3-p11.2	21.8	16.2-26.7	<i>TP53</i>	60	9	35
	17q21.32-q25.3	33.7	15.2-19.0				
ESAD	9q31.1-q34.3	37.6	27.8				
	17p13.3-p11.2	21.5	27.8-55.6	<i>TP53</i>	17	9	7
	19p13.3-p13.11	16.1	27.8-33.3				
ESCC	3p26.3-p22.3	33.7	20.5				
	9p24.3-q34.3	141	28.2-64.1	<i>NOTCH1</i>	4	4	0
	17p13.3-q25.3	81.2	25.6-53.8	<i>TP53</i>	32	15	8
STAD	9p24.3-q34.3	141	10.7-21.5				
	17p13.3-q25.3	81.2	14.9-25.6	<i>TP53</i>	57	12	21
	19p26.3-p22.3	24.5	15.7-19.0				

Table 2. FISH and ASCAT ploidy

Sample	Chr 7 ^a	Chr 15 ^a	Chr 18 ^a	Chr 20 ^a	FISH ploidy ^b	ASCAT ploidy
2T	3.21	n.a.	1.61	n.a.	2.41	2.24
5T	4.42	1.79	1.63	2.30	2.53	1.93
6T	3.73	3.16	1.53	4.39	3.20	3.87
8T	3.38	1.18	1.16	2.83	2.14	2.89
9T	3.56	1.48	1.56	3.75	2.59	2.14
10T	2.95	1.90	1.09	1.99	1.98	1.98
11T	3.92	2.24	1.49	5.39	3.26	3.59
12T	2.38	n.a.	3.45	n.a.	2.92	4.04
13T	3.76	1.81	1.45	4.37	2.85	4.29
33T	3.12	1.83	1.19	1.88	2.01	2.10
37T	3.59	n.a.	1.62	n.a.	2.61	3.29
40T	2.68	1.80	1.48	2.07	2.01	1.97
41T	5.07	2.10	1.32	6.01	3.63	2.37
42T	4.63	1.46	1.67	3.24	2.75	3.33
43T	3.21	2.11	1.19	2.90	2.35	2.37
44T	2.22	3.01	1.74	3.69	2.67	2.68
45T	3.99	2.16	1.47	4.95	3.14	4.24
46T	3.39	2.15	1.35	2.69	2.39	2.15
47T	2.02	n.a.	1.96	n.a.	1.99	2.45
50T	3.81	n.a.	1.41	n.a.	2.61	3.31

^aValues indicated are weighted means of copy number of the corresponding chromosome.

^bMean of chromosomes 7, 15, 18 and 20 values.

n.a stands for not available.

Bibliography

- Bernstein H, Bernstein C, Payne C, et al (2005) Bile acids as carcinogens in human gastrointestinal cancers. *Mutat Res Mutat Res* 589:47–65.
- Beroukhir R, Getz G, Nghiemphu L, et al (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 104:20007–12.
- Beroukhir R, Mermel CH, Porter D, et al (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* 463:899–905.
- Bignell GR, Greenman CD, Davies H, et al (2010) Signatures of mutation and selection in the cancer genome. *Nature* 463:893–8.
- Carter SL, Cibulskis K, Helman E, et al (2012) Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 30:413–421.
- DePristo MA, Banks E, Poplin R, et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498.
- Dulak AM, Schumacher SE, van Lieshout J, et al (2012) Gastrointestinal Adenocarcinomas of the Esophagus, Stomach, and Colon Exhibit Distinct Patterns of Genome Instability and Oncogenesis. *Cancer Res* 72:4383–4393.
- Engel E (1980) A new genetic concept: uniparental disomy and its potential effect, isodisomy. *Am J Med Genet* 6:137–43.
- Howarth K, Ranta S, Winter E, et al (2009) A mitotic recombination map proximal to the APC locus on chromosome 5q and assessment of influences on colorectal cancer risk. *BMC Med Genet* 10:54.
- Knudson a G (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68:820–3.
- Lapunzina P, Monk D (2011) The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biol Cell* 103:303–317.
- Lips EH, de Graaf EJ, Tollenaar RAEM, et al (2007) Single nucleotide polymorphism array analysis of chromosomal instability patterns discriminates rectal adenomas from carcinomas. *J Pathol* 212:269–77.
- Makishima H, Maciejewski JP (2011) Pathogenesis and Consequences of Uniparental Disomy in Cancer. *Clin Cancer Res* 17:3913–3923.
- Melcher R, Hartmann E, Zopf W, et al (2011) LOH and copy neutral LOH (cnLOH) act as alternative mechanism in sporadic colorectal cancers with chromosomal and microsatellite instability. *Carcinogenesis* 32:636–42.
- O’Keefe C, McDevitt MA, Maciejewski JP (2010) Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. *Blood* 115:2731–9.
- Olshen AB, Bengtsson H, Neuvial P, et al (2011) Parent-specific copy number in paired tumor–normal studies using circular binary segmentation. *Bioinformatics* 27:2038–2046.
- R Core Team (2017) R: A Language and Environment for Statistical Computing.
- Rajaram M, Zhang J, Wang T, et al (2013) Two Distinct Categories of Focal Deletions in Cancer Genomes. *PLoS One* 8:e66264.
- Segditsas S, Rowan AJ, Howarth K, et al (2009) APC and the three-hit hypothesis. *Oncogene* 28:146–55.
- Storchova Z, Kuffer C (2008) The consequences of tetraploidy and aneuploidy. *J Cell Sci* 121:3859–3866.
- Storchova Z, Pellman D (2004) From polyploidy to aneuploidy, genome instability and cancer. *Nat Rev Mol Cell Biol* 5:45–54.

-
- Torabi K, Miró R, Fernández-Jiménez N, et al (2015) Patterns of somatic uniparental disomy identify novel tumor suppressor genes in colorectal cancer. *Carcinogenesis* 36:1103–1110.
- Tuna M, Knuutila S, Mills GB (2009) Uniparental disomy in cancer. *Trends Mol Med* 15:120–8.
- Van Loo P, Nordgard SH, Lingjaerde OC, et al (2010) Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci* 107:16910–16915.
- Zack TI, Schumacher SE, Carter SL, et al (2013) Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 45:1134–1140.
- Zarzour P, Boelen L, Luciani F, et al (2015) Single nucleotide polymorphism array profiling identifies distinct chromosomal aberration patterns across colorectal adenomas and carcinomas. *Genes Chromosomes Cancer* 54:303–14.
- Zhao M, Kim P, Mitra R, et al (2016) TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res* 44:D1023-31.

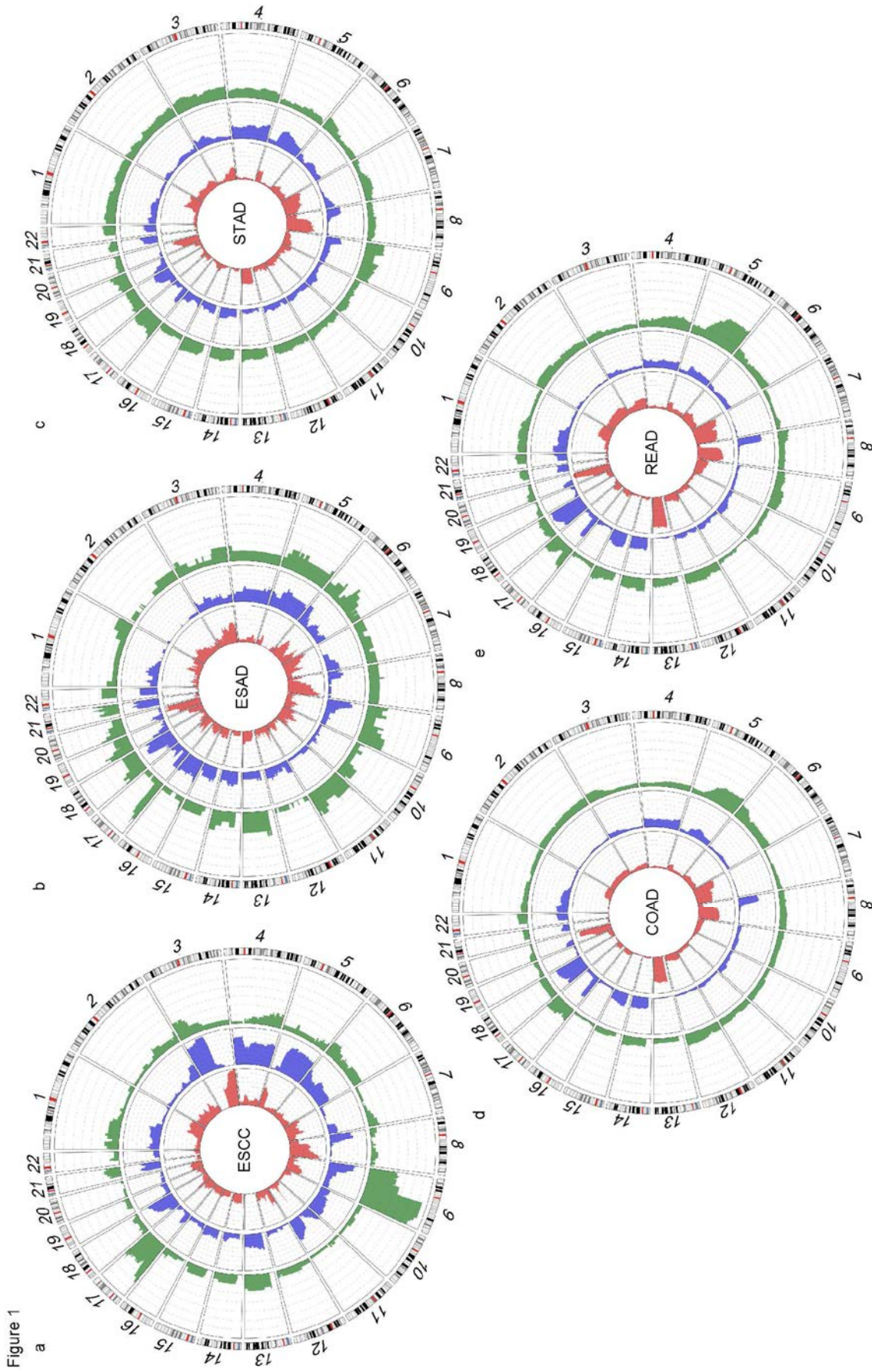


Figure 1

Figure 2

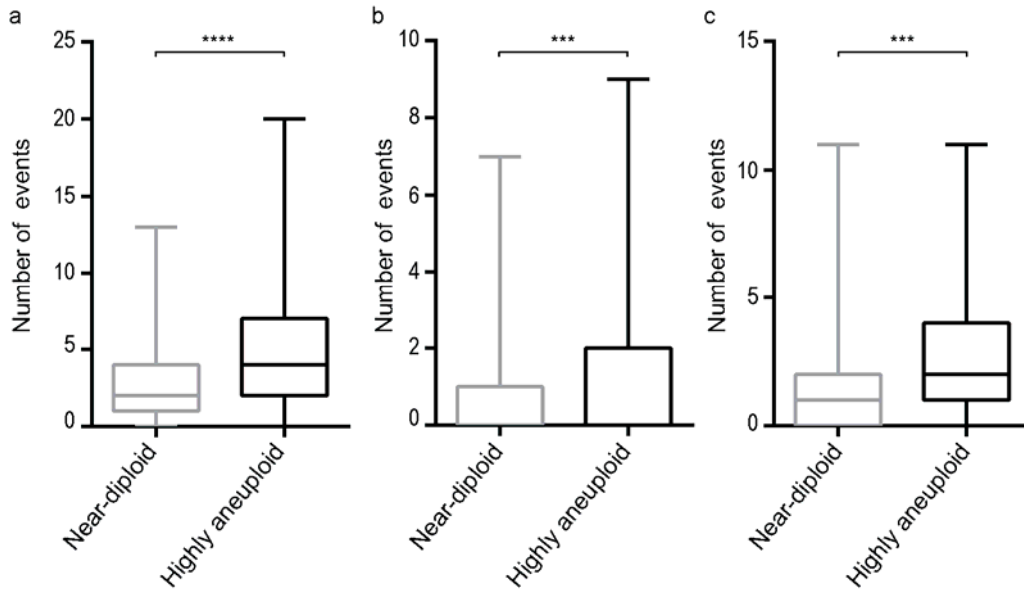
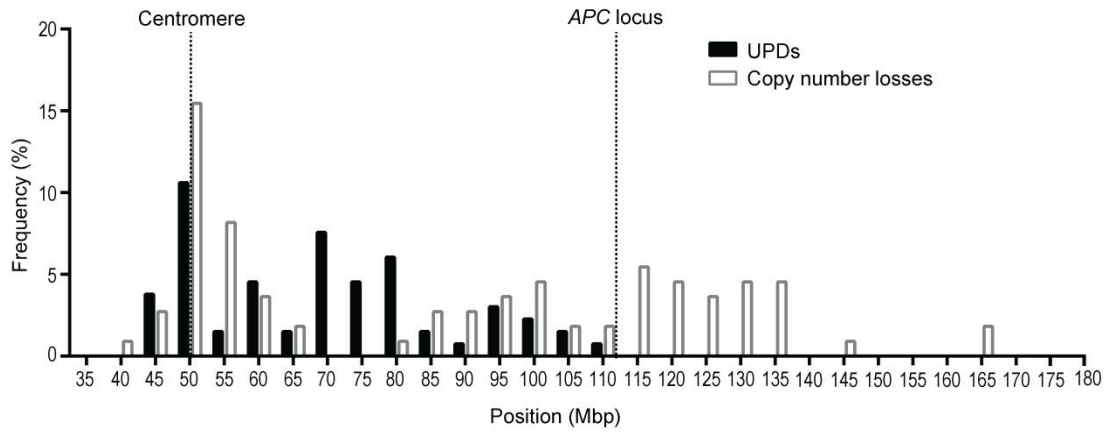
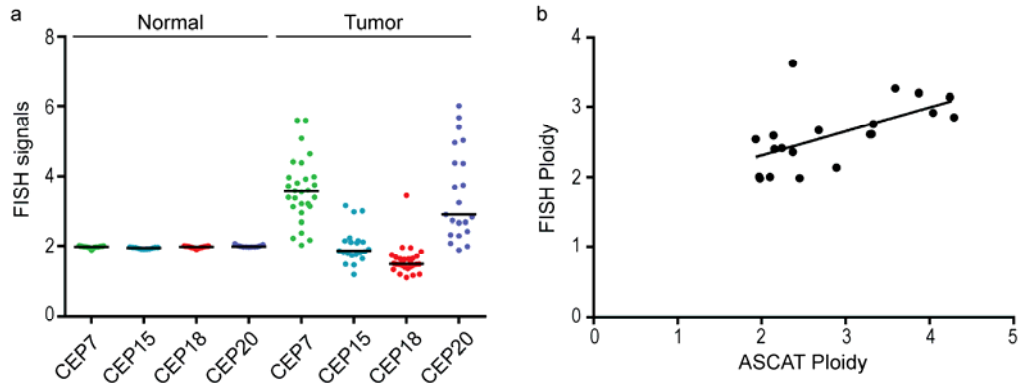


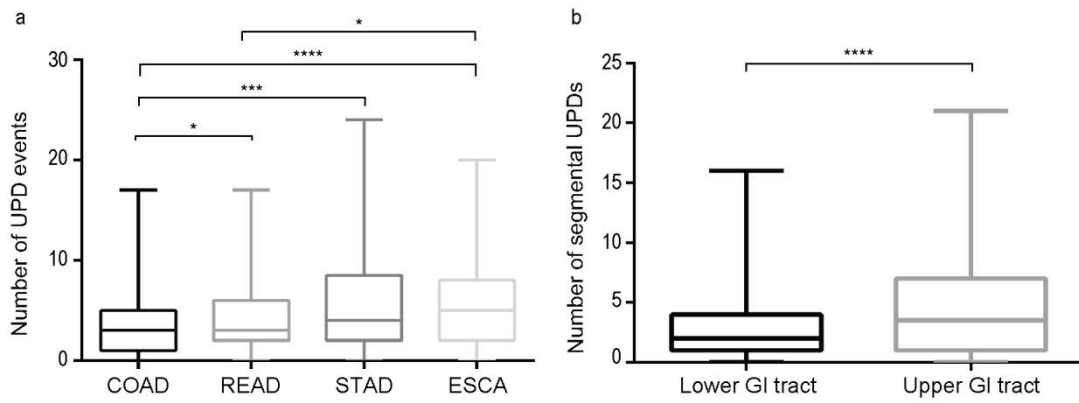
Figure 3



Supplementary Figure 2



Supplementary Figure 1



3.4 Capítol III

Keyvan Torabi, Darawalee Wangsa, Immaculada Ponsa, Markus Brown, Anna Bosch, Maria Vila-Casadesús, Tatiana S. Karpova, Maria Calvo, Antoni Castells, Rosa Miró, Thomas Ried and Jordi Camps.

Transcription-dependent radial distribution of TCF7L2 regulated genes in chromosome territories.

Chromosoma (2017). doi: 10.1007/s00412-017-0629-6.

Impact Factor (2015): 4.303



ORIGINAL ARTICLE

Transcription-dependent radial distribution of *TCF7L2* regulated genes in chromosome territories

Keyvan Torabi^{1,2} · Darawalee Wangsa³ · Immaculada Ponsa^{1,4} · Markus Brown³ · Anna Bosch⁵ · Maria Vila-Casadesús² · Tatiana S. Karpova⁶ · Maria Calvo⁵ · Antoni Castells² · Rosa Miró^{1,6} · Thomas Ried³ · Jordi Camps^{1,2}

Received: 11 November 2016 / Revised: 20 February 2017 / Accepted: 7 March 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract Human chromosomes occupy distinct territories in the interphase nucleus. Such chromosome territories (CTs) are positioned according to gene density. Gene-rich CTs are generally located in the center of the nucleus, while gene-poor CTs are positioned more towards the nuclear periphery. However, the association between gene expression levels and the radial positioning of genes within the CT is still under debate. In the present study, we performed three-dimensional fluorescence in situ hybridization experiments in the colorectal cancer cell lines DLD-1 and LoVo using whole chromosome painting probes for chromosomes 8 and 11 and BAC clones targeting four genes with different expression levels assessed by gene expression arrays and RT-PCR. Our results confirmed that the two over-expressed genes, *MYC* on chromosome 8 and *CCND1* on chromosome 11, are located significantly further away from the center of the CT compared to under-expressed genes on the same chromosomes, i.e., *DLC1*

and *SCN3B*. When *CCND1* expression was reduced after silencing the major transcription factor of the WNT/ β -catenin signaling pathway, *TCF7L2*, the gene was repositioned and mostly detected in the interior of the CT. Thus, we suggest a non-random distribution in which over-expressed genes are located more towards the periphery of the respective CTs.

Keywords 3D-FISH · Chromosome territory · Radial positioning · Gene expression · Nuclear architecture · Colorectal cancer

Introduction

Chromosomes are organized in distinct territories in the interphase nucleus, also known as chromosome territories (CTs) (Cremer and Cremer 2010). Based on three-dimensional

Keyvan Torabi, Darawalee Wangsa, and Immaculada Ponsa are co-first authors.

Electronic supplementary material The online version of this article (doi:10.1007/s00412-017-0629-6) contains supplementary material, which is available to authorized users.

✉ Thomas Ried
riedt@mail.nih.gov

✉ Jordi Camps
jcamps@clinic.cat

¹ Unitat de Biologia Cel·lular i Genètica Mèdica, Departament de Biologia Cel·lular, Fisiologia i Immunologia, Facultat de Medicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain

² Gastrointestinal and Pancreatic Oncology Group, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Institut d'Investigacions Biomèdiques

August Pi i Sunyer (IDIBAPS), Rosselló 153, 4th floor, 08036 Barcelona, Catalonia, Spain

³ Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-8010, USA

⁴ Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain

⁵ Advanced Optical Microscopy Unit, Scientific and Technologic Centers, Facultat de Medicina, Universitat de Barcelona, 08036 Barcelona, Catalonia, Spain

⁶ CCR/LRBGE Optical Microscopy Core, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

fluorescence in situ hybridization (3D-FISH) of chromosome-specific probes in interphase human lymphocyte nuclei and confocal laser scanning microscopy, the existence of a gene density radial distribution is now generally accepted. Accordingly, gene-rich chromosomes are positioned in the center of the nucleus compared to gene-poor chromosomes, which are located closer to the nuclear periphery (Croft et al. 1999; Cremer et al. 2001; Boyle et al. 2001; Cremer et al. 2003). In contrast, in flat human fibroblast nuclei, a radial distribution according to the chromosome size has been reported (Sun et al. 2000; Bolzer et al. 2005). In different mouse cell types, non-random correlations of chromosomes with gene density as well as chromosome size have been described (Mayer et al. 2005). Moreover, the radial distribution of CTs is mostly conserved in primates and in cells carrying rearranged chromosomes and is maintained in cells with artificially introduced chromosomes and in human-mouse hybrid cell lines (Tanabe et al. 2002; Sengupta et al. 2007; Meaburn et al. 2008; Sengupta et al. 2008). Other chromosomal features such as regional gene-dense chromatin segments yielding R-bands and gene-poor regions represented by G-bands have also been considered decisive parameters determining the radial positioning of chromatin in the nucleus (Küpper et al. 2007; Gindin et al. 2014). Furthermore, gene density also correlates with chromatin structure; whereas gene-rich chromosomes display less condensation and presence of open chromatin, gene-poor chromosomes show high levels of condensation and less accessible chromatin (Croft et al. 1999; Gilbert et al. 2004).

A relationship between gene positioning and function in individual CTs, as defined by transcriptional activity, has been established. Several studies provided evidence that chromatin regions containing highly over-expressed genes (e.g., major histocompatibility complex (MHC) locus, epidermal differentiation complex (EDC) cluster, and the 11p15.5 region) either looped out from the CT (Volpi et al. 2000; Williams et al. 2002; Mahy et al. 2002) or recurrently positioned in the periphery of the CT (Dietzel et al. 1999). However, so far, long-range chromatin loops have not been described as a common conformation in interphase nuclei (Küpper et al. 2007). Recently developed biochemical approaches, such as chromosome conformation capture and its technical derivations, provided evidence that chromosomes form compartments composed of topological associated domains (TADs), which are chromatin domains in the order of 100 kbp to 1 Mbp, potentially related to the regulation of gene expression (Dekker and Heard 2015). In agreement with this, the presence of an interchromatin compartment that expands between ~1 Mbp chromatin domain clusters would provide access of the transcriptionally competent chromatin to transcription factories and might represent a model to explain the functional organization of higher order chromatin structures (Popken et al. 2014). Several authors have claimed that the high-

throughput detection of actively transcribed chromatin contacts in *trans* could be due to the gene density radial distribution, as chromosomes encompassing large numbers of active genes are prone to be located in the center of the nucleus, thus in closer contact to each other than gene-poor chromosomes (Kalhor et al. 2012; Bickmore and van Steensel 2013). Taken together, the relationship between structure and function of genes within their CTs and their transcriptional activity still remains elusive.

A major transcription factor in colorectal cancer is the effector of the WNT/ β -catenin signaling pathway, *TCF7L2* (sometimes referred to as TCF4) (Mosimann et al. 2009). Upon nuclear internalization, β -catenin functions as a transcriptional co-activator, inducing the expression of several genes that have T cell factor/lymphoid enhancer family (TCF/LEF) DNA-binding sites in their regulatory regions. The collection of genes regulated by β -catenin/TCF includes several proto-oncogenes involved in the cell cycle progression such as *MYC* and *CCND1*, metalloproteinases such as *MMP7*, members of the fibroblast growth factor family, and other genes involved in the development of colorectal cancer (Hatzis et al. 2008).

In the present study, we aimed at assessing whether the position of over-expressed genes is different from genes expressed at lower levels with respect to the CTs in which they reside. Towards this goal, we performed 3D-FISH in two karyotypically stable colorectal cancer cell lines: DLD-1 and LoVo. Cells were synchronized at G1- and S-phases, and chromosome distribution and gene topology were analyzed. BAC clones covering four genes with profoundly different transcription levels were used to evaluate the distance from each gene to the geometrical center of the corresponding CT. Our data revealed that over-expressed genes are located more towards the periphery of their respective CTs, while under-expressed genes reside in the interior of the territories. Most importantly, we could show that genes, when silenced through RNA interference, moved towards the interior compartment of the CT.

Materials and methods

Cell culture and cell cycle synchronization

DLD-1 and LoVo colorectal adenocarcinoma cell lines were obtained from American Type Culture Collection and grown at 37 °C and 5% of CO₂ in RPMI-1640 and Ham's/F12 media, respectively, supplemented with 1% L-glutamine, penicillin, and 10% fetal bovine serum (Life Technologies). Both cell lines are diploid and with stable karyotypes (<http://www.ncbi.nlm.nih.gov/sky/>). For synchronizing cells in S-phase, double thymidine block was performed. Briefly, 70,000 DLD-1 cells were plated for at least 4 h and then blocked with

thymidine (2 mM) for 18 h. After three washes with PBS, cells were grown again with fresh media for 9 h. Then, a second block with thymidine (2 mM) for 17 h was performed. For LoVo, both thymidine blocks were for 12 h at the same concentration after an initial incubation of 48 h of 200,000 cells. After the second thymidine block, cells were released for 4 h in normal media to let the cells progress into S-phase. In unsynchronized cultures, the majority of cells were in G1; hence, synchronization was not required. Fluorescence-activated cell sorting (FACS) was used to assess cell percentage for each cell cycle phase after methanol fixation and propidium iodide (50 µg/mL) staining. Moreover, those cells with doublet FISH signals in unsynchronized cell cultures, indicative of sister chromatid separation after S-phase, were not included in our analysis.

Small interfering RNA transfection

A small interfering RNA (siRNA) molecule was used against *TCF7L2* (5'-GGUCAACCAGUGUACCCAA-3') (Ambion) to generate transient silencing of this gene using Lipofectamine RNAiMAX (Life Technologies). The transfection was performed in a 6-well plate according to the suggested procedure from Life Technologies. For each well, 25 pmol of siRNA were complexed with 7.5 µL of RNAiMAX in 250 µL of Opti-MEM Reduced Serum Medium (ThermoFisher Scientific) for 5 min at room temperature. The solution was then added in its entirety to 2 mL of RPMI-1640 to each well. Plates were directly transferred to an incubator at 37 °C and 5% of CO₂. Total RNA was collected after 72 h following transfection. As a negative control, we used the AllStars Negative Control siRNA (Qiagen), which basically consists of a validated non-silencing siRNA with no homology to any known mammalian gene.

RNA extraction and qRT-PCR

RNA extraction was performed with the RNeasy Mini Kit (Qiagen). RNA quantity was assessed using NanoDrop 1000 spectrophotometer (ThermoFisher Scientific). Briefly, 800 µL of total RNA were reverse transcribed to complementary DNA (cDNA) with the Verso cDNA Synthesis Kit (ThermoFisher Scientific). Then, quantitative reverse transcription PCR (qRT-PCR) using Power SYBR Green (Applied Biosystems) was performed as previously described (Camps et al. 2013), using specific primers for each of the four genes studied and the housekeeping genes *HPRT1* and *YWHAZ* (Eurofins Genomics). The $2^{-\Delta\Delta CT}$ algorithm was applied to estimate the relative gene expression changes (Livak and Schmittgen 2001). All the primers used in this study are shown in Table S1.

3D-FISH

Cells were plated onto 20-mm² sterile glass coverslips in 6-well plates and incubated overnight until 60–70% confluency was reached. When indicated, cells were treated with thymidine and washed out for synchronization at S-phase. First, cells were washed in 1× PBS and fixed in 4% paraformaldehyde (PFA) for 10 min at room temperature. Two minutes before finalizing the PFA treatment, two to three drops of 0.5% Triton X-100 were added. Then, PFA was washed out in 0.05% Triton X-100 for 5 min, and subsequently, cells were permeabilized in 0.5% Triton X-100 for 20 min, followed by an overnight incubation in 20% glycerol at room temperature. Next, three cycles of freezing and thawing in liquid nitrogen were performed, while the coverslips were incubated in 20% glycerol in between cycles. Afterwards, the cells were incubated in HCl for 10–20 min, washed in 2× saline sodium citrate (SSC), and incubated in RNase A (200 µg/mL) for 1 h at 37 °C. Finally, cells were incubated in 50% formamide/2× SSC at 4 °C for at least 1 week before hybridizing.

Whole chromosome painting probes (WCPs) for chromosome 8 and 11 were prepared from flow-sorted chromosomes and labeled with DY-505-dUTP (Dyomics). Four gene-specific probes were generated using contigs of several bacterial artificial chromosome (BAC) clones for the genes *MYC* (chr8:128,748,315–128,753,680) (RP11-1136L8, CTD-3056O22, RP11-55J15) and *DLCL1* (chr8:12,940,872–13,134,057) (RP11-92C1, RP11-722B21, RP11-684F19), both genes mapping 115 Mbp apart on chromosome 8, and *CCND1* (chr11:69,455,873–69,469,242) (CTD-2507F7, RP11-30016, RP11-186D19) and *SCN3B* (chr11:123,499,895–123,525,315) (RP11-793D12, RP11-764P17), both genes mapping 54 Mbp apart on chromosome 11. *Escherichia coli* containing the BAC clones were grown following standard procedures and with the selective antibiotic recommended by the BACPAC Resources Center. DNA was extracted using the Qiagen Maxi Kit (Qiagen) and quantified with NanoDrop. Each contig was differentially labeled by the nick translation enzymatic reaction using Spectrum Orange-dUTP (Abbott Molecular) for the genes *DLCL1* and *SCN3B* and DY-651 dUTP (Dyomics) for the genes *MYC* and *CCND1*. Cot-1 DNA and salmon sperm DNA were added to the resulting enzymatic reaction in order to suppress repetitive sequences. As a result, CTs might not be totally represented when using a whole chromosome painting probe. Probes were combined in two panels (panel 1: WCP8, *MYC*, and *DLCL1*; panel 2: WCP11, *CCND1*, and *SCN3B*) and hybridized to pretreated samples. Coverslips and probes were co-denatured at 72 °C for 5 min in a ThermoBrite StatSpin system (Abbott Molecular) and hybridized at 37 °C for 48 h in a humidified chamber. After hybridization, coverslips were washed with 2× SSC/NP-40 0.1% at 45 °C for 2 min and twice with 2× SSC at room temperature for 3 min. ProLong

Gold Antifade Mountant with 4',6-diamidino-2-phenylindole (DAPI) (ThermoFisher Scientific) was applied to the coverslips which were mounted on standard glass microscope slides.

Microscope imaging

Images were acquired using a DeltaVision Elite High Resolution Microscope (GE Healthcare Life Sciences), with a 60×/1.42 Oil Plan APO and a 100×/1.4 Oil Super-Plan APO oil immersion objectives and processed with the softWoRx software (v6) (GE Healthcare Life Sciences) at 896 × 896 size and 16 bits. The voxel size was 0.1076 × 0.1076 × 0.2 μm for ×60 images and 0.0648 × 0.0648 × 0.2 μm for ×100 images. Image deconvolution was performed using the Resolve3D software integrated with softWoRx.

Three-dimensional distance measurements

3D-FISH measurements of at least 25 individual nuclei per condition were performed using customized macros of Fiji-ImageJ software package, based on 3D Objects Counter and 3D ImageJ Suite plugins (Bolte and Cordelières 2006; Ollion et al. 2013). First, individual nuclei were cropped from a merged image stack. Next, channels were split, and different algorithms were used for each channel to calculate thresholds of each individual object (i.e., Li for nucleus, Max entropy for CTs, and Intermode for gene locus signals). In order to adjust the thresholding for CT segmentation, we utilized the original RGB image as a template to ensure that background was sufficiently removed and that the created object reconstructed the entire fluorescence signal. Threshold values were typically found to vary between 10 and 15 units intraexperiment. Once individual 3D objects were generated, they were imported to the 3D Roi Manager plugin (Fiji-ImageJ), and the following measurements were calculated. First, based on the number of voxels contained in each object, volumes were assigned to each nucleus and CT. In order to evaluate radial distances of CTs and gene loci within the nucleus, eccentricity values were obtained. Eccentricity is a measurement provided by the Roi Manager plugin and corresponds to the ratio between the distance from the geometric center of the nucleus (N_c) to the center of the CT (CT_c) or gene locus (G_c) and the nuclear radius (N_r). N_r is defined by the distance between the N_c and the CT_c or G_c plus the extended distance to the nuclear periphery. Distances are presented as a scale where 0 corresponds to the N_c and 1 the nuclear periphery. Next, radial distances from the geometric center of the CT (CT_c) to the center of the gene locus (G_c) were also measured by calculating eccentricity values. In this case, CT radius (CT_r) is defined by the distance between the CT_c and the G_c plus the extended distance to the CT periphery. Similarly, radial distances are presented as a scale where 0 corresponds the CT_c and 1 the

periphery. Of note, as CT shapes are highly irregular, distances above 1 can be observed, especially in those instances where the gene loci loop out from the CT.

Finally, absolute distances from the G_c to the nearest border of the corresponding CT were also scored. In addition, measurements of signal co-localization as a surrogate of the positioning of the gene within its CT were calculated. In this regard, the software quantified the number of voxels corresponding to the gene locus, which overlapped with the CT. When the level of co-localization was lower than 50%, the distance between the gene locus and the border of the CT was considered to be outside of the CT. On the other hand, if the level of co-localization between the gene locus and the CT was higher than 50%, the distance between this gene locus and the CT was considered to be inside the territory, and this value was multiplied by -1.

Statistical analysis

The paired Wilcoxon test was used in order to compare gene radial distances within CTs and distances from the G_c to the closest border. Proportion test was applied when comparing gene loop proportions. Mann-Whitney sum-rank test was applied when comparing gene radial distances and distances from the G_c to the closest border between siRNA, siNEG, and untreated cultures. Kruskal-Wallis test was used in order to compare gene radial distances in a gene density manner. p values and graphs were calculated and designed using the software GraphPad Prism 6.0 (GraphPad) and R version 3.2.2. All distributions were computed using the kernel density estimator method.

Results

Gene-rich chromosomes occupy a more internal position in the nucleus. Yet, the spatial positioning of specific genes in their respective chromosome territories during interphase is still a matter of debate, and so is the question whether transcriptional activity determines their localization. In the present study, three-color 3D-FISH experiments were performed in DLD-1 and LoVo cells to assess the localization of transcriptionally active and inactive genes within chromosome territories CT8 and CT11 in different phases of the cell cycle. Cells were synchronized at G1- and S-phases in order to assess higher order of chromatin arrangements through the cell cycle. Representative maximum intensity projections of a three-dimensional image stack nuclei are shown in Fig. 1a, b. Furthermore, a schematic illustration of a three-dimensional reconstruction for a nucleus and all distance measurements analyzed in this study is depicted in Fig. 1c–e. Our results provide insights into the structure-function relationship of genomic loci in the interphase nucleus.

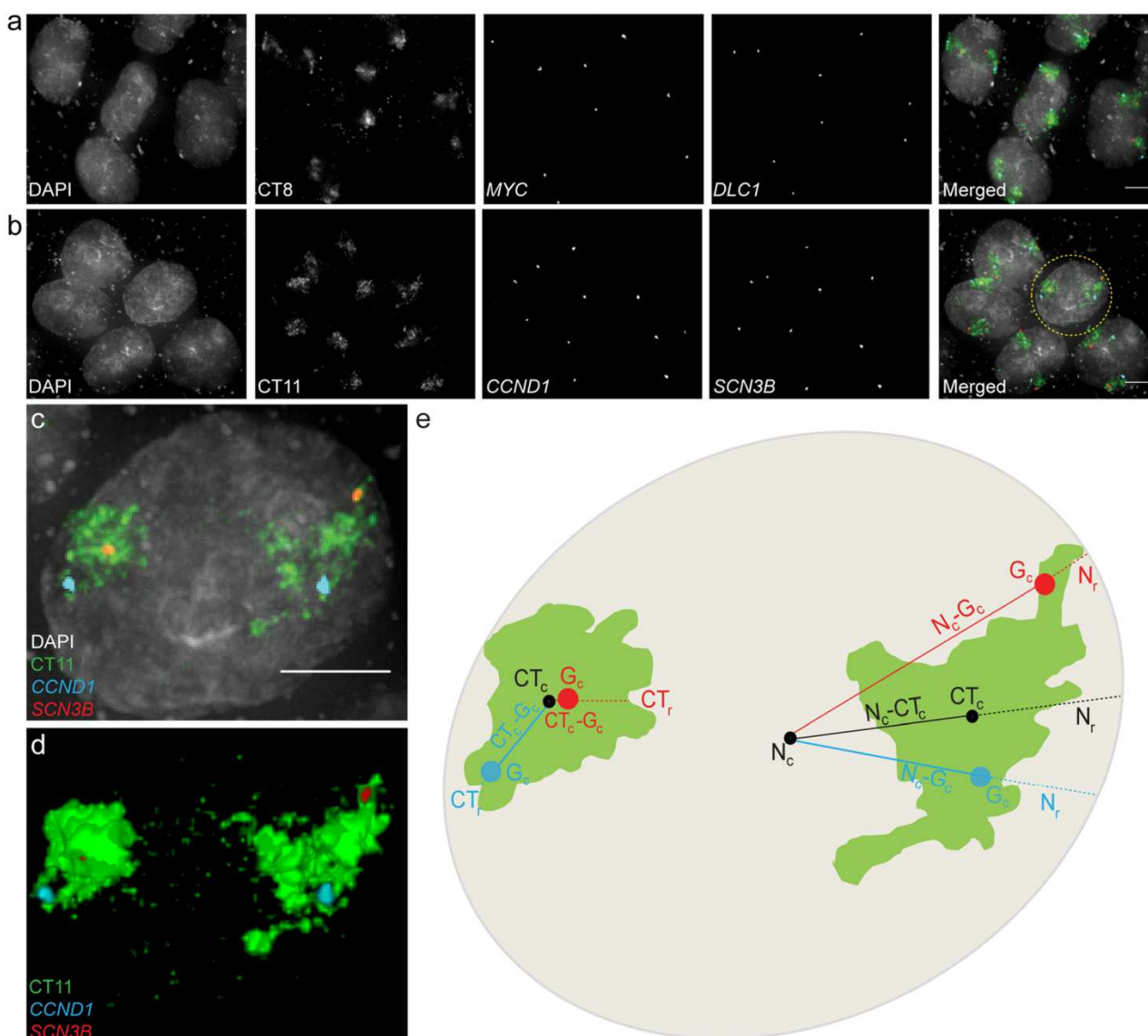


Fig. 1 Image representations of 3D-FISH analysis. **a** A representative maximum projection of a pseudoconfocal image stack for each fluorochrome is depicted for DLD-1 nuclei in S-phase showing **a** WCP8, *MYC*, and *DLC1* and **b** WCP11, *CCND1*, and *SCN3B*. DAPI was used as DNA counterstain. **c** Enlarged picture of a selected nucleus. **d** Three-dimensional reconstruction of the CTs and the corresponding FISH signals for gene probes. **e** A schematic representation illustrating all measurements assessed. Nuclear radial

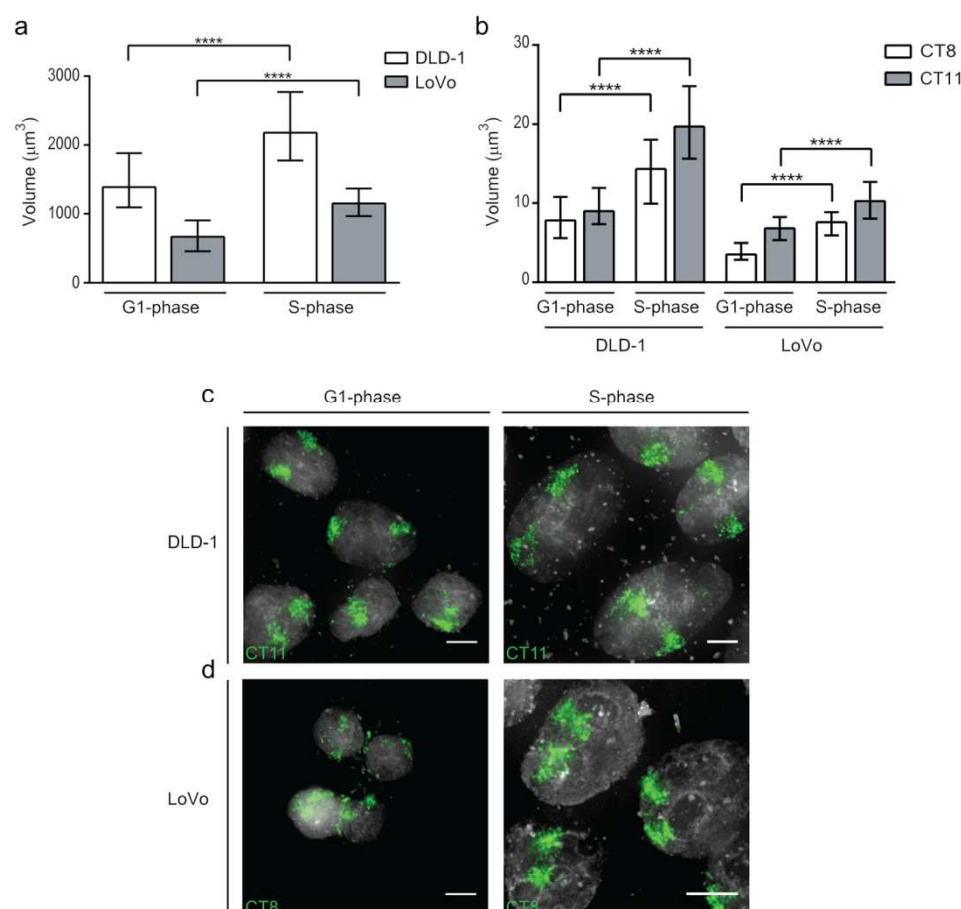
position of chromosome territory (N_c-CT_c : center of the nucleus (N_c), center of the chromosome territory (CT_c); nuclear radial position of over-expressed or under-expressed gene (N_c-G_c : G_c colored in blue represents over-expressed gene, G_c colored in red represents under-expressed gene); over-expressed or under-expressed gene radial position with respect to the center of the territory (CT_c-G_c). Both types of measurements were assessed in the two homologous CTs. Scale bars correspond to 5 μm

Nuclear and CT volumes through different stages of the cell cycle

First, we sought to investigate whether the size of interphase nuclei and CTs differed between G1- and synchronized S-phase cells. An enrichment of at least 65% of cells in either G1- or S-phase was achieved in both cell lines. G1- phase cells showed a median nuclear volume of 1388 and 665 μm^3 , in DLD-1 and LoVo, respectively, while cells in S- phase showed a median

nuclear volume of 2180 and 1151 μm^3 (Fig. 2a). In order to evaluate if this increase in nuclear size was influenced by changes in the chromatin as a consequence of possible genome doubling, we assessed the volume of specific CTs. Our results showed a statistically significant increase in the size of CTs in S-phase compared to G1- phase cells. Thus, while DLD-1 cells in G1 showed a median volume of 7.81 and 8.99 μm^3 for CT8 and CT11, respectively, DLD-1 S- phase cells showed a median of 14.34 and 19.72 μm^3 (Fig. 2b, c).

Fig. 2 Measurements of nuclear and CT volumes. **a** Bar plot showing the volume measurements of G1- and S-phase DLD-1 and LoVo nuclei. **b** Shown in a bar plot are the volume measurements of CTs in G1- and S- phase DLD-1 and LoVo cells. A representative maximum projection of a pseudoconfocal image stack of G1- and S- phase in **c** DLD-1 and **d** LoVo nuclei. Scale bars correspond to 5 μm . Mann-Whitney sum-rank test was used to compare nuclei and CT volumes. **** $p < 0.0001$



Similarly, LoVo cells in G1 showed a median volume of 3.53 and 6.81 μm^3 for CT8 and CT11, and S- phase cells showed median volumes of 7.6 and 10.27 μm^3 for CT8 and CT11, respectively (Fig. 2b, d). Moreover, when comparing CTs, our analysis showed that chromosome 11, although it is smaller in size compared to chromosome 8 (134 vs 145 Mbp, respectively), displayed a significantly greater volume (19.72 and 10.27 μm^3 for chromosome 11 and 14.34 and 7.60 μm^3 for chromosome 8 in S- phase DLD-1 and LoVo cells, respectively; $p < 0.0001$) (Fig. 2b). In addition, further analysis of the nuclear radial positioning showed that CT8 remained invariably nearby to the nuclear periphery (range from 0.67 to 0.71 in DLD-1 and LoVo cells) through different stages of the cell cycle, while CT11 in S- phase cells was preferentially positioned more interior compared with G1- phase (from 0.71 in G1- to 0.67 in S- phase DLD-1 cells, $p = 0.069$ and from 0.64 in G1- to 0.55 in S- phase LoVo cells, $p < 0.001$) (Fig. S1a, b).

Nuclear radial gene distribution is independent of expression levels

In order to assess the nuclear positioning of genes that are either transcribed or not in these cell lines, we chose four differentially

expressed genes based on microarray data previously generated in our laboratory (Camps et al. 2009). Gene expression levels of the genes *MYC*, *CCND1*, *DLC1*, and *SCN3B* were assessed by semi-quantitative RT-PCR in both cell lines and different cell cycle phases. Our results showed a consistent over-expression of *MYC* and *CCND1*, while *DLC1* and *SCN3B* were under-expressed normalized to the housekeeping gene *HPRT1*. These differences were highly evident in G1- phase cells (Fig. S2).

First, we measured the distance from each locus-specific gene probe set to the geometrical center of the nucleus. Our results suggested that the nuclear radial position of differentially expressed genes was statistically independent of the gene expression levels, both in DLD-1 and LoVo cells. Specifically, our data indicated that the highly expressed gene *MYC* was positioned towards the nuclear periphery to a similar level than the under-expressed genes *DLC1* and *SCN3B* (Fig. 3). In addition, this lack of correlation between nuclear location and transcriptional activity was independent of the cell cycle stage. Next, we assessed whether other factors such as local gene density were determining gene-specific nuclear arrangements. To this end, we examined the local gene density of a 10-Mbp genomic window covering our genes of interest and define whether they were located in a gene-dense R-band

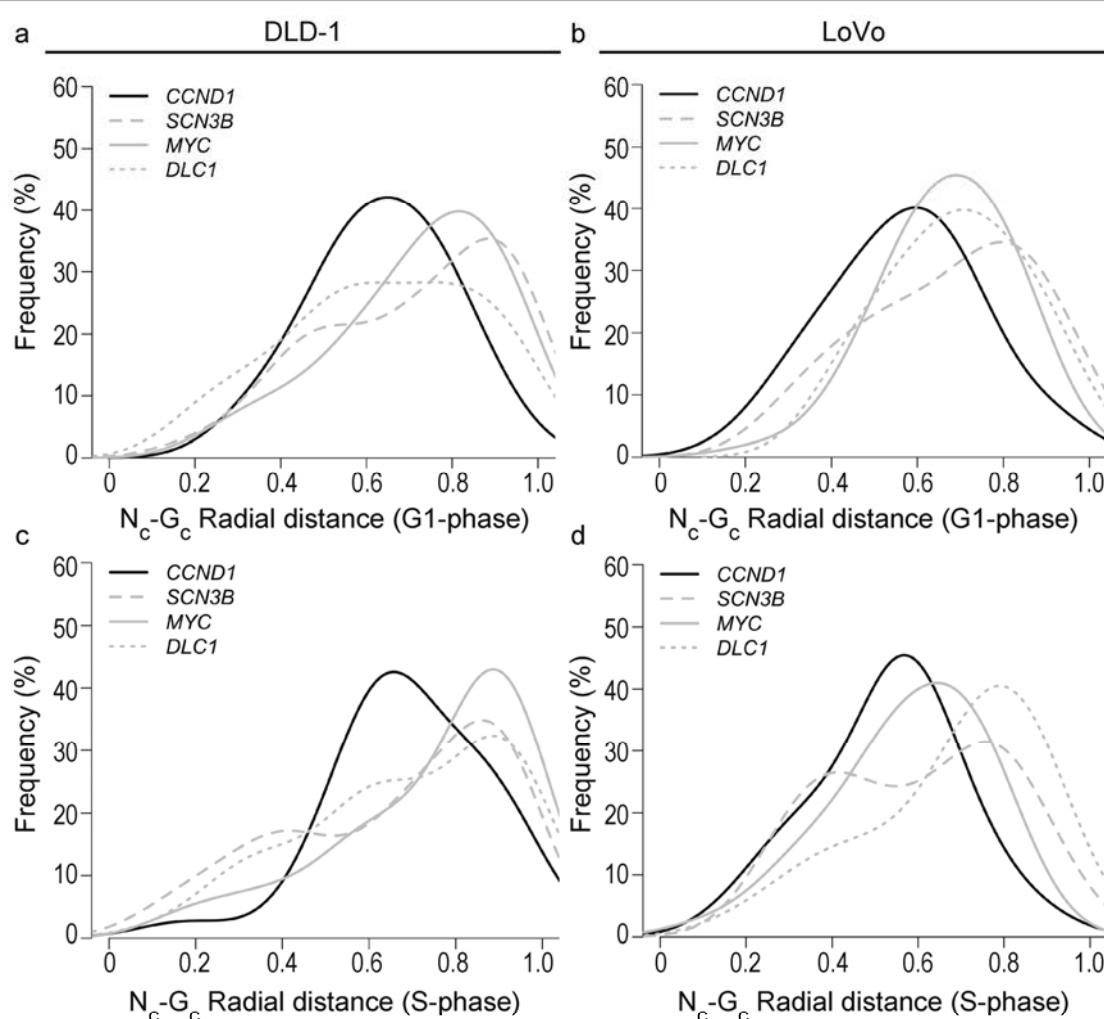


Fig. 3 Gene density-based nuclear radial positioning. Distances between the genes of interest and the center of the nucleus (N_c - G_c) in **a** G1-phase DLD-1 cells, **b** G1-phase LoVo cells, **c** S-phase DLD-1 cells, and **d** S-phase LoVo cells. *Black line* indicates genes within R-bands, and *gray*

lines correspond to genes contained in G-bands. The Kruskal-Wallis test was used in order to determine whether *CCND1* median values were different compared to *MYC*, *DLC1*, and *SCN3B* medians ($n > 50$)

or gene-poor G-band. Specifically, *MYC*, *DLC1*, and *SCN3B*, which all are located in chromatin segments assigned to G-bands (12.9, 18.4, and 26.1 genes/Mbp, respectively), were positioned closer to the nuclear periphery compared to *CCND1*, which is located in a gene-dense R-band (37.1 genes/Mbp) and towards to the nuclear center. Although this difference in nuclear positioning was statistically significant in the G1-phase of DLD-1 and LoVo cells ($p < 0.01$) (Fig. 3a, b), it was less evident in S-phase cells (Fig. 3c, d).

Positioning of genes within the CT is associated with transcription activity

We then measured the position of individual genes with respect to their corresponding CT (CT_c - G_c , i.e., eccentricity values) to assess whether the topological

distribution of genes is affected by their transcriptional activity. Our results indicated that the two highly expressed genes (*MYC* and *CCND1*) were radially positioned further away from the center of their corresponding CT compared to the under-expressed genes (*DLC1* and *SCN3B*) in both G1- and S-phase DLD-1 (Fig. 4a, b) and LoVo cells (Fig. S3a, b). Specifically, while the median radial measurements of the under-expressed genes in DLD-1 and LoVo were 0.816 and 1.005 in G1- and in S-phase cells, respectively, the over-expressed genes were located 1.32 and 1.27-fold more external of the CT ($p < 0.0001$) (Table S2). Next, we also analyzed the co-localization between individual gene probes and their corresponding CT, assuming that the higher the co-localization, the more internal the gene was positioned. In contrast, when the

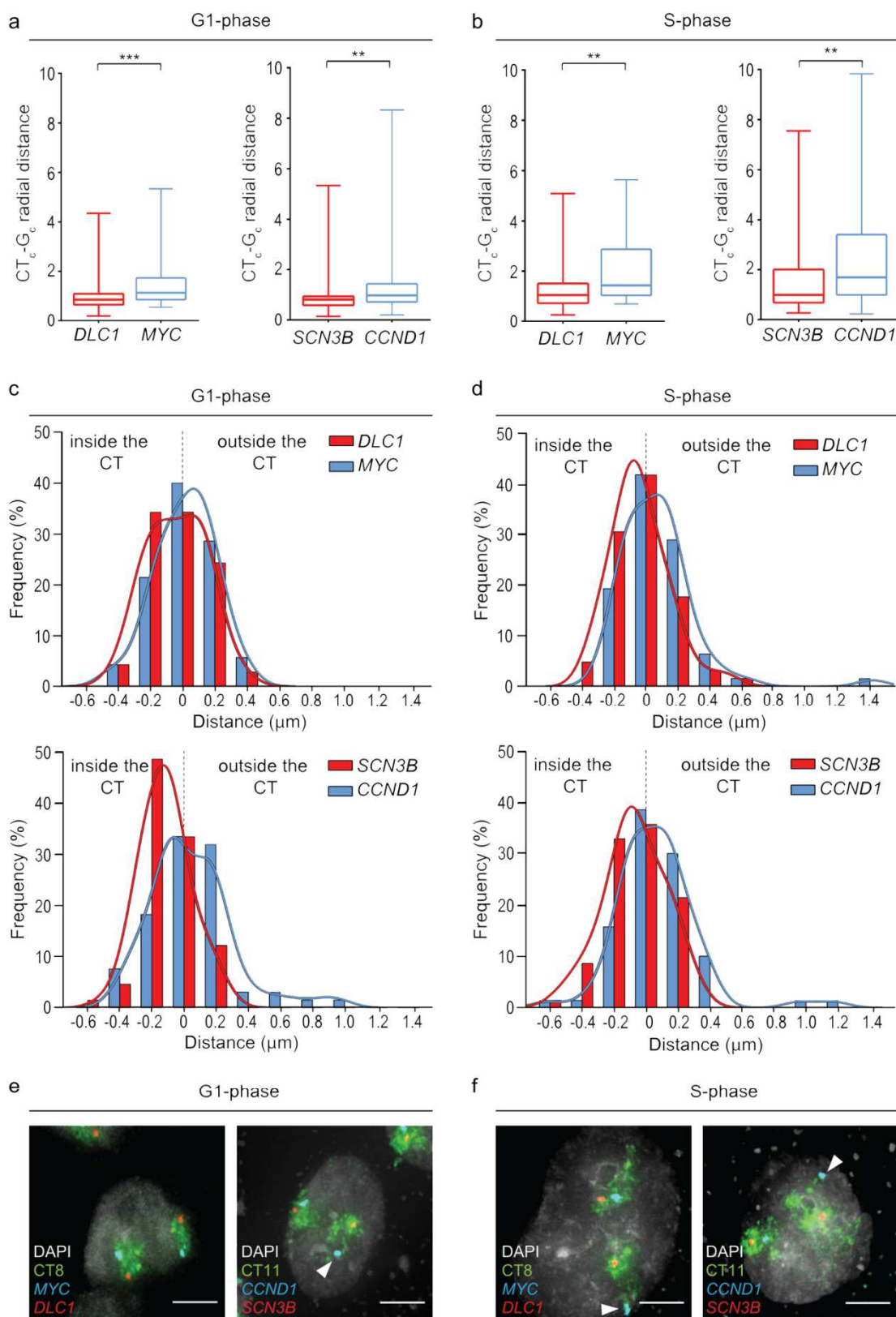


Fig. 4 Positioning of differentially expressed genes within their corresponding CTs. Box plots showing radial distances (i.e., eccentricity values) of *MYC* vs *DLCL1* and *CCND1* vs *SCN3B* in a G1- and b S- phase DLD-1 cells. Histograms and distributions of the distances between the center of the gene and the closest CT border depicted in bins of 0.2 μm are plotted for c G1- and d S- phase DLD-1 cells. Dashed line at 0 indicates the border of the CT; thus, positive values indicate that co-localization of the gene signal and the CT painting probe was lower than 50%. In contrast, negative values indicate that co-localization of the gene signal and the CT painting probe was higher than 50%. Adjusted distributions computed using kernel density estimator method have been added to the plots. Indicated by arrowheads in the representative examples are loci looping out of the corresponding CT in e G1- and f S- phase DLD-1 cells. Scale bars correspond to 5 μm . The Wilcoxon test was used to compare median radial measurements between over-expressed and under-expressed genes ($n > 50$). * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$

level of co-localization was low, the gene was positioned in a more peripheral region of the CT and thus more exposed to the interchromatin compartment. By measuring the overlapping volume of FISH signals for the gene probes and their corresponding CTs, we could demonstrate that *CCND1* and *MYC* co-localization values were significantly lower compared to *SCN3B* and *DLCL1* in both G1- and S- phase DLD-1 cells (Fig. S4a, b). By combining the co-localization analysis and measurements between G_c and the nearest border of the corresponding CT, we further evaluated the position of the over-expressed genes *MYC* and *CCND1* compared with their under-expressed counterparts *DLCL1* and *SCN3B* in both cell cycle stages in DLD-1 (Fig. 4c, d) and LoVo (Fig. S3c, d). Our data showed that the under-expressed genes exhibit a median distance of $-0.070 \mu\text{m}$, indicating a more interior positioning within the CT, whereas the over-expressed genes displayed a median distance of $0.047 \mu\text{m}$, indicative of a more external positioning irrespective of the cell

cycle phase (Table 1). In addition, since FISH is a single-cell level experiment, we assessed in how many CTs the over-expressed gene is located more externally than the under-expressed (Table 1). These results confirmed that *MYC* and *CCND1* were positioned more externally in 60.5 and 64.7% of the CTs analyzed, respectively.

Finally, we then assessed patterns that would resemble gene looping out from their corresponding CT by considering the loci that showed a radial distance above 3, given the fact that 1 is the border of the CT, and were located at least $0.3 \mu\text{m}$ outside the CT. Based on these criteria, we could identify patterns of chromatin suggesting looping out preferentially affecting the over-expressed genes (4.35% for *MYC*, 4.85% for *CCND1*, 0.72% for *DLCL1*, and 0% for *SCN3B*; $p < 0.01$) in DLD-1 (Fig. 4e, f) and LoVo (Fig. S3e, f).

Topological reposition of over-expressed genes after silencing TCF7L2

In order to functionally determine whether such gene-specific chromosomal localization was indeed dependent on transcriptional levels, we took advantage of a constitutive activated signaling pathway in the majority of colorectal cancer cells, i.e., the WNT/ β -catenin pathway. Using siRNA-mediated gene silencing, we inhibited the expression of *TCF7L2* (siTCF7L2), which is the molecular effector of the WNT/ β -catenin pathway and acts as a modulator of the transcriptional activity of *MYC* and *CCND1*. To assess the gene silencing efficiency, we measured the levels of the remaining *TCF7L2* messenger RNA (mRNA) after 72 h following transfection and showed that silencing was above 70% in S- phase DLD-1 cells (Fig. 5a). We performed this analysis in

Table 1 Gene distances with respect to the border of CTs

Cell line	Chr.	Cell cycle phase	Over-expressed gene (μm) ^{a,b}	Under-expressed gene (μm) ^{a,b}	p val ^c	Gene positioning (%) ^d
DLD-1	CT8	G1	0.050	-0.052	0.0208	61.43
		S	0.071	-0.066	0.0059	59.68
	CT11	G1	-0.046	-0.109	0.0001	66.67
		S	0.080	-0.073	0.0025	62.86
LoVo	CT8	G1	0.050	-0.090	0.0042	63.33
		S	-0.046	-0.044	0.1811	53.57
	CT11	G1	0.043	-0.133	0.0001	85.14
		S	0.010	0.005	0.1340	58.62

^aDistances are indicated as the median of $n > 50$ territories

^bPositive values correspond to loci outside the CTs, while negative values correspond to loci inside the CTs

^cWilcoxon matched pairs signed rank test

^dPercentage of CTs with the over-expressed gene located more externally than the under-expressed gene

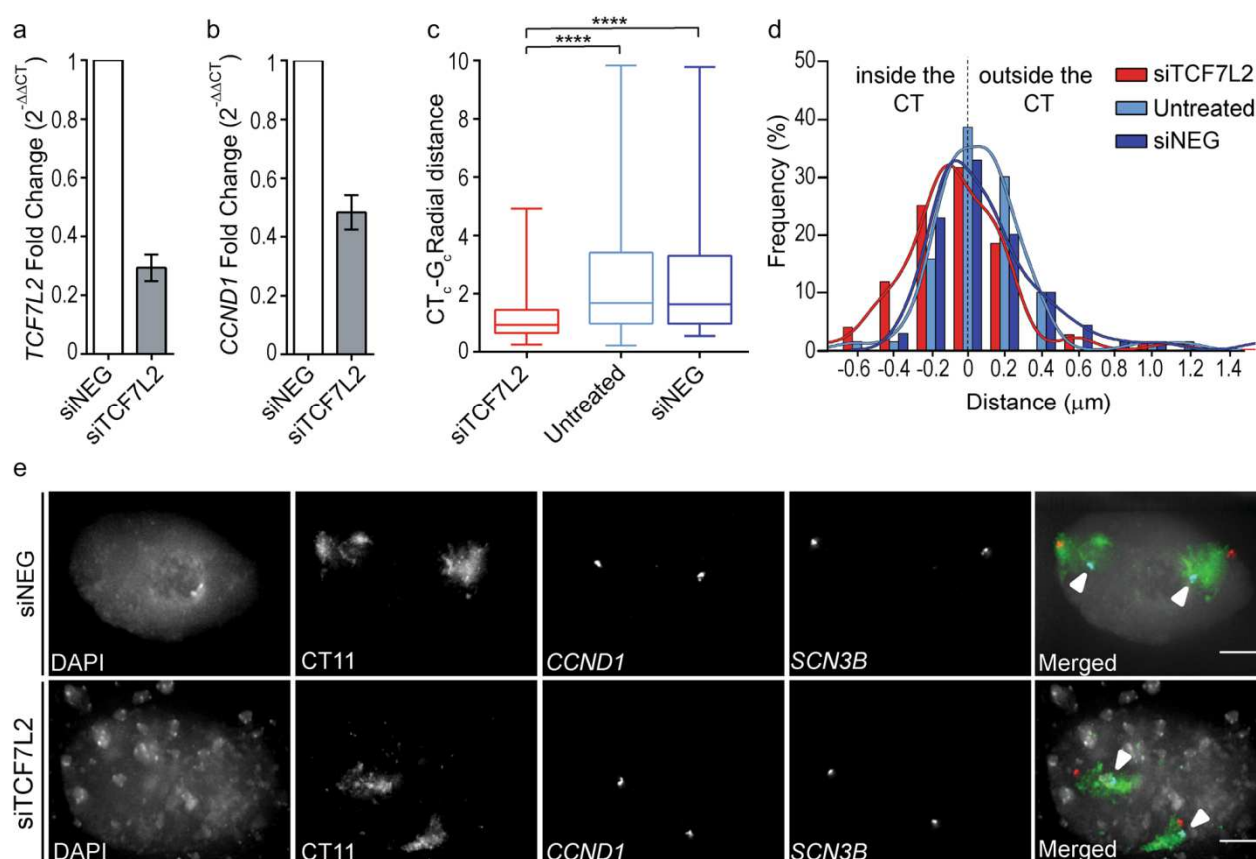


Fig. 5 Repositioning of *CCND1* upon silencing of *TCF7L2* in DLD-1. Bar plots showing the fold change expression levels of **a** *TCF7L2* and **b** *CCND1* in siRNA-treated cells. In both analyses, expression levels of *TCF7L2* were compared against the siNEG transfection. *YWAZ* was used as a normalization gene. **c** Box plot showing the significant differences of *CCND1* radial distances in siTCF7L2 compared to siNEG and untreated cells. **d** Histogram showing the distribution of the distance between *CCND1* and the closest border of CT11 depicted in bins of 0.2 μm of siTCF7L2, siNEG, and untreated cells. Adjusted distributions computed using kernel density estimator method have

been added to the plots. **e** Representative maximum projections of all channels of siNEG and siTCF7L2-treated cells showing the repositioning of *CCND1* from the periphery of the CT to the interior. Arrowheads in the merged image show how *CCND1* FISH signals are located in the periphery of CTs in the siNEG nuclei and in the interior of the CTs in the siTCF7L2-treated cells. Scale bars correspond to 5 μm. Mann-Whitney sum-rank test was used to compare *CCND1* median radial distances between different experiments (untreated, siNEG, and siRNA). **** $p < 0.0001$

synchronized S- phase cells in order to achieve the highest level of gene silencing. Thus, we interrogated the effect of silencing *TCF7L2* on the mRNA levels of *MYC* and *CCND1*. In contrast to previous reported data, we did not observe a downregulation of *MYC* upon silencing *TCF7L2* in any of the cell lines we analyzed. On the other hand, we did detect a 50% downregulation of *CCND1* in DLD-1 cells upon silencing of *TCF7L2* (Fig. 5b). Therefore, we utilized this model to examine whether the radial distribution of *CCND1* within CT11 was modified upon silencing of its transcriptional regulator *TCF7L2*.

Next, we measured the radial distance between the center of the CT11 and *CCND1* (i.e., eccentricity values) and the distance from *CCND1* to the nearest border of CT11 in siTCF7L2-treated DLD-1 cells compared to the mock transfection (i.e., siNEG). The localization of *SCN3B* was used as a

negative control in the radial positioning analysis of 3D FISH images. While the radial positioning of *CCND1* in the siNEG-transfected cells remained unchanged compared with untreated cells, the radial positioning of *CCND1* in siTCF7L2 exhibited a more central position (1.64 vs 0.93 in siNEG vs siTCF7L2-treated cells, respectively; $p < 0.0001$) (Fig. 5c). Specifically, the comparison between radial positioning of *CCND1* and *SCN3B* showed that upon silencing of *TCF7L2*, *CCND1* was more proximal to the center of the CT than *SCN3B* (0.93 and 1.16, respectively), thus inverting the original distribution observed in wild-type cells. Likewise, the distance measurements from *CCND1* to the border of the CT11 confirmed that in siTCF7L2-treated cells, *CCND1* was located more in the interior of CT11 compared to siNEG (from 0.059 to -0.076 μm in siNEG vs siTCF7L2-treated cells, respectively; $p < 0.01$) (Fig. 5d, e). Accordingly, as the expression of *MYC* was not reduced upon

silencing of *TCF7L2*, the radial positioning was conserved (1.71 vs 1.33 in siNEG vs siTCF7L2-treated cells, respectively) (Fig. S5a). However, the combination of co-localization and distance analysis with respect to the CT8 border did detect a slight relocation of *MYC* to the interior of the CT in siTCF7L2-treated cells (from 0.101 to 0.024 μm siNEG vs siTCF7L2-treated cells, respectively; $p = 0.051$) (Fig. S5b).

Altogether, these results suggest that the repositioning of *CCND1* in the CT is triggered by its transcriptional inactivation upon silencing *TCF7L2*. This provides experimental evidence in isogenic controlled models that transcriptional activity of specific genes is correlated, and perhaps dependent, on their position in the CT in which they reside.

Discussion

It is generally accepted that gene density is the main feature in organizing the distribution of chromosomes within the interphase nucleus (Cremer and Cremer 2001). In the present study, we performed 3D-FISH analysis in two different cell cycle stages to assess nuclear architecture of specific genes and CTs in a dynamic system. Our data revealed that S-phase cells displayed at least 1.5-fold increase in nuclear volume compared to G1-phase cells. Chromosome territories CT8 and CT11 scaled up accordingly. Despite similar chromosome lengths, CT11 volumes were significantly larger than CT8 in both cell lines. Such discrepancy in the volume of both equally sized chromosomes might be explained by differences in the gene density. This would be consistent with previous reports that indicate that gene-rich chromosomes show larger volumes, which potentially result from a less condensed chromatin state (Croft et al. 1999; Müller et al. 2010). Enrichment for hypersensitive DNase I sites is characteristic of gene-dense genomic segments and thus could be one of the causes of a more open chromatin conformation (Gilbert et al. 2004; Sehgal et al. 2014). However, limitations on the chromosome painting probe resolution and image processing need to be considered while performing 3D-FISH experiments (Fritz et al. 2016). In addition, the finding of very subtle differences in the radial position of CTs in G1- and S-phase cells indicates that the radial position of the CTs is highly invariable during interphase (Gerlich et al. 2003; Strickfaden et al. 2010; Müller et al. 2010). In fact, long-range repositioning of CTs has been shown during DNA damage repair induction, in nuclear lamina dysfunctions, in lymphocyte activation, and in cell differentiation (Cremer et al. 2006; Mehta et al. 2010; Mehta et al. 2013; Ioannou et al. 2015; Ranade et al. 2016; Sehgal et al. 2016b). Other studies focused on the three-dimensional topology of each CT reported that the uniqueness of each chromosome pair provides specific non-random intrachromosomal arrangements and structural organization, which can change

during cell cycle and might be related to global expression programs of individual chromosomes (Sehgal et al. 2016a).

Several studies assessed the correlation between levels of gene expression and the radial nuclear distribution. Some findings support a more nuclear internal location of expressed genes compared with the more peripheral location of silent genes (Zink et al. 2004; Williams et al. 2006), while such correlation was not found by other authors (Nielsen et al. 2002; Murmann et al. 2005; Meaburn and Misteli 2008). Recently, the use of chromosome conformation capture techniques has identified TADs as the main unit of chromatin spatial organization (Dixon et al. 2012). These TADs are defined by the number of sequenced ligation products, which reflects the amount of DNA interactions and thus chromatin proximity. In addition, TADs have been divided into those corresponding to enriched areas of active chromatin and those representing a silent state (Wang et al. 2016). In our study, despite the over-expression of *MYC*, we did not observe a more interior nuclear positioning compared to *DLCI*, whose expression is much lower. One possible explanation could be linked to the recent observations suggesting that transcriptional levels in ~ 1 Mbp chromatin domains, rather than in individual genes, determine the nuclear radial position (Kölbl et al. 2012). Thus, we argue that the sum of transcription of all genes in the corresponding chromatin domains encompassing *MYC*, *DLCI*, and *SCN3B* could be similar. These chromatin domains might be consistent with the existence of TADs. Nevertheless, our data do show that *CCND1*, which is the only one of the four genes we analyzed that resides in a gene-dense R-band, is located more towards the interior of the nucleus compared to *SCN3B*, *MYC*, and *DLCI*, which are all located in a gene-poor G-band. This is in agreement with previous reports showing that regional gene density is a decisive parameter to determine the nuclear radial positioning in lymphocytes and fibroblasts (Küpfer et al. 2007). Therefore, we further support that positioning of genes within the interphase nucleus, not the CT, is independent on expression but dependent on local gene density.

The relationship between transcriptional activity and gene positioning with respect to the corresponding CT has been recurrently questioned. This association has been described for gene loci with cell type-related functions and coordinated patterns of gene expression such as MHC and EDC and for gene loci with distinctive expression levels during cell differentiation (Volpi et al. 2000; Williams et al. 2002; Stadler et al. 2004; Wiblin et al. 2005). Moreover, other studies have demonstrated that genes reproducibly locate at the border of the inactive X chromosome irrespective of their transcriptional activity (Clemson et al. 2006). The recent application of Hi-C to assess single-cell chromosome conformation suggested that the localization of transcriptionally active gene domains to the edges of CTs is a hallmark of large-scale chromosome structure and spatial nuclear arrangement (Nagano et al. 2013). Nevertheless,

limited experimental data have been reported for specific genes involved in physiological conditions, including cell cycle regulation. Our data show that cell cycle-related over-expressed genes are positioned further away from the center of its CT compared to under-expressed genes in two different *in vitro* models of adherent colorectal cancer cells. In fact, overall there was no circumstance in which the under-expressed gene was located more externally in the CT compared to the transcriptionally active gene. One of the most well-studied regions in the human genome is the gene-rich chromosome band 11p15.5, which contains ubiquitous active genes in several cell types. Chromatin at 11p15.5 frequently extends away from the chromosome 11 territory, and treatment with global transcription inhibitor actinomycin D and the RNA polymerase II inhibitor 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole partially decreases loci located beyond the chromosome edge (Mahy et al. 2002). The observation of chromatin looping out from the territory was not uncommon in our models. The co-localization analysis confirmed that over-expressed genes preferentially occupied positions away from the CT, indicating long-range chromatin decondensation to mediate transcription activity. However, it is noticeable that chromatin loops were preferentially observed in only one of the homologous chromosomes for both *MYC* and *CCND1*. Whether this implies that one locus is more transcriptionally active than the other (i.e., monoallelic expression) as it has been suggested in X-inactivated and a few autosomal genes needs further exploration (Gimelbrant et al. 2007; Takizawa et al. 2008). The biological correlation between form and function is further proved when *TCF7L2* downstream-regulated genes are assessed for their radial positioning with respect to the corresponding CT. Not unexpectedly, silencing of the transcription factor *TCF7L2* resulted in downregulation of *CCND1* (Hatzis et al. 2008). We observed repositioning of *CCND1* to a radial position closer to the center of the CT. Of note, the position of *CCND1* in *TCF7L2* downregulated cells was closer to the center than the transcriptionally inactive gene *SCN3B*. To the best of our knowledge, this is the first experimental evidence in a controlled system to show how induced transcriptional inactivation resulted in repositioning of genes in their CT. To this end, genome-wide transcriptional profiling would provide the chance to explore global gene expression changes upon deregulation of *TCF7L2* or any other transcription factor and their chromatin conformation rearrangements.

In summary, our results provide strong evidence of a gene positioning pattern that reflects expression levels: over-expressed genes show a more external position within the corresponding CT compared to under-expressed genes. Moreover, despite differences in chromatin condensation throughout the cell cycle, no major long-range dynamics has been observed for the CTs. This is one of the first studies where such measurements have been performed in cancer cell lines after silencing a specific transcription factor to advocate that gene radial position in CTs is determined by gene expression levels. Further studies

to understand relationships of nuclear structure and function in space and time, i.e., the 4D Nucleome, are needed to reveal the general principles of nuclear organization.

Acknowledgements The authors would like to thank Dr. Subhadra Banerjee from the Flow Cytometry Core Facility at NCI.

Compliance with ethical standards

Funding This work was supported by grants from the European Commission (COLONGEVA to J.C.), the Instituto de Salud Carlos III and co-funded by the European Regional Development Fund (ERDF) (CP13/00160 to J.C.), the CIBEREHD program, the Agència de Gestió d'Ajuts Universitaris i de Recerca, Generalitat de Catalunya (2014 SGR 135 and 2014 SGR 903), and from the Intramural Research Program of National Institutes of Health/NCI. J.C. has received a travel grant from the Instituto de Salud Carlos III and was co-funded by the European Regional Development Fund (ERDF) to perform experiments at NIH (MV15/00026). K.T. received a PIF fellowship and a mobility grant from the Universitat Autònoma de Barcelona (456-01-02/2013).

Conflict of interest The authors declare that they have no conflict of interest.

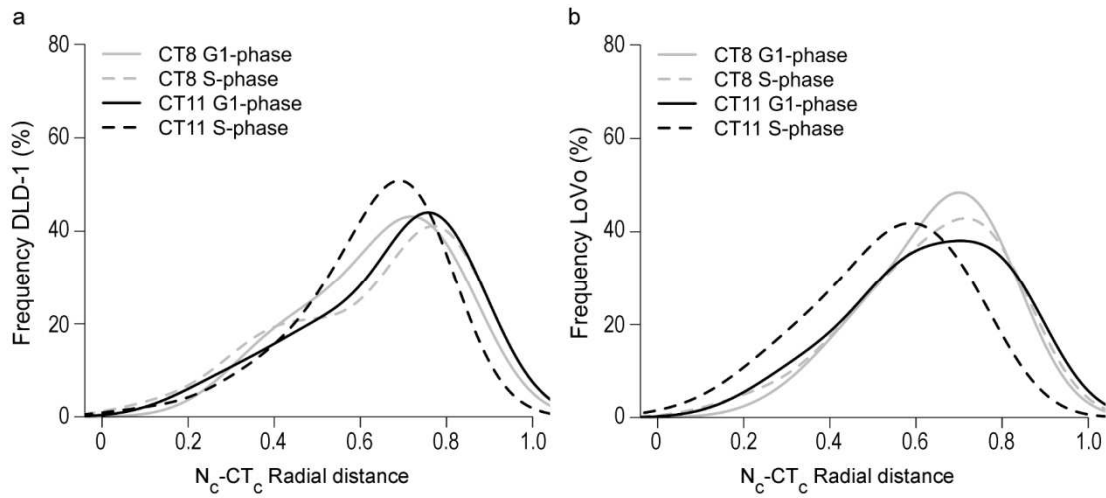
Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

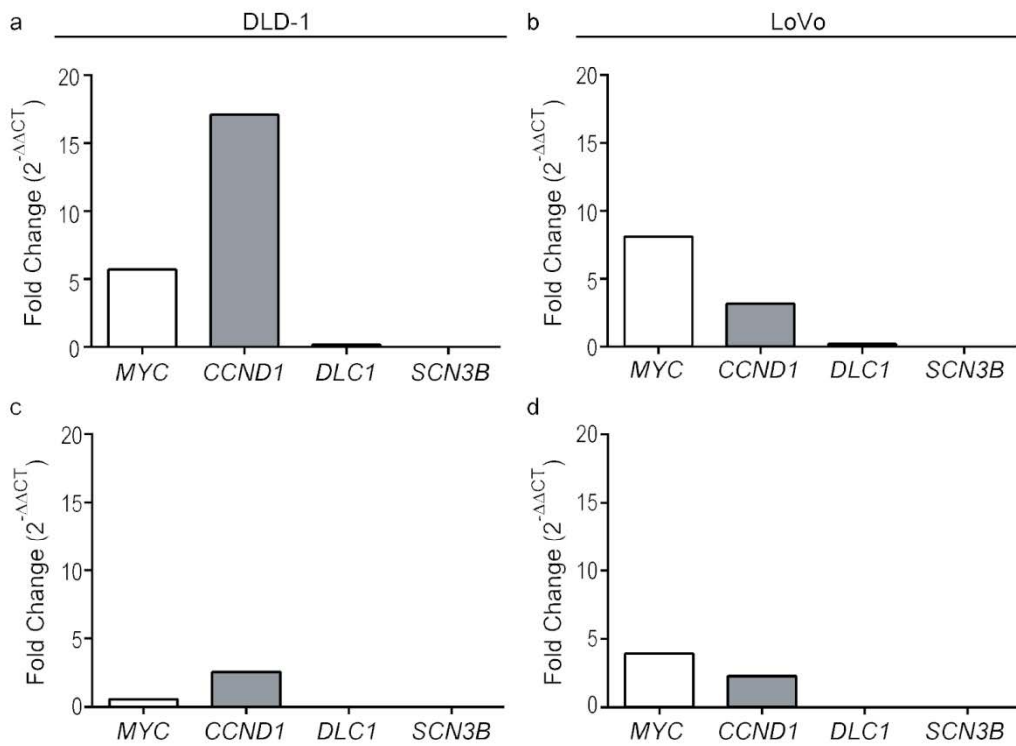
- Bickmore WA, van Steensel B (2013) Genome architecture: domain organization of interphase chromosomes. *Cell* 152:1270–1284
- Bohte S, Cordelières FP (2006) A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* 224:213–232
- Bolzer A, Kreth G, Solovei I et al (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* 3:e157
- Boyle S, Gilchrist S, Bridger JM et al (2001) The spatial organization of human chromosomes within the nuclei of normal and emerimutant cells. *Hum Mol Genet* 10:211–219
- Camps J, Nguyen QT, Padilla-Nash HM et al (2009) Integrative genomics reveals mechanisms of copy number alterations responsible for transcriptional deregulation in colorectal cancer. *Genes Chromosomes Cancer* 48:1002–1017
- Camps J, Pitt JJ, Emons G et al (2013) Genetic amplification of the NOTCH modulator LNX2 upregulates the WNT/ β -catenin pathway in colorectal cancer. *Cancer Res* 73:2003–2013
- Clemson CM, Hall LL, Byron M et al (2006) The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. *Proc Natl Acad Sci U S A* 103:7688–7693
- Cremer M, Küpper K, Wagler B et al (2003) Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. *J Cell Biol* 162:809–820
- Cremer M, von Hase J, Völm T et al (2001) Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res* 9:541–567
- Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2:292–301
- Cremer T, Cremer M (2010) Chromosome territories. *Cold Spring Harb Perspect Biol* 2:a003889
- Cremer T, Cremer M, Dietzel S et al (2006) Chromosome territories—a functional nuclear landscape. *Curr Opin Cell Biol* 18:307–316

- Croft JA, Bridger JM, Boyle S et al (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* 145:1119–1131
- Dekker J, Heard E (2015) Structural and functional diversity of topologically associating domains. *FEBS Lett* 589:2877–2884
- Dietzel S, Schiebel K, Little G et al (1999) The 3D positioning of ANT2 and ANT3 genes within female X chromosome territories correlates with gene activity. *Exp Cell Res* 252:363–375
- Dixon JR, Selvaraj S, Yue F et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–380
- Fritz AJ, Barutcu AR, Martin-Buley L et al (2016) Chromosomes at work: organization of chromosome territories in the interphase nucleus. *J Cell Biochem* 117:9–19
- Gerlich D, Beaudouin J, Kalbfuss B et al (2003) Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell* 112:751–764
- Gilbert N, Boyle S, Fiegler H et al (2004) Chromatin architecture of the human genome. *Cell* 118:555–566
- Gimelbrant A, Hutchinson JN, Thompson BR, Chess A (2007) Widespread monoallelic expression on human autosomes. *Science* 318:1136–1140
- Gindin Y, Valenzuela MS, Aladjem MI et al (2014) A chromatin structure-based model accurately predicts DNA replication timing in human cells. *Mol Syst Biol* 10:722
- Hatzis P, van der Flier LG, van Driel MA et al (2008) Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. *Mol Cell Biol* 28:2732–2744
- Ioannou D, Kandukuri L, Simpson JL, Tempest HG (2015) Chromosome territory repositioning induced by PHA-activation of lymphocytes: a 2D and 3D appraisal. *Mol Cytogenet* 8:47
- Kalhor R, Tjong H, Jayathilaka N et al (2012) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol* 30:90–98
- Kölbl AC, Weigl D, Mulaw M et al (2012) The radial nuclear positioning of genes correlates with features of megabase-sized chromatin domains. *Chromosome Res* 20:735–752
- Küpper K, Kölbl A, Biener D et al (2007) Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma* 116:285–306
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* 4:402–408
- Mahy NL, Perry PE, Bickmore WA (2002) Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol* 159:753–763
- Mayer R, Brero A, von Hase J et al (2005) Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biol* 6:44
- Meabum KJ, Misteli T (2008) Locus-specific and activity-independent gene repositioning during early tumorigenesis. *J Cell Biol* 180:39–50
- Meabum KJ, Newbold RF, Bridger JM (2008) Positioning of human chromosomes in murine cell hybrids according to synteny. *Chromosoma* 117:579–591
- Mehta IS, Amira M, Harvey AJ, Bridger JM (2010) Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts. *Genome Biol* 11:R5
- Mehta IS, Kulashreshtha M, Chakraborty S et al (2013) Chromosome territories reposition during DNA damage-repair response. *Genome Biol* 14:R135
- Mosimann C, Hausmann G, Basler K (2009) Beta-catenin hits chromatin: regulation of Wnt target gene activation. *Nat Rev Mol Cell Biol* 10:276–286
- Müller I, Boyle S, Singer RH et al (2010) Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells. *PLoS One* 5:e11560
- Murmann AE, Gao J, Encinosa M et al (2005) Local gene density predicts the spatial position of genetic loci in the interphase nucleus. *Exp Cell Res* 311:14–26
- Nagano T, Lubling Y, Stevens TJ et al (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502:59–64
- Nielsen JA, Hudson LD, Armstrong RC (2002) Nuclear organization in differentiating oligodendrocytes. *J Cell Sci* 115:4071–4079
- Ollion J, Cochenne J, Loll F et al (2013) TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* 29:1840–1841
- Popken J, Brero A, Koehler D et al (2014) Reprogramming of fibroblast nuclei in cloned bovine embryos involves major structural remodeling with both striking similarities and differences to nuclear phenotypes of in vitro fertilized embryos. *Nucleus* 5:555–589
- Ranade D, Koul S, Thompson J et al (2016) Chromosomal aneuploidies induced upon Lamin B2 depletion are mislocalized in the interphase nucleus. *Chromosoma*. doi:10.1007/s00412-016-0580-y
- Sehgal N, Fritz AJ, Morris K et al (2014) Gene density and chromosome territory shape. *Chromosoma* 123:499–513
- Sehgal N, Fritz AJ, Vecerova J et al (2016a) Large-scale probabilistic 3D organization of human chromosome territories. *Hum Mol Genet* 25:419–436
- Sehgal N, Seifert B, Ding H et al (2016b) Reorganization of the inter-chromosomal network during keratinocyte differentiation. *Chromosoma* 125:389–403
- Sengupta K, Camps J, Mathews P et al (2008) Position of human chromosomes is conserved in mouse nuclei indicating a species-independent mechanism for maintaining genome organization. *Chromosoma* 117:499–509
- Sengupta K, Upender MB, Barenboim-Stapleton L et al (2007) Artificially introduced aneuploid chromosomes assume a conserved position in colon cancer cells. *PLoS One* 2:e199
- Stadler S, Schnapp V, Mayer R et al (2004) The architecture of chicken chromosome territories changes during differentiation. *BMC Cell Biol* 5:44
- Strickfaden H, Zunhammer A, van Koningsbruggen S et al (2010) 4D chromatin dynamics in cycling cells: Theodor Boveri's hypotheses revisited. *Nucleus* 1:284–297
- Sun HB, Shen J, Yokota H (2000) Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys J* 79:184–190
- Takizawa T, Gudla PR, Guo L et al (2008) Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes Dev* 22:489–498
- Tanabe H, Müller S, Neusser M et al (2002) Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A* 99:4424–4429
- Volpi EV, Chevret E, Jones T et al (2000) Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci* 113:1565–1576
- Wang S, Su J-H, Beliveau BJ et al (2016) Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353:598–602
- Wiblin AE, Cui W, Clark AJ, Bickmore WA (2005) Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells. *J Cell Sci* 118:3861–3868
- Williams RRE, Azuara V, Perry P et al (2006) Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *J Cell Sci* 119:132–140
- Williams RRE, Broad S, Sheer D, Ragoussis J (2002) Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp Cell Res* 272:163–175
- Zink D, Amaral MD, Englmann A et al (2004) Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *J Cell Biol* 166:815–825

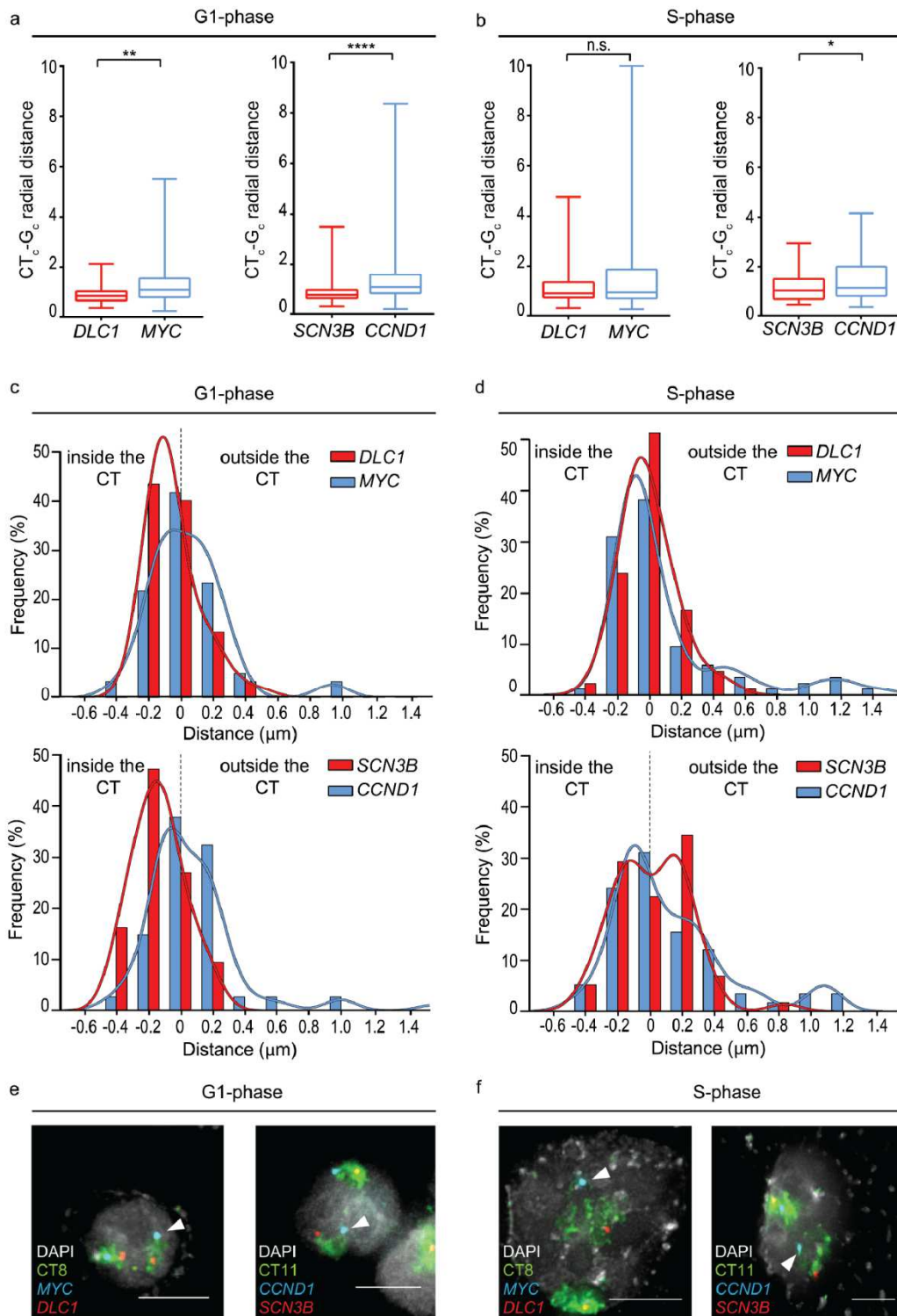
Supplementary Figure 1



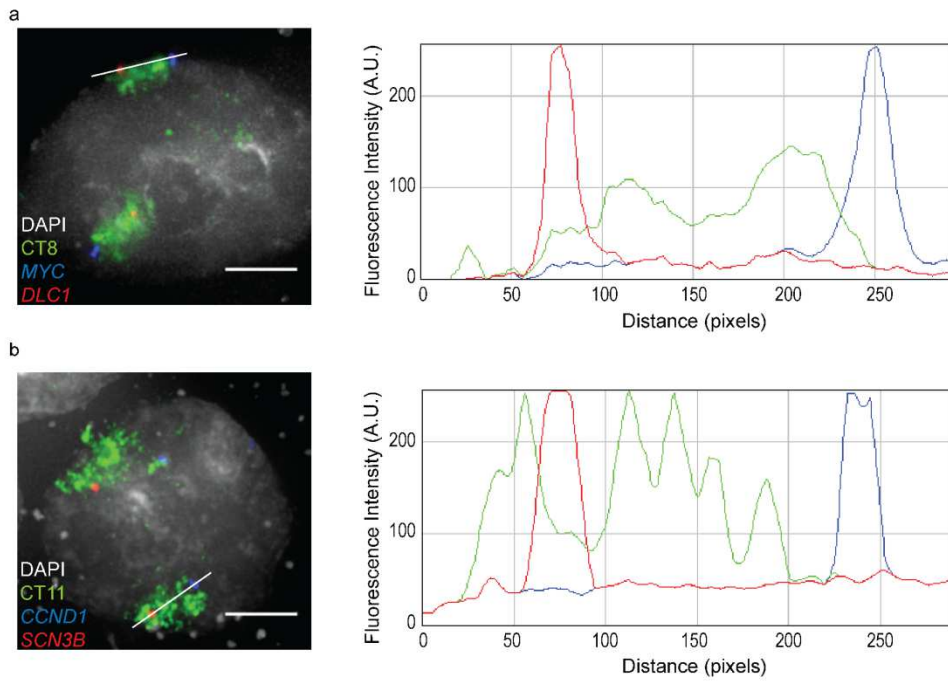
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5

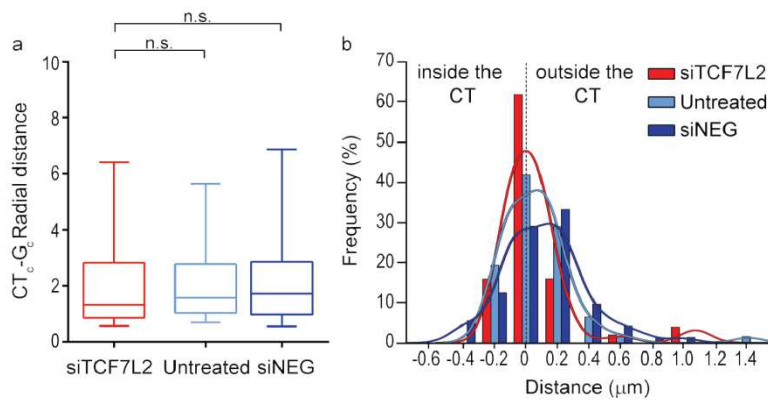


Table S1. Primer sequences for qRT-PCR

Gene	Forward	Reverse
<i>MYC</i>	GGACCCGCTTCTCTGAAAGG	TAACGTTGAGGGGCATCGTC
<i>DLC1</i>	AGGAACGCTGCCATGCTAGA	AAAATCCAGTGCTGGAGCAGA
<i>CCND1</i>	CAGAAGAGCGCGAGGGAG	CACAGGAGCTGGTGTTCAT
<i>SCN3B</i>	AACGCGTCTGACTACCTTGC	GTCAGCTGGCCTCATGTCAC
<i>HPRT1</i>	TGACACTGGCAAACAATGCA	GGTCCTTTTCACCAGCAAGCT
<i>YWHAZ</i>	ACTTTTGGTACATTGTGGCTTCAA	CCGCCAGGACAAACCAGTAT

Table S2. Gene radial distances with respect to the center of the CT

Cell line	Chr.	Cell cycle phase	Over-expressed gene ^a	Under-expressed gene ^a	<i>p</i> -value ^b
DLD-1	CT8	G1	1.131	0.852	0.0004
		S	1.429	1.042	0.0106
	CT11	G1	0.974	0.808	0.01
		S	1.687	0.985	0.0074
LoVo	CT8	G1	1.080	0.825	0.0019
		S	0.935	0.902	0.0743
	CT11	G1	1.075	0.765	0.0001
		S	1.124	1.024	0.0109

^a Values are indicated as the median of $n > 50$ territories.

^b Wilcoxon matched-pairs signed rank test.

4 Discussió

4.1 Eines d'estudi de la CIN al CRC

La presència de CNAs és una característica de la majoria de tumors sòlids (Beroukhim et al. 2010a). Aquestes alteracions es caracteritzen per ser guanys i pèrdues somàtiques de cromosomes sencers causats per errors de segregació en mitosi. Alhora, les CNAs també inclouen guanys i pèrdues de regions sub-cromosòmiques, causades en aquest cas per diversos mecanismes estructurals. Les CNAs són una característica especialment important al CRC, ja que al voltant del 85% de tumors s'engloba dins de l'entitat del CRC amb inestabilitat cromosòmica (Pino and Chung 2010). Tal i com ja s'ha descrit en la introducció, els tumors amb aquest tipus d'inestabilitat presenten una elevada taxa d'alteracions cromosòmiques, conduint-los a presentar uns cariotips altament aneuploides (Lengauer et al. 1998).

L'ús de tècniques de citogenètica clàssica i molecular va permetre la identificació d'aquests cariotips altament aneuploides per tal d'entendre el complex escenari que comporta adquirir inestabilitat cromosòmica. Posteriorment, aquests tipus d'estudis van deixar pas a l'anàlisi mitjançant metodologia basada en *arrays* de CGH. Aquest tipus de metodologia, juntament amb millores tecnològiques com els *arrays* d'SNPs o la seqüenciació de nova generació (NGS, de l'anglès *Next Generation Sequencing*), són els procediments més usuals avui en dia a l'hora d'estudiar les CNAs, entre altres tipus d'alteracions. Així, durant les últimes dues dècades, la irrupció i consolidació d'aquesta metodologia ha portat a la identificació de patrons de CNAs específics per diferents tipus de càncer (Zack et al. 2013a). No obstant, el primer gran repte que apareix en el moment d'avaluar aquests perfils característics de guanys i pèrdues cromosòmiques, està en determinar les CNAs responsables de la iniciació i/o progressió del tumor respecte els esdeveniments més eventuals fruit de la inestabilitat cromosòmica innata que presenta el càncer. Una estratègia per assolir aquest repte és la d'estudiar grans cohorts de tumors i així proporcionar evidències que les alteracions més freqüents puguin ser els esdeveniments amb un paper més funcional pel mateix càncer. D'aquesta manera, juntament amb la millora tecnològica, s'ha de

destacar el paper de grans projectes que han proporcionat l'oportunitat d'estudiar el càncer a gran escala, mitjançant l'estudi potencial de milers de mostres. El *The Cancer Genome Atlas* (TCGA), coordinat pel *National Cancer Institute* (NCI) i pel *National Human Genome Research Institute* (NHGRI), és un d'aquests grans projectes, vigent des de 2005. L'objectiu principal del TCGA és el de caracteritzar diferents tipus de càncer, 33 fins a data d'avui, en termes de CNAs, expressió gènica, metilació, microRNAs, variants de seqüència, entre d'altres anomalies característiques del càncer (Chang et al. 2013). L'ús recurrent de cohorts extretes del TCGA ha portat a la publicació de múltiples estudis amb el suficient poder estadístic per aportar conclusions biològicament rellevants sobre el coneixement d'aquesta malaltia (Cancer Genome Atlas Network et al. 2012).

És en aquest escenari on se situen part dels resultats presentats en aquesta tesi. L'estudi inicial, corresponent al capítol I, analitza dues cohorts de CRC diferents, una d'elles extreta del TCGA (n=303). A nivell experimental, les mostres de l'altra cohort (n=30) corresponen a DNA de tumors primaris procedents de pacients amb CRC. Aquestes mostres s'han estudiat al nostre laboratori mitjançant l'ús de dues plataformes d'*array* diferents, la primera correspon a un *array* de CGH anomenada *Human Genome CGH Microarray 244A* (Agilent) i la segona a un *array* d'SNPs, anomenada *Genechip Human Mapping 250K Sty array* (Affymetrix). L'ús d'aquestes dues plataformes ha permès la comparació i validació de la detecció de segments de CNAs entre els dos tipus d'*arrays*. Així, els resultats extrets d'ambdues plataformes han estat molt similars, on el 92% del segments de CNAs han concordat entre les dues anàlisis. Paral·lelament, per a la comparació entre la nostra cohort i la del TCGA, s'han usat els resultats extrets dels *arrays* d'SNPs, ja que la cohort del TCGA es basa en mostres analitzades mitjançant aquesta metodologia, específicament la plataforma *Genome-Wide Human SNP array 6.0* (Affymetrix).

4.2 Perfils genòmics de CNAs al CRC

La nostra cohort consisteix en mostres procedents de 30 pacients amb adenocarcinoma de còlon. Cadascuna de les mostres tumorals ha estat comparada amb una mostra de la respectiva mucosa normal adjacent. Aquesta

condició ha portat a la detecció única de CNAs a nivell somàtic, i per tant, específiques de la iniciació i/o progressió del tumor. Anàlisis d'*arrays* comparant el DNA tumoral amb un DNA normal de referència, tenen el risc potencial de sobreestimar la detecció de CNAs a causa de la multitud de CNVs, a nivell polimòrfic, que es troben repartits per tot el genoma (Zhang et al. 2009).

El perfil genòmic de CNAs, a nivell general, extret d'aquestes anàlisis mostra guanys recurrents als cromosomes i/o regions cromosòmiques 7, 8q, 13 i 20q, i pèrdues recurrents als cromosomes i/o regions cromosòmiques 4, 8p, 14, 15, 17p i 18, confirmant així el perfil característic de CNAs en CRC descrit amb anterioritat a la literatura (Camps et al. 2009; Cancer Genome Atlas Network et al. 2012). Aquest perfil tant conservat amb guanys i pèrdues tant freqüents com la dels cromosomes 7 i 18, respectivament, confirma la importància d'aquests esdeveniments en la progressió del CRC (Ried et al. 1996; Nakao et al. 2004). De fet, els guanys del cromosoma 7 i de la regió cromosòmica 20q s'han identificat múltiples vegades com a alteracions que es donen en fases primerenques, adenomes de baix grau, de la seqüència adenoma-carcinoma del CRC (Ried et al. 1996; Meijer et al. 1998; Höglund et al. 2002; Carvalho et al. 2009; Shi et al. 2012). Els nostres resultats mostren guany als cromosomes 7 i 20q en un 66,7% de les mostres. Per altra banda, l'anàlisi de FISH en teixit parafinat dels centròmers d'aquests mateixos cromosomes, exposat al capítol II d'aquesta tesi, mostra un nombre modal superior a dos en un 82% i un 62% de les mostres analitzades pels cromosomes 7 i 20, respectivament. Aquestes diferències en percentatge entre les diferents metodologies d'anàlisi es deuen a casos potencials de genomes poliploides, en els quals 3 o 4 senyals de FISH d'un centròmer en concret, equivalen a regions neutrals a nivell de nombre de còpies en els resultats de l'*array*. Pel què fa a la pèrdua del cromosoma 18, es considera que aquesta és un pas clau en la progressió d'adenoma avançat a carcinoma (Grade et al. 2006). Globalment, la gran majoria d'aquestes alteracions, descrites en adenomes i carcinomes, s'han identificat ben conservades en metàstasi i línies cel·lulars provinents de tumors primaris (Korn et al. 1999; Platzer et al. 2002). Els nostres resultats confirmen la recurrència d'aquesta alteració, ja que un 50% dels tumors analitzats per *arrays* mostren pèrdues al cromosoma 18. Les anàlisis de FISH per aquest mateix cromosoma

confirmen aquesta pèrdua recurrent mostrant una mediana de 1,48 en número de còpies per tota la cohort. Cal destacar que altres alteracions observades freqüentment a la nostra cohort, com el guany del cromosoma 13 o la pèrdua del braç curt del cromosoma 17, també formen part de les alteracions més comunes en la progressió del CRC (Meijer et al. 1998; Diep et al. 2006).

Amb la finalitat de validar els resultats obtinguts en la nostra cohort vam comparar els perfils obtinguts de CNAs amb la cohort extreta del TCGA. Qualitativament, els resultats de la nostra cohort mostren un alt grau de coincidència amb la cohort de CRC extreta del TCGA. No obstant, s'observen algunes discrepàncies a nivell quantitatiu. En general, la gran majoria de regions implicades en guanys o pèrdues en la nostra cohort mostren una freqüència més baixa de l'alteració en concret en la cohort del TCGA. Això pot ser degut, simplement, a la diferència de mostres que engloba cadascuna de les dues cohorts. D'altra banda, també pot ser degut a la potencial inclusió recurrent de mostres amb inestabilitat de microsatèl·lits (MSI) en la cohort del TCGA. Mentre que en la nostra cohort disposem d'informació clínica molt detallada de gran part dels pacients, incloses les anàlisis per detectar si els tumors es classifiquen dins el grup amb MSI, la informació clínica de les mostres del TCGA no sempre és prou completa. De fet, en la nostra cohort, tan sols hi ha dos casos que no mostren cap CNA, un de les quals sí que presenta MSI i un altre que correspon a un pacient amb CRC de tipus X, els quals tampoc acostumen a presentar CNAs (Goel et al. 2010). Mentre que la resta de tumors de la nostra cohort, presenten un perfil de CNAs compatibles amb el CRC amb CIN, alguns casos del TCGA no mostren cap o gairebé cap CNA.

El segon gran repte a assolir per part dels estudis genòmics és el d'identificar els gens clau, possibles dianes terapèutiques, que formen part de les CNAs, entre altres tipus d'alteracions, i que aporten una selecció positiva al tumor. Així, les regions amplificades han estat recurrentment associades a la presència d'oncogens. Els guanys característics del CRC com el del cromosoma 7, o les regions cromosòmiques 8q o 20q, contenen oncogens reconeguts en aquest tipus tumoral com són, *EGFR*, *MYC* o *AURKA*, respectivament (Carvalho et al. 2009). D'altra banda, basat en la teoria dels *two-hit* de Knudson, una de les aproximacions més usades a l'hora d'identificar

nous TSGs ha estat la de determinar la regió mínima de pèrdua en estudis de grans cohorts de tumors i avaluar els gens localitzats en aquesta regió per tal d'assignar si la pèrdua d'alguns d'aquests pot ser un esdeveniment important per al desenvolupament del càncer. Un clar exemple és la presència de dos TSGs, com són *SMAD2* i *SMAD4*, a la regió cromosòmica 18q que recurrentment es troba perduda al CRC. En aquest cas, tots dos gens formen part de la via de senyalització de TGF- β i es troben mutats al CRC en un 6 i 10%, respectivament (Cancer Genome Atlas Network et al. 2012). Una altra estratègia ha estat la identificació de delecions focals, tant en hemi- com homozigosi, involucrant normalment un gen, freqüentment observades al càncer. S'han desenvolupat diverses metodologies bioinformàtiques per tal de determinar regions focals afectant significativament diversos tipus de càncers (Beroukhim et al. 2007b). Aquesta estratègia ha portat a la successiva identificació de diversos TSGs en diversos tipus tumorals. Tanmateix, alguns casos de delecions focals es localitzen en llocs fràgils, els quals són regions cromosòmiques amb tendència a trencar-se (Rajaram et al. 2013b). En la nostra cohort hem identificat diversos exemples d'aquests esdeveniments focals, els quals inclouen els gens *FHIT*, *RBFOX1* i *C20orf133*, localitzats en les regions cromosòmiques 3p14.2, 16p13.3 i 20p12.1, respectivament. Tots tres casos són gens molt grans, entre 1,5 i 2 Mb, localitzats en llocs fràgils i amb altes freqüències de pèrdua en diferents tipus de càncer, a més del CRC (Smith et al. 2006; Camps et al. 2008; Beroukhim et al. 2010a). Diversos estudis suggereixen que aquestes alteracions corresponen a esdeveniments fortuïts i estan altament representades a causa d'una expansió clonal, però que no aporten cap mena d'avantatge selectiva al tumor. Així, aquestes delecions es formen simplement per la fragilitat de la regió cromosòmica en la qual es localitzen (Beroukhim et al. 2010a; Bignell et al. 2010a; Dereli-Öz et al. 2011). Contràriament, altres autors suggereixen que els gens localitzats a aquestes delecions focals presenten un paper funcionalment important per a la iniciació i/o progressió del càncer (Hazan et al. 2016). La hipòtesi d'aquest estudi és la de què els gens situats en llocs fràgils tenen un paper modulador de la resposta al dany al DNA (DDR, de l'anglès *DNA-Damage Response*) quan estan actius. A l'inactivar-se via deleció perden aquesta funció i es desencadena un

descontrol de la DDR a causa d'estrès replicatiu provocant consegüentment l'adquisició d'instabilitat genòmica (Hazan et al. 2016).

4.3 Disomia uniparental: alteració cromosòmica comuna al càncer

La UPD es va descriure per primera vegada a l'any 1980 com l'herència de dos cromosomes homòlegs que provenen de la mateixa via parental (Engel 1980b). Aquest esdeveniment ha estat llargament associat a alteracions a nivell constitucional causades, principalment, per errors en meiosi. Tanmateix, durant els últims anys la UPD ha anat prenent molt interès en l'estudi del càncer (Tuna et al. 2009). L'ús generalitzat dels arrays d'SNPs ha millorat la detecció de UPDs en estudis genòmics del càncer, mostrant regions cromosòmiques recurrentment afectades per aquest tipus d'esdeveniment en diferents tipus tumorals (Makishima and Maciejewski 2011). De la mateixa manera que en estudis de CNAs, el repte dels estudis genòmics de UPDs en càncer, és el d'identificar les alteracions que confereixen una avantatge selectiva a la cèl·lula tumoral, i diferenciar-les dels simples esdeveniments eventuais fruit de la instabilitat genòmica pròpia del tumor. Una d'aquestes avantatges pot ser el paper que la UPD adquireix esdevenint un mecanisme per a la inactivació de TSGs al càncer. Basat en la teoria dels *two-hit* de Knudson, un TSG pot estar inactivat a causa d'una mutació, somàtica o germinal, en un dels seus dos al·lells (Knudson 1971). Posteriorment, mitjançant la pèrdua del cromosoma que conté l'al·lel normal i la subseqüent duplicació del cromosoma que conté l'al·lel mutat, la cèl·lula tumoral acaba amb la presència d'un TSG inactivat, en ambdós al·lells, però sense pèrdua de material genòmic (Tuna et al. 2009). Tot i la potencial transcendència d'aquest esdeveniment, pocs estudis genòmics de grans cohorts similars als que s'han fet per a l'avaluació de CNAs, han descrit regions cromosòmiques afectades per UPD en càncer.

És arrel d'aquest context que hem establert un perfil genòmic de UPDs en dues cohorts diferents de CRC, descrit en el capítol I d'aquesta tesi. De manera similar que en el cas de les CNAs, l'ús aparellat del DNA de la mucosa normal i el DNA tumoral de cadascun dels pacients és essencial a l'hora d'identificar exclusivament les UPDs somàtiques, específiques de la iniciació i/o progressió del tumor (Bacolod et al. 2009). En aquest sentit, hem pogut descartar regions

cromosòmiques homozigotes constitucionals, compartides en ambdós teixits i conegudes de l'anglès com *identical-by-descent (IBD)* (Bacolod et al. 2009). De fet, la majoria d'algoritmes bioinformàtics usats en l'actualitat per inferir segments de CNAs i UPDs funcionen molt millor en estudis aparellats (Olshen et al. 2011; Li et al. 2014). A més, l'avaluació de la ràtio al·lèlica per part dels algoritmes basats en dades d'*arrays* d'SNPs, té una gran avantatge respecte els basats en *arrays* de CGH, i és el fet que es pot inferir la puresa del tumor (Van Loo et al. 2010; Carter et al. 2012). Tot i els esforços del patòlegs, la no inclusió de cèl·lules normals o cèl·lules no canceroses en mostres tumorals és pràcticament impossible. El grau de presència d'aquests tipus cel·lulars és una variable clau en la detecció d'alteracions cromosòmiques. Mentre que en una mostra amb poca presència de cèl·lules normals serà molt fàcil determinar les alteracions cromosòmiques pròpies del tumor, en una mostra on el percentatge d'aquest tipus cel·lular sigui molt elevat, les mateixes alteracions s'observaran de manera més diluïda (Van Loo et al. 2010; Li et al. 2014). A aquesta dificultat, s'afegeix el fet que entre les cèl·lules d'un mateix tumor existeix, sovint, una gran variabilitat a nivell d'anomalies cromosòmiques, fet conegut amb el terme heterogeneïtat intra-tumoral (Burrell et al. 2013b; Williams et al. 2016). Així, la presència de cèl·lules normals, sumat a l'heterogeneïtat tumoral inherent del càncer, comporta l'ús necessari d'algoritmes complexos que considerin la cel·lularitat específica de cada mostra tumoral (Van Loo et al. 2010; Li et al. 2014). Tenint en compte aquests factors, tots els resultats descrits referents a perfils de UPDs, estan basats en l'ús de l'algoritme PSCBS, el qual computa llimdars diferents en funció de la puresa específica de cada tumor (Olshen et al. 2011).

4.4 Perfils genòmics de UPD al CRC

Malgrat el recent interès en caracteritzar regions de UPD genòmiques al càncer, el nombre d'estudis publicats fins a l'actualitat sembla ser insuficient per tal de determinar les alteracions específiques que confereixen un avantatge positiu a la carcinogènesi colorectal. Contràriament, l'escenari actual mostra una variabilitat de regions cromosòmiques descrites com a freqüents, les quals depenen de la metodologia i cohorts utilitzades en cada cas. D'entrada, l'estudi d'Andersen i col·laboradors va ser dels primers en determinar que la UPD és

freqüent al CRC. El grup va utilitzar la plataforma *Genechip Mapping 10K* (Affymetrix) per analitzar 15 tumors primaris microdisseccionats i identificar com a regions de UPD freqüents els braços cromosòmics 3p, 8p, 13q i 20q (Andersen et al. 2007). Posteriorment, Kurashina i col·laboradors, van analitzar 94 tumors primaris, comparats amb les respectives mucoses normals, i van destacar les regions cromosòmiques 5q, 8p, 11, 14, 15, 17p i 18q com a freqüentment afectades per UPD (Kurashina et al. 2008). Ambdós estudis comparteixen únicament la identificació de 8p com a regió freqüent de UPD. Cal destacar, que l'observació recurrent de LOH en aquest braç cromosòmic en diversos tipus tumorals ha incrementat la cerca de gens candidats a ser TSGs en aquesta regió (Pole et al. 2006; Chua et al. 2009). Els nostres resultats confirmen l'afectació recurrent a 8p en CRC, ja que la meitat dels casos mostren pèrdues o UPDs en aquesta regió cromosòmica. En un altre estudi, Melcher i col·laboradors analitzen 16 tumors amb CIN, i identifiquen les regions cromosòmiques 5q21.1-q22.3 i 17p13.2-p13.1 com a les regions més freqüents de UPD en CRC esporàdic, amb un 31 i un 25%, respectivament (Melcher et al. 2011). Anteriorment, Lips i col·laboradors van analitzar una cohort específica de 78 tumors rectals dividits en adenomes i carcinomes. Mitjançant l'ús de la plataforma *GeneChip Mapping 10K 2.0 arrays* (Affymetrix) van observar que el grup de carcinomes mostrava UPDs involucrant els cromosomes 9 i 12, i les regions cromosòmiques 5q i 17p. Tanmateix, l'única regió de UPD destacable a nivell de freqüència al grup d'adenomes es localitzava a 5q (Lips et al. 2007). Recentment, l'estudi de Zarzour i col·laboradors ha analitzat, mitjançant la plataforma *Human610-Quad chips* (Illumina), 216 mostres en les que s'inclouen, 60 adenomes, 36 carcinomes d'estadiatge I i 120 carcinomes d'estadiatge III (Zarzour et al. 2015). Mentre que en la cohort de carcinomes d'estadiatge III s'identifiquen UPDs a 5q, 11q, 15q, 17p, 18q, 18q, 20p i 22q amb una freqüència superior al 10%, en la cohort d'adenomes la UPD a 5q és l'única que supera el mateix llindar de freqüència. Un estudi més recent no detecta aquesta alteració en mucosa normal comparat amb els respectius carcinomes, on la UPD a 5q torna a ser l'esdeveniment més freqüent (Conconi et al. 2016). Referent al nostre estudi, el perfil global de UPDs mostra regions recurrents superiors al 10%, als braços cromosòmics 1p, 3p, 5q, 8p, 9q, 10q, 14q, 17p, 20p, 21q i 22q i als cromosomes 18 i 19. A més de la UPD a 8p, cal

destacar les regions cromosòmiques 5q14.1-q31.2, 17p13.3-q25.3 i 18p11.32-q23, ja que són les tres més afectades per UPD i les més descrites en els estudis ja publicats. Aquestes tres regions de UPD s'han identificat també en diverses línies cel·lulars de CRC (Gaasenbeek et al. 2006; Melcher et al. 2007). Les tres regions inclouen TSGs molt involucrats en les vies de progressió del CRC com *APC* a 5q, *TP53* a 17p i *SMAD4* a 18q (Baker et al. 1989; Fearon and Vogelstein 1990). No obstant, la UPD a 5q sembla ser la més freqüent d'entre totes les descrites a CRC fins al moment (Tuna et al. 2009). A més, la identificació freqüent de UPD a 5q en adenomes, suggereix que aquest esdeveniment és una de les primeres alteracions del procés de carcinogènesi del còlon i del recte (Lips et al. 2007; Zarzour et al. 2015). Segons aquests autors, la UPD a 5q és el principal mecanisme pel qual el TSG *APC* perd la seva funció, especialment en la fase d'adenoma. De fet, la inactivació d'*APC* s'ha definit com a un dels esdeveniments més primerencs en la progressió d'adenoma de baix grau a adenoma avançat del model adenoma-carcinoma del CRC (Fearon and Vogelstein 1990).

4.5 UPD com a mecanisme alternatiu d'inactivació de TSGs

Els estudis genòmics en grans cohorts de càncers han permès la identificació de patrons de CNAs específics per tipus tumoral (Beroukhim et al. 2010a). Tanmateix, manquen estudis similars on es caracteritzi la UPD. A més, la integració amb dades mutacionals és essencial a l'hora de determinar quin paper funcional aporten aquestes alteracions al tumor. Amb el propòsit de confirmar que els perfils de UPD observats en CRC són específics d'aquest tipus tumoral, hem estudiat diferents cohorts tumorals del tracte gastrointestinal extretes del TCGA. A més, hem integrat aquestes dades amb l'anàlisi mutacional de les mateixes mostres per tal d'identificar la presència de TSGs inactivats en les regions de UPD. Així, al capítol II d'aquesta tesi hem comparat els perfils de UPDs, a més dels de CNAs, en tumors de còlon (COAD, de l'anglès *Colon Adenocarcinoma*) i recte (READ, de l'anglès *Rectum Adenocarcinoma*) per separat, estómac (STAD, de l'anglès *Stomach Adenocarcinoma*) i esòfag, que alhora es subdivideix en adenocarcinoma (ESAD, de l'anglès *Esophagus Adenocarcinoma*) i escamós (ESCC, de l'anglès *Esophagus Squamous Cell Carcinoma*).

En esòfag, Nancarrow i col·laboradors van determinar que les regions més freqüents de UPD en ESAD es localitzaven als braços cromosòmics 2p, 2q, 9q, 17 i 17q (Nancarrow et al. 2008). Posteriorment, el mateix grup va analitzar una altra cohort d'ESAD mostrant resultats molt similars als obtinguts anteriorment i confirmant que la UPD és un esdeveniment molt freqüent en l'adenocarcinoma d'esòfag (Frankel et al. 2014). D'altra banda, el perfil genòmic de UPDs en ESCC és força diferent segons un estudi on s'identifiquen els braços cromosòmics 2p, 5p, 19p, i 20q afectats per aquest tipus d'alteració (Hu et al. 2010). A nivell general, les nostres anàlisis extretes de les cohorts del TCGA mostren que a part d'algunes diferències quantitatives, els dos tipus de tumors d'esòfag comparteixen les dues regions més freqüents de UPD, localitzades als braços cromosòmics 9q i 17p. Referent a la regió 17p, en totes dues cohorts s'observen mutacions inactivants de *TP53* en aquells tumors que mostren UPD. De fet, *TP53* és el gen amb més variants identificades en regions afectades per UPD en ambdues cohorts. Pel què fa al cromosoma 9, cal destacar que la cohort d'ESAD tendeix a presentar UPD al braç llarg, mentre que en ESCC la UPD involucra el cromosoma sencer. Curiosament, la cohort d'ESCC mostra que en tots els casos on hem identificat una mutació inactivant al TSG *NOTCH1*, hem detectat també una UPD al cromosoma 9. Contràriament, tot i l'elevada freqüència de UPDs al braç 9q, no hem detectat mutacions inactivants de *NOTCH1* en ESAD. Mentre que variants puntuals en aquest gen s'han identificat prèviament en ESCC, no s'han descrit en ESAD (Agrawal et al. 2012). De fet, la via de senyalització de Notch, mitjançant la inactivació de *NOTCH1*, es troba afectada en la majoria de tumors epitelials escamosos, inclosos els carcinomes de pell, cavitat oral, esòfag i pulmó (Sakamoto 2016). Aquests resultats confirmen que la UPD influeix en la via Notch, inactivant *NOTCH1* específicament en ESCC.

A nivell de càncer gàstric, s'ha descrit molt recentment que els braços cromosòmics 5q, 11p, 11q i 12q presenten freqüències elevades de UPD (Arakawa et al. 2017). El mateix estudi compara els perfils genòmics de tumors avançats amb els perfils de tumors primerencs, els quals mostren CNAs però cap UPD. Les nostres anàlisis mostren que la regió més afectada en STAD es localitza a 17p. Similar a les cohorts descrites anteriorment, *TP53* torna a ser el

gen amb més mutacions inactivants en els tumors on s'observa UPD a 17p. En aquesta cohort, *ARID1A* és el segon gen amb més mutacions inactivants localitzades en regions de UPDs. Cal destacar, que la regió cromosòmica 1p35.3, on es localitza el gen, no s'inclou entre les més freqüents de UPD en STAD. Aquest gen és un membre de la família de remodeladors de la cromatina SWI/SNF, encarregats de regular la transcripció de *MYC* entre altres gens (Wu et al. 2014). Diversos autors han descrit mutacions recurrents a *ARID1A*, i l'han associat amb funcions supressives de tumor en càncer gàstric i colorectal, entre d'altres (Wang et al. 2011; Cancer Genome Atlas Network et al. 2012).

Finalment, en còlon i recte la UPD més freqüent torna a involucrar el cromosoma 5, com ja s'ha descrit al capítol I d'aquesta tesi, amb pics regionals de fins al 20% i 32%, respectivament. En ambdós tipus tumorals, *APC* és el gen amb més mutacions inactivants en les mostres on s'observa UPD a 5q. El gen *APC* té un paper funcional clau en la via de senyalització Wnt, regulant β -catenina, i es considera com a TSG per excel·lència al CRC esporàdic, trobant-se mutat fins a un 81% dels casos sense fenotip hipermutador (Cancer Genome Atlas Network et al. 2012). A més, variants que afecten aquest gen s'associen a diverses formes hereditàries del CRC (Lynch et al. 2003).

Globalment, els nostres resultats confirmen l'existència d'anomalies característiques per tipus tumoral. Mentre que en CRC la UPD a 5q és l'alteració més observada, als altres tipus tumorals la pèrdua del mateix braç és sempre més freqüent que la UPD. A nivell d'exemple, *APC* està involucrat en pèrdues cromosòmiques en un 38,8%, 59% i 32,2%, en ESAD, ESCC i STAD, respectivament. Contràriament, aquest TSG està involucrat en UPDs en un 16,7%, 5,1% i 7,4% en els mateixos tipus tumorals. La pèrdua del braç llarg del cromosoma 5 s'ha descrit en altres neoplàsies, especialment en síndromes mielodisplàsiques, on la presència d'aquesta i d'altres anomalies citogenètiques estan associades a un pitjor pronòstic (Fenaux and Kelaidi 2006). En canvi, la UPD a la mateixa regió sembla ser un tret característic i únic del CRC (Tuna et al. 2009). L'elevada freqüència d'aquesta anomalia en CRC, no només confirma l'especificitat de l'alteració per tal d'inactivar *APC*, sinó que també suggereix que la cèl·lula necessita mantenir gran part del cromosoma 5 amb

més d'una còpia, per tal de mantenir intacte la seva viabilitat. Si existeixen gens localitzats al braç 5q claus per a la supervivència de la cèl·lula tumoral en CRC caldrà demostrar-ho amb més anàlisis funcionals. A part de la inactivació recurrent d'*APC* en COAD i READ, cal fer èmfasi en les inactivacions de *NOTCH1* en ESCC i *ARID1A* en STAD, confirmant l'especificitat de la UPD a l'hora d'inactivar vies de senyalització que condueixen la carcinogènesi d'un tipus tumoral concret. Tanmateix, és important tenir en compte que aquest mecanisme és responsable també d'inactivar *TP53* TSG implicat en múltiples tipus tumorals.

A més de les diferències qualitatives, també hem observat diferències quantitatives entre tipus tumorals. Específicament, la cohort d'esòfag és el tipus tumoral amb major nombre de UPDs, amb una mitjana de 6,18 esdeveniments per cas, confirmant estudis anteriors on es suggereix que aquest tipus d'alteració és la més freqüent en càncer d'esòfag (Frankel et al. 2014). Tot seguit, se situa la cohort d'estómac amb una mitjana de 5,62 esdeveniments per mostra. Finalment, se situen recte i còlon amb unes mitjanes de 4,46 i 3,28 esdeveniments per mostra, respectivament. Aquestes diferències es mantenen estables quan només es tenen en compte UPDs segmentals, ja siguin telomèriques o intersiticials, indicant que el càncer d'esòfag i el d'estómac acumulen més inestabilitat cromosòmica estructural comparat amb el CRC. Prèviament, ja s'havia descrit un increment de CIN estructural en adenocarcinomes de la part alta del tracte gastrointestinal, d'esòfag i estómac, comparat amb adenocarcinomes colorectals (Dulak et al. 2012). En aquest estudi, Dulak i col·laboradors observaven un increment d'amplificacions focals en tumors d'esòfag i estómac comparat amb tumors colorectals. S'ha suggerit que la constant exposició a àcids biliars, a la qual estan sotmesos l'esòfag i l'estómac, col·labora a la carcinogènesi mitjançant la formació de DSBs (Bernstein et al. 2005). Aquest increment de formació de DSBs contribueix en l'elevada inestabilitat cromosòmica estructural observada en tumors esofàgics i gàstrics comparat amb tumors colorectals (Dulak et al. 2012).

4.6 La UPD és més freqüent en casos altament aneuploides

Tradicionalment l'única manera de determinar la ploïdia d'un tumor era mitjançant l'obtenció de metafases de les cèl·lules tumorals en divisió i comptar el número total de cromosomes observats. Arrel de l'ús dels *arrays* d'SNPs, diverses eines bioinformàtiques s'han desenvolupat per tal de determinar el nombre de còpies a nivell al·lèlic i així quantificar la ploïdia general del tumor (Van Loo et al. 2010; Carter et al. 2012). Gràcies a l'ús d'aquestes eines s'ha descrit la ploïdia genòmica en grans cohorts de diferents tipus de càncer (Zack et al. 2013a). L'elevada freqüència de genomes altament aneuploides, entre els que s'inclouen genomes propers a ser triploides i superiors (més de 58 cromosomes), fa pensar que la tetraploidització del genoma seguit de la pèrdua consecutiva de diversos cromosomes és habitual en la progressió del càncer, especialment en tumors d'origen epitelial (Carter et al. 2012). De fet, s'ha descrit que les cèl·lules tetraploides mostren més CIN numèrica, ja que aquestes cèl·lules tenen més cromosomes per poder perdre que no pas un genoma diploide, on la monosomia té efectes molt perjudicials (Storchova and Pellman 2004; Storchova and Kuffer 2008). L'ús de l'ASCAT en la nostra i en les cohorts del TCGA ha confirmat que gran part dels casos, entre un 44,3% i un 63,9% dependent del tipus tumoral, corresponen a genomes amb valors de ploïdies superiors a 2,5. Un genoma amb una ploïdia de 2,5 correspon a un cariotip de 58 cromosomes definit com a cariotip hipotriploide (entre 58-68 cromosomes). Els nostres resultats mostren que els casos altament aneuploides presenten un nombre de UPDs significativament més elevat que els casos propers a ser diploides, confirmant que els tumors que provenen de duplicacions del seu genoma presenten una elevada taxa d'errors de segregació en mitosi. Si en aquests tumors existeix una preferència per perdre cromosomes que provenen de la mateixa via parental caldrà demostrar-ho a partir de més estudis. D'altra banda, la diferència en nombre de UPDs entre diferents grups de ploïdies tumorals es manté quan es comparen UPDs telomèriques i intersticials. Ambdós tipus d'esdeveniments, són anomalies iniciades per defectes estructurals i no pas per defectes de segregació mitòtica. Aquests resultats suggereixen que els genomes que potencialment provenen d'una tetraploidització adquireixen paral·lelament a l'increment de CIN

numèrica un increment de CIN estructural. De fet, l'adquisició de CIN estructural, mitjançant la identificació d'un increment de l'estrès replicatiu de la cèl·lula, s'ha descrit recentment en tumors amb CIN elevada, relacionant ambdós tipus d'inestabilitats, numèrica i estructural (Burrell et al. 2013a).

4.7 La UPD telomèrica és el tipus d'esdeveniment més present, afectant principalment a *APC*

Tot i la recurrent descripció de la UPD com a un esdeveniment que involucra un cromosoma sencer, la gran majoria d'alteracions observades en càncer corresponen a anomalies segmentals (Tuna et al. 2009; O'Keefe et al. 2010). Específicament, les UPDs telomèriques, aquelles que inclouen la gran majoria d'un braç cromosòmic, són les més descrites a la literatura (Makishima and Maciejewski 2011). Les diferents cohorts del TCGA que hem analitzat confirmen que la telomèrica és el tipus de UPD més freqüent amb una mitjana de 1,89, 2,31, 2,47 i 3,12 esdeveniments per còlon, recte, estómac i esòfag, respectivament. La UPD telomèrica involucrant el braç 5q és de les més freqüents, considerant totes les cohorts estudiades del TCGA. La inactivació recurrent d'*APC* causada per aquest esdeveniment, suggereix que aquesta és una alteració important per a la iniciació i/o progressió del CRC. De fet, la UPD a 5q s'ha descrit com un esdeveniment primerenc, afectant a adenomes freqüentment (Zarzour et al. 2015). D'altra banda, en múltiples casos on no s'observa UPD a 5q, una pèrdua a nivell de braç cromosòmic o a nivell intersticial pot resultar també en la inactivació d'*APC*. Entre les diferents cohorts de CRC analitzades en aquesta tesi, s'han recollit més de cent casos amb *APC* afectat, ja sigui per deleció o per UPD. Les freqüències dels dos tipus d'alteracions són similars, tot i que la regió cromosòmica que afecta cadascun dels esdeveniments és diferent. Mentre que les pèrdues a nivell de braç tenen com a punt de trencament localitzacions properes al centròmer, les delecions intersticials s'agrupen en regions més properes al *locus* d'*APC*. Contràriament, la gran majoria de casos on *APC* està involucrat en regions de UPD, aquestes són telomèriques, afectant una regió força àmplia a nivell de braç cromosòmic. Cal destacar que la distribució dels punts de trencament al llarg del braç 5q associats a la UPD telomèrica difereix de l'observada per les pèrdues. Els nostres resultats mostren que en la gran majoria de casos els punts de

trencament es localitzen entre les 65 i les 80 Mb del braç 5q. Aquesta anàlisi, confirma els resultats publicats per Howarth i col·laboradors, els quals van observar una distribució similar referent als punts de trencament a 5q associada a tipus d'alteració (Howarth et al. 2009). Aquests autors van analitzar pacients FAP, els quals es caracteritzen per presentar inactivació germinal d'APC. El mateix grup aporta evidències sobre la presència de punts de recombinació meiòtica a la localització cromosòmica on s'acumulen els trencaments associats a la UPD telomèrica de 5q. De fet, observen un enriquiment de punts de trencament entre les 68 i 71 Mb del braç 5q, on s'acumulen seqüències de repetició del tipus LCR (de l'anglès, *Low Copy Repeats*). Aquests LCRs, incrementen la probabilitat de trencaments al DNA i de l'estancament de la força de replicació, que conduirà finalment la cèl·lula a intentar reparar el trencament mitjançant replicació associada a trencament (BIR, de l'anglès *Break Induced Replication*) creant finalment la regió de UPD telomèrica. D'altra banda, els autors suggereixen que els trencaments associats a pèrdues no es poden reparar a causa de la proximitat dels mateixos al centròmer, on les seqüències altament repetitives poden conduir a l'aparellament de seqüències no específiques. Contràriament, els nostres resultats no reforcen l'últim punt d'aquesta hipòtesi, ja que molts dels casos on s'observa una UPD telomèrica els punts de trencament es localitzen en regions pericentromèriques. Finalment, els autors suggereixen que la pèrdua de 5q acostuma a succeir en casos molt aneuploides, on la conseqüència de l'alteració no és tant perjudicial per la cèl·lula, mentre que la UPD a la mateixa regió és un esdeveniment primerenc que inicia la carcinogènesi (Howarth et al. 2009). Aquests resultats, explicarien l'elevada freqüència de UPD a 5q que s'observa en adenomes (Lips et al. 2007; Zarzour et al. 2015). Alguns dels tumors de les cohorts que hem estudiat no presenten CNAs, però mostren la UPD telomèrica a 5q o una deleció afectant específicament APC. D'altra banda, la pèrdua del braç sencer tendeix a presentar-se acompanyada d'altres anomalies citogenètiques, resultant en genomes força aneuploides.

Un altre estudi en CRC identifica casos de UPDs telomèriques no associades al braç cromosòmic 5q (Middeldorp et al. 2008). En aquest cas, Middeldorp i col·laboradors observen una freqüència molt elevada de UPDs en carcinomes

de pacients amb MAP (de l'anglès, *MUTYH-Associated Polyposis*), comparat amb la freqüència mitjana trobada en casos esporàdics. Aquests tumors presenten una elevada taxa de mutacions associada als defectes en la via de reparació per excisió de bases (BER, de l'anglès *Base Excision Repair*), característica dels pacients amb mutació germinal del gen *MUTYH*. La gran majoria dels esdeveniments que es descriuen en aquests carcinomes són UPDs segmentals i inclouen les regions cromosòmiques 17p (57%), 18q (52%), 15q (52%) i 6p (36%). Els autors suggereixen que l'increment de mutacions que presenten aquests tumors, associats als defectes de la via BER, afavoreixen els esdeveniments de la recombinació mitòtica com a via de reparació alternativa (Middeldorp et al. 2008). Aquest fet podria ser la causa de l'elevada freqüència de UPDs segmentals que presenten aquests carcinomes comparat amb casos esporàdics.

Altres estudis similars s'han realitzat en pacients amb neurofibromatosi de tipus I (Garcia-Linares et al. 2011). Aquests pacients desenvolupen neurofibromes causats per la inactivació dels dos al·lels del gen *NF1* (Serra et al. 1997). En aquest estudi, Garcia-Linares i col·laboradors identifiquen LOH a 17q, afectant principalment el TSG *NF1* en un 24.9% dels neurofibromes analitzats. Entre aquests, un 62% mostren UPD telomèrica, mentre que el 38% restant mostren delecions focals, entre 80 kb i 8Mb, afectant específicament *NF1*. Els autors conclouen que la recombinació mitòtica és el mecanisme més freqüent per formar les UPDs a 17q i, conseqüentment, inactivar *NF1*. Aquests resultats, com els observats en *APC* en les nostres cohorts de CRC, confirmen la inactivació mitjançant UPD telomèrica de TSGs específics de tipus tumoral.

4.8 La importància de l'arquitectura nuclear en la formació de UPDs segmentals

La recombinació homòloga (HR, de l'anglès *Homologous Recombination*) és un mecanisme especialitzat per a la reparació de DSBs (Moynahan and Jasin 2010). Normalment, aquest mecanisme utilitza seqüències de DNA idèntiques com a motlle per a reparar aquests DSBs. Així, després de la replicació del DNA aquest mecanisme disposa d'una cromàtide germana per recombinar i poder reparar el trencament conservant la informació genètica. Tanmateix,

altres seqüències homòlogues, com les corresponents al cromosoma homòleg, també poden utilitzar-se com a motlles per a la reparació (Moynahan and Jasin 1997; Neuwirth et al. 2007). Aquest tipus de recombinació resulta en una pèrdua d'heterozigositat, la qual pot comportar la inactivació d'un gen en concret. Estudis en llevats han demostrat que, generalment, la cèl·lula prioritza l'ús de cromàtides germanes a l'hora de recombinar (Kadyk and Hartwell 1992). No obstant, s'ha observat que la inactivació de RecQ, una família d'helicases amb un paper clau en l'HR, és suficient per incrementar l'intercanvi entre cromosomes homòlegs (Carotenuto and Liberi 2010). Tot i aquesta inactivació, la recombinació entre cromàtides germanes continua sent més freqüent. Cal destacar, que mutacions germinals en tres gens de la família d'helicases RecQ estan implicades en cinc malalties d'herència recessiva com són les síndromes de Bloom pel gen *BLM*, de Werner pel gen *WRN*, i de Rothmund-Thomson, Rapadilino i Baller-Gerold pel gen *RECQ4* (Chu and Hickson 2009). Entre aquestes síndromes, la més estudiada és la de Bloom, els individus de la qual presenten un increment d'instabilitat genòmica, caracteritzat per un excés no només d'intercanvi entre cromàtides germanes sinó també entre cromosomes homòlegs, per la qual cosa desenvolupen predisposició al càncer (Ellis and German 1996).

Un altre dels factors que pot estar implicat en la generació d'intercanvi entre cromosomes homòlegs és la proximitat espacial d'aquests cromosomes en el nucli interfàsic. Així es pot hipotetitzar que la prioritat que té la cèl·lula de recombinar amb la cromàtide germana es pot deure a la baixa probabilitat de contacte entre les molècules de DNA de cromosomes homòlegs (Moynahan and Jasin 2010). En qualsevol cas, l'elevada freqüència de UPDs segmentals, potencialment formades per HR, observades al càncer fan pensar que ja sigui la proximitat espacial o mutacions en RecQ poden ser determinants en afavorir la recombinació entre cromosomes homòlegs. En quant a la possible implicació de la distribució espacial dels cromosomes homòlegs, el fet que les UPDs siguin específiques de cromosoma podria suggerir que certs cromosomes homòlegs es troben en relativa proximitat al nucli interfàsic. Per exemple, l'increment en el nombre de UPDs segmentals en genomes altament aneuploides, descrit en el capítol II d'aquesta tesi, podria suggerir que en

aquests casos els cromosomes homòlegs es troben més propers dins del nucli interfàsic. Arrel d'aquesta hipòtesi, és necessari entendre diversos conceptes de l'arquitectura nuclear per tal d'adquirir el coneixement necessari sobre el possible paper que té l'estructura del nucli en la formació de UPDs. El dogma central en el camp de l'arquitectura nuclear exposa que els cromosomes ocupen unes àrees concretes dins el nucli interfàsic anomenades territoris cromosòmics (CTs, de l'anglès *Chromosome Territories*) (Cremer and Cremer 2010). Aquests CTs es disposen dins el nucli d'una manera no aleatòria, seguint una organització radial en funció de la densitat gènica del cromosoma (Croft et al. 1999; Boyle et al. 2001; Cremer and Cremer 2001). Aquesta organització nuclear ha estat confirmada mitjançant l'ús de tècniques de Hi-C, les quals evidencien que els CTs més rics en gens, a l'estar recurrentment posicionats al centre del nucli tenen una freqüència de contacte entre ells molt més elevada que els CTs més pobres en gens (Lieberman-Aiden et al. 2009; Kalhor et al. 2011). En aquest sentit, els nostres resultats mostren una clara associació entre densitat gènica del cromosoma, i la freqüència de UPDs telomèriques en CRC (Annex; Figura 16). Així, la proximitat entre cromosomes homòlegs en el nucli interfàsic podria explicar una elevada freqüència de UPDs telomèriques en determinats cromosomes. Tanmateix, cromosomes que no són especialment rics en gens, també mostren regions intra-cromosòmiques freqüentment afectades per UPD en CRC, com és el cas del braç cromosòmic 8p, la qual cosa obre la porta a pensar que no solament la posició del CT, sinó per exemple l'estructura de la cromatina pugui estar implicada en la generació de UPDs.

Tal i com es presenta en el capítol III d'aquesta tesi, l'ús de pintats cromosòmics mitjançant 3D-FISH ens ha permès avaluar la localització nuclear del cromosoma 8 i 11, els quals presenten freqüències de UPD en CRC molt diferents entre ells. Mentre que el cromosoma 8 presenta freqüències elevades de UPD en CRC, el cromosoma 11 gairebé no en presenta. D'entrada, tot i la diferència en densitat gènica, no hem observat patrons diferencials de posicionament radial nuclear entre els dos territoris, localitzant-se ambdós en zones més aviat perifèriques i mostrant distàncies similars entre cromosomes homòlegs. A més, la posició radial és similar durant el cicle cel·lular, confirmant

així que el posicionament dins el nucli interfàsic dels CTs s'estableix a G1 i es manté molt estable durant tot el cicle cel·lular (Strickfaden et al. 2010; Müller et al. 2010). Aquesta distribució tan conservada s'ha descrit en diverses espècies de mamífers, evidenciant la importància que aquesta organització nuclear té en la funcionalitat de la cèl·lula (Tanabe et al. 2002). De fet, grans moviments o canvis en la localització de CTs dins el nucli interfàsic, modificant aquesta arquitectura tant conservada, s'han descrit únicament en condicions molt específiques, com per exemple en càncer, algunes laminopaties i processos de diferenciació cel·lular (Bridger et al. 2014; Ioannou et al. 2015b; Sehgal et al. 2016). Recentment, un treball realitzat en línies cel·lulars de CRC ha confirmat que Lamina B2 (LMNB2), una proteïna de la làmina nuclear, té un paper important en el manteniment de la distribució territorial de cromosomes aneuploides en el nucli interfàsic (Ranade et al. 2017). Un altre estudi confirma que existeixen moviments de CTs dins el nucli interfàsic arrel d'induir dany al DNA (Mehta et al. 2013). Els autors suggereixen que el moviment dels CTs succeeix per tal de reparar el possible dany. A més, el grup identifica que generalment són els CTs més rics en gens els que porten a terme aquest moviment, suggerint una reparació guiada per la densitat gènica específica de cada cromosoma. Més recentment, en un estudi molt similar a l'anterior, Ioannou i col·laboradors aporten més evidències de moviments de CTs arrel de la inducció de dany al DNA (Ioannou et al. 2015a). En aquest últim estudi el cromosoma 8, entre d'altres, presenta diferències significatives de moviments intra-nuclears entre limfòcits exposats a dany al DNA i limfòcits no exposats. Paral·lelament, un altre grup ha induït DSBs en regions de DNA associades a la làmina nuclear (LADs, de l'anglès *Lamina Associated Domains*), i ha determinat que la unió d'extrems no homòlegs (NHEJ, de l'anglès *Non-Homologous End Joining*) és el mecanisme pel qual la cèl·lula repara aquests trencaments (Lemaître et al. 2014). Aquest mecanisme, a diferència de l'HR, no requereix de seqüències homòlogues per tal de reparar els DSBs ja que lliga directament els extrems trencats (Moore and Haber 1996). Contràriament, quan el DSB es localitza en DNA de l'interior del nucli o dels porus nuclears, la cèl·lula activa la resposta al dany de DNA i el trencament es repara mitjançant HR (Lemaître et al. 2014). Així, els autors suggereixen un model on la posició nuclear és un factor clau per a la determinació del mecanisme de reparació del

DNA per part de la cèl·lula. En aquest sentit són necessaris més estudis per tal d'entendre si la distribució dels cromosomes en el nucli interfàsic pot estar relacionada amb l'etiologia de UPDs i si factors implicats en l'arquitectura nuclear poden ser dependents de tipus tumoral, la qual cosa podria explicar l'aparició de patrons de UPD específics, tal i com s'ha descrit en els capítols I i II de la secció de resultats.

A més de l'estudi de distribució radial dels CTs, l'anàlisi de quatre gens localitzats en regions diferents dels cromosomes 8 i 11 ens ha permès l'avaluació del paper de diferents variables en el posicionament de la cromatina, a nivell gènic, dins el nucli interfàsic. Per començar, tot i que el braç curt del cromosoma 8 tendeix a presentar més UPDs que el braç llarg en CRC, el gen *DLC1*, localitzat a 8p22, no mostra diferències significatives de posicionament nuclear respecte *MYC*, localitzat a 8q24.21. Contràriament, sí que s'observen diferències significatives en el cas de la comparació entre *CCND1* i *SCN3B*, localitzats a 11q13.3 i 11q24.1, respectivament. Així *CCND1*, es localitza més a l'interior del nucli que el gen *SCN3B*. Globalment, aquests resultats contradiuen una teoria llargament discutida a la literatura, la qual s'associa una elevada expressió de gens individuals amb un posicionament més a l'interior del nucli interfàsic (Zink et al. 2004; Williams et al. 2006). Les nostres anàlisis mostren que el gen *MYC*, altament expressat en les dues línies cel·lulars estudiades, es posiciona freqüentment a prop de la perifèria del nucli, de la mateixa manera que ho fan els dos gens infra-expressats que s'han estudiat, *DLC1* al mateix cromosoma 8, i *SCN3B* al cromosoma 11. Treballs més recents utilitzant microscòpia d'alta resolució argumenten que la posició radial dins el nucli no s'estructura a nivell de gens individuals sinó a nivell de dominis cromatínics definits com a regions aproximades d'1Mb (Kölbl et al. 2012). L'expressió global dels gens inclosos en aquests dominis podria ser el factor que determina la posició dins el nucli. Recentment, l'ús de la Hi-C, ha definit uns dominis cromatínics anomenats TADs (de l'anglès, *Topologically Associated Domains*), confirmant la mida dels dominis descrits per microscòpia (1Mb), com a unitat bàsica d'organització nuclear a nivell de la cromatina (Dixon et al. 2012). D'altra banda, els nostres resultats mostren que mentre *MYC*, *DLC1* i *SCN3B*, localitzats en regions cromosòmiques pobres en gens o

bandes G, se situen més cap a la perifèria del nucli, *CCND1*, el qual està localitzat en una regió cromosòmica rica en gens o banda R, se situa més cap al centre. Aquests resultats suggereixen una distribució radial dins el nucli en funció de la densitat gènica local, confirmant així estudis previs que descriuen que la densitat gènica local és un paràmetre decisiu a l'hora d'establir la posició radial dins el nucli interfàsic (Küpper et al. 2007).

4.9 L'expressió gènica és un factor clau per a la localització intracromosòmica dels gens

A continuació vam estudiar la relació entre l'activitat transcripcional i la posició dels gens dins els propis CTs. El model d'arquitectura nuclear anomenat CT-IC diferencia dos compartiments cromatínics dins els CTs, l'actiu (ANC, de l'anglès *Active Nuclear Compartment*) i l'inactiu (INC, de l'anglès *Inactive Nuclear Compartment*). Els autors del model proposen que l'ANC està format principalment per cromatina transcripcionalment activa i descondensada. A més, l'ANC també el formen canals intercromatínics que connecten l'interior i exterior del CT amb el citoplasma mitjançant el porus nuclears, proporcionant un espai on la transcripció dels gens es pot dur a terme. Contràriament, el component inactiu està format per cromatina més compactada. La funcionalitat d'aquest model es basa en el fet que els gens altament expressats es localitzen en regions més properes al compartiment intercromatínic, on poden tenir accés més fàcil a les maquinàries de transcripció. Així, el grup del Dr. Cremer, va ser pioner en suggerir que gens altament expressats tendeixen a situar-se a l'exterior del territori comparat amb gens amb nivells d'expressió més baixos (Zirbel et al. 1993). Tanmateix, diversos treballs no han observat cap mena de relació entre l'expressió gènica i la localització dins el propi CT (Kurz et al. 1996; Clemson et al. 2006). Els nostres resultats mostren que tot i que de manera general els quatre gens estudiats es troben a la perifèria del propi CT, els gens sobre-expressats, *MYC* i *CCND1*, es troben radialment disposats més lluny respecte el centre del CT comparat amb els gens infra-expressats, *DLC1* i *SCN3B*, respectivament. A més, mitjançant una anàlisi de co-localització de píxels en imatges tridimensionals, hem identificat que les senyals de fluorescència dels gens sobre-expressats i el CT corresponent co-localitzen menys, comparat amb les senyals dels gens infra-expressats i el mateix CT.

Aquests resultats confirmen que *MYC* i *CCND1* es troben recurrentment a la cara exterior del propi CT, mentre que *DLC1* i *SCN3B* es mantenen a la cara interior. Tot i la gran heterogeneïtat que s'observa en experiments on s'avaluen cèl·lules úniques, cap de les condicions analitzades mostra, a nivell global, que els gens infra-expressats es localitzin més freqüentment a l'exterior del CT comparat amb el gens sobre-expressats. Diversos estudis han descrit regions cromatíniques que contenen gens altament expressats i que es localitzen a la perifèria del CT corresponent (Dietzel et al. 1999; Volpi et al. 2000; Williams et al. 2002; Mahy et al. 2002). Alguns d'aquests estudis identifiquen no només que aquestes regions se situen més a l'exterior del CT, sinó que a més ho fan en forma de llargues estructures cromatíniques definides com a *loops* (Volpi et al. 2000; Williams et al. 2002; Mahy et al. 2002). No obstant, la poca evidència microscòpica de l'existència d'aquestes estructures cromatíniques és un dels punts de discussió més evidents entre els diferents models d'arquitectura nuclear proposats fins a dia d'avui (Küpper et al. 2007; Cremer and Cremer 2010). Un dels models més contraris a l'existència freqüent d'aquestes estructures és el CT-IC, mencionat anteriorment (Cremer et al. 2015). En aquest model, es considera altament improbable el contacte generalitzat entre CTs. Contràriament, diversos models s'oposen a aquesta individualitat i suggereixen que existeix un contacte global entre CTs. Entre aquests models destaca l'anomenat ICN (Branco and Pombo 2006). Els autors d'aquest model proposen que els *loops* són estructures prou comunes i que ajuden a la interconnexió entre CTs. Aquests *loops* col·laboren en la co-regulació de múltiples gens en punts concrets del nucli on es localitzen les maquinàries necessàries per a la transcripció (Kosak and Groudine 2004b; Fraser and Bickmore 2007b). L'ús actual i més generalitzat de tècniques derivades de la 3C aporten evidències sobre l'arquitectura nuclear des d'un punt de vista més probabilístic que els estudis de 3D-FISH (Dekker et al. 2002). Un dels models d'organització més reconeguts extrets mitjançant Hi-C és el proposat per Lieberman-Aiden i col·laboradors, el qual suggereix que la cromatina es plega dins el nucli interfàsic organitzant-se com a un glòbul fractal (Lieberman-Aiden et al. 2009). L'avantatge d'aquest tipus d'organització és que la cromatina pot descondensar-se fàcil i ràpidament gràcies a la inexistència d'entrecreuaments cromatínics que defineix el model del glòbul fractal. Aquest estudi confirma

l'existència de compartiments actius on la cromatina es troba més descondensada que la cromatina localitzada als compartiments inactius (Lieberman-Aiden et al. 2009). Recentment, un estudi de Hi-C, a nivell de cèl·lules individuals, confirma la presència de cromatina transcripcionalment activa en dominis propers a la perifèria dels CTs (Nagano et al. 2013). A més, l'estudi proposa la poca probabilitat de contacte entre dominis actius mitjançant *loops*. D'altra banda, estudis de models de polímers confirmen l'existència generalitzada de diferents mides de *loops* associats a co-regulació (Mateos-Langerak et al. 2009; Di Stefano et al. 2013). Tot i la discrepància que existeix en la presència generalitzada de *loops* en nuclis interfàsics, les nostres anàlisis mostren unes estructures cromatíniques compatibles amb la descripció de *loops*, i representades de manera específica als gens sobre-expressats. A més, en gran part dels nuclis on hem identificat una d'aquestes estructures, aquesta es localitza al CT més descondensat. Més estudis són necessaris per tal d'entendre si aquests gens s'expressen de manera mono-al·lèlica, i si els al·lèls que s'expressen són els que formen aquestes estructures cromatíniques.

Finalment, vam poder demostrar la relació entre l'activitat gènica i el posicionament respecte el CT mitjançant l'ús de tècniques de RNA d'interferència induint el silenciament del factor de transcripció TCF7L2. Aquest factor de transcripció té un paper modul·lador de la via de senyalització Wnt/ β -catenina, la qual està constitucionalment activada en la majoria de cèl·lules de CRC (Mosimann et al. 2009). TCF7L2 s'associa amb β -catenina quan aquesta entra al nucli, i regulen *MYC* i *CCND1*, entre d'altres gens involucrats en el desenvolupament del CRC (Hatzis et al. 2008). No obstant, els nostres experiments no van demostrar un descens en l'expressió de *MYC* després de la inactivació temporal de TCF7L2. Conseqüentment, en aquestes cèl·lules *MYC* no mostra diferències de localització respecte el CT8 comparat amb les cèl·lules no tractades. Contràriament, sí que vam poder demostrar un descens aproximat del 50% en l'expressió de *CCND1*. D'aquesta manera, els nostres resultats mostren que a causa del descens d'expressió de *CCND1*, aquest gen es re-posiciona cap a l'interior del CT11. De fet, aquesta va ser l'única condició estudiada on el gen *SCN3B* es va observar freqüentment més lluny del centre del CT que *CCND1*. Resultats similars es van obtenir en l'estudi de Mahy i

col·laboradors, on a conseqüència de l'addició d'un inhibidor global de la transcripció, com és l'actinomicina D, es va observar un descens dels *loops* de la regió cromatínica 11p15.5 en el CT11 (Mahy et al. 2002). Tanmateix, els nostres resultats són els primers en avaluar aquest re-posicionament usant una inhibició transcripcional més específica. Globalment, aquests resultats evidencien la funcionalitat d'una distribució cromatínica a nivell de CT, confirmant l'existència de la relació estructura-funció dins el nucli interfàsic.

5 Conclusions

Chapter I

1. CRC shows a specific pattern of UPDs, mainly involving chromosome arms 5q, 14q, 17p, 18q and 21q.
2. UPD regions match with copy number losses profiles, providing regions with a high incidence of loss of heterozygosity in CRC.
3. TSGs are significantly located in recurrent regions of UPD.
4. UPD acts as a second hit in order to biallelically inactivate *APC* in CRC.

Chapter II

5. UPD genomic profiles are cancer-specific.
6. The integrative analysis of regions affected by UPDs and copy number losses with gene mutational status provides further evidence that UPD acts as the “second hit” in order to inactivate TSGs.
7. Highly aneuploid genomes present a higher number of whole-chromosome and segmental UPDs events compared to near-diploid genomes.
8. Segmental UPDs including both, telomeric and interstitial, are more frequent events compared to whole-chromosome UPDs.
9. *APC* inactivation in CRC is mainly involving a telomeric UPD at chromosome arm 5q, clustering between 60 and 80 Mbp, potentially caused by a mitotic recombination event.

Chapter III

10. CT8 and CT11 displayed similar radial position in the interphase nucleus in CRC cell lines.
11. Nuclear radial gene distribution is independent of expression levels
12. Over-expressed genes show a more external position with respect the CT compared to under-expressed genes.
13. Silencing of *TCF7L2* cause down-regulation of *CCND1* and reposition of this gene to a more internal location within CT11.

6 Bibliografia

- Agrawal N, Jiao Y, Bettgowda C, et al (2012) Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discov* 2:899–905.
- Al-Tassan N, Chmiel NH, Maynard J, et al (2002) Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat Genet* 30:227–232.
- Albiez H, Cremer M, Tiberi C, et al (2006) Chromatin domains and the interchromatin compartment form structurally defined and functionally interacting nuclear networks. *Chromosome Res* 14:707–33.
- Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet* 12:363–76.
- Andersen CL, Wiuf C, Kruhoffer M, et al (2007) Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis* 28:38–48.
- Arakawa N, Sugai T, Habano W, et al (2017) Genome-wide analysis of DNA copy number alterations in early and advanced gastric cancers. *Mol Carcinog* 56:527–537.
- Astler VB, Coller FA (1954) The prognostic significance of direct extension of carcinoma of the colon and rectum. *Ann Surg* 139:846–52.
- Bacolod MD, Schemmann GS, Giardina SF, et al (2009) Emerging paradigms in cancer genetics: some important findings from high-density single nucleotide polymorphism array studies. *Cancer Res* 69:723–7.
- Baia GS, Stifani S, Kimura ET, et al (2008) Notch activation is associated with tetraploidy and enhanced chromosomal instability in meningiomas. *Neoplasia* 10:604–12.
- Baker SJ, Fearon ER, Nigro JM, et al (1989) Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* 244:217–21.
- Bauman JG, Wiegant J, Borst P, van Duijn P (1980) A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. *Exp Cell Res* 128:485–90.
- Bea S, Salaverria I, Armengol L, et al (2009) Uniparental disomies, homozygous deletions, amplifications, and target genes in mantle cell lymphoma revealed by integrative high-resolution whole-genome profiling. *Blood* 113:3059–3069.
- Bernstein H, Bernstein C, Payne C, et al (2005) Bile acids as carcinogens in human gastrointestinal cancers. *Mutat Res Mutat Res* 589:47–65.
- Beroukhim R, Getz G, Nghiemphu L, et al (2007a) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 104:20007–12.

- Beroukhim R, Getz G, Nghiemphu L, et al (2007b) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 104:20007–12.
- Beroukhim R, Mermel CH, Porter D, et al (2010a) The landscape of somatic copy-number alteration across human cancers. *Nature* 463:899–905.
- Beroukhim R, Mermel CH, Porter D, et al (2010b) The landscape of somatic copy-number alteration across human cancers. *Nature* 463:899–905.
- Bertoin F, Letouzé E, Grignani P, et al (2014) Genome-Wide Paternal Uniparental Disomy as a Cause of Beckwith-Wiedemann Syndrome Associated with Recurrent Virilizing Adrenocortical Tumors. *Horm Metab Res* 47:497–503.
- Bignell GR, Greenman CD, Davies H, et al (2010a) Signatures of mutation and selection in the cancer genome. *Nature* 463:893–898.
- Bignell GR, Greenman CD, Davies H, et al (2010b) Signatures of mutation and selection in the cancer genome. *Nature* 463:893–8.
- Boland CR, Goel A (2010) Microsatellite Instability in Colorectal Cancer. *Gastroenterology* 138:2073–2087.e3.
- Bolzer A, Kreth G, Solovei I, et al (2005) Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. *PLoS Biol* 3:e157.
- Boveri T (1902) Über mehrpolige mitosen als mittel zur analyse des zellkerns. *Verh Phys Med Ges Würzburg* 35:67–90.
- Boveri T (1909) Die Blastomerenkerne von *Ascaris megalocephala* und die Theorie der Chromosomenindividualität. *Arch Zellforsch* 3:181–268.
- Boyle S, Gilchrist S, Bridger JM, et al (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* 10:211–9.
- Branco MR, Pombo A (2006) Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* 4:e138.
- Bridger JM, Arican-Gotkas HD, Foster HA, et al (2014) The Non-random Repositioning of Whole Chromosomes and Individual Gene Loci in Interphase Nuclei and Its Relevance in Disease, Infection, Aging, and Cancer. In: *Advances in experimental medicine and biology*. pp 263–279
- Brittan M, Wright NA (2004) The gastrointestinal stem cell. *Cell Prolif* 37:35–53.
- Burrell RA, McClelland SE, Endesfelder D, et al (2013a) Replication stress links structural and numerical cancer chromosomal instability. *Nature* 494:492–6.
- Burrell RA, McGranahan N, Bartek J, Swanton C (2013b) The causes and

- consequences of genetic heterogeneity in cancer evolution. *Nature* 501:338–345.
- Caldwell CM, Green RA, Kaplan KB (2007) APC mutations lead to cytokinetic failures in vitro and tetraploid genotypes in Min mice. *J Cell Biol* 178:1109–20.
- Camps J, Grade M, Nguyen QT, et al (2008) Chromosomal Breakpoints in Primary Colon Cancer Cluster at Sites of Structural Variants in the Genome. *Cancer Res* 68:1284–1295.
- Camps J, Morales C, Prat E, et al (2004) Genetic evolution in colon cancer KM12 cells and metastatic derivatives. *Int J cancer* 110:869–74.
- Camps J, Nguyen QT, Padilla-Nash HM, et al (2009) Integrative genomics reveals mechanisms of copy number alterations responsible for transcriptional deregulation in colorectal cancer. *Genes Chromosomes Cancer* 48:1002–17.
- Cancer Genome Atlas Network DM, Bainbridge MN, Chang K, et al (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487:330–7.
- Carén H, Erichsen J, Olsson L, et al (2008) High-resolution array copy number analyses for detection of deletion, gain, amplification and copy-neutral LOH in primary neuroblastoma tumors: four cases of homozygous deletions of the CDKN2A gene. *BMC Genomics* 9:353.
- Carotenuto W, Liberi G (2010) Mitotic inter-homologue junctions accumulate at damaged DNA replication forks in recQ mutants. *DNA Repair (Amst)* 9:661–669.
- Carter SL, Cibulskis K, Helman E, et al (2012) Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 30:413–421.
- Carvalho B, Postma C, Mongera S, et al (2009) Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. *Gut* 58:79–89.
- Cassidy SB, Schwartz S (1998) Prader-Willi and Angelman syndromes. *Disorders of genomic imprinting. Medicine (Baltimore)* 77:140–51.
- Chambeyron S, Bickmore WA (2004) Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev* 18:1119–30.
- Chang K, Creighton CJ, Davis C, et al (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45:1113–1120.
- Cheung VG, Spielman RS, Ewens KG, et al (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 437:1365–9.
- Chu WK, Hickson ID (2009) RecQ helicases: multifunctional genome

- caretakers. *Nat Rev Cancer* 9:644–654.
- Chua YL, Ito Y, Pole JCM, et al (2009) The *NRG1* gene is frequently silenced by methylation in breast cancers and is a strong candidate for the 8p tumour suppressor gene. *Oncogene* 28:4041–52.
- Chubb JR, Bickmore WA (2003) Considering nuclear compartmentalization in the light of nuclear dynamics. *Cell* 112:403–6.
- Clemson CM, Hall LL, Byron M, et al (2006) The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. *Proc Natl Acad Sci U S A* 103:7688–93.
- Coe BP, Ylstra B, Carvalho B, et al (2007) Resolving the resolution of array CGH. *Genomics* 89:647–53.
- Colussi D, Brandi G, Bazzoli F, Ricciardiello L (2013) Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. *Int J Mol Sci* 14:16365–85.
- Conconi D, Redaelli S, Bovo G, et al (2016) Unexpected frequency of genomic alterations in histologically normal colonic tissue from colon cancer patients. *Tumour Biol* 37:13831–13842.
- Conlin LK, Thiel BD, Bonnemann CG, et al (2010) Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum Mol Genet* 19:1263–75.
- Cremer C, Rappold G, Gray JW, et al (1984) Preparative dual-beam sorting of the human Y chromosome and in situ hybridization of cloned DNA probes. *Cytometry* 5:572–9.
- Cremer M, Grasser F, Lanctôt C, et al (2008) Multicolor 3D Fluorescence In Situ Hybridization for Imaging Interphase Chromosomes. In: *Methods in molecular biology* (Clifton, N.J.). pp 205–239
- Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2:292–301.
- Cremer T, Cremer C, Lichter P (2014) Recollections of a scientific journey published in human genetics: from chromosome territories to interphase cytogenetics and comparative genome hybridization. *Hum Genet* 133:403–16.
- Cremer T, Cremer M (2010) Chromosome territories. *Cold Spring Harb Perspect Biol* 2:a003889.
- Cremer T, Cremer M, Hübner B, et al (2015) The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. *FEBS Lett* 589:2931–43.
- Croft JA, Bridger JM, Boyle S, et al (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* 145:1119–31.

- Cunningham JM, Christensen ER, Tester DJ, et al (1998) Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Res* 58:3455–60.
- Davoli T, de Lange T (2011) The Causes and Consequences of Polyploidy in Normal Development and Cancer. *Annu Rev Cell Dev Biol* 27:585–610.
- Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* 295:1306–11.
- del Rey J, Prat E, Ponsa I, et al (2010) Centrosome clustering and cyclin D1 gene amplification in double minutes are common events in chromosomal unstable bladder tumors. *BMC Cancer* 10:280.
- DePristo MA, Banks E, Poplin R, et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498.
- Dereli-Öz A, Versini G, Halazonetis TD (2011) Studies of genomic copy number changes in human cancers reveal signatures of DNA replication stress. *Mol Oncol* 5:308–314.
- Desai DC, Neale KF, Talbot IC, et al (1995) Juvenile polyposis. *Br J Surg* 82:14–7.
- DeSantis CE, Lin CC, Mariotto AB, et al (2014) Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin* 64:252–271.
- Di Stefano M, Rosa A, Belcastro V, et al (2013) Colocalization of coregulated genes: a steered molecular dynamics study of human chromosome 19. *PLoS Comput Biol* 9:e1003019.
- Diep CB, Kleivi K, Ribeiro FR, et al (2006) The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. *Genes, Chromosom Cancer* 45:31–41.
- Dietzel S, Schiebel K, Little G, et al (1999) The 3D positioning of ANT2 and ANT3 genes within female X chromosome territories correlates with gene activity. *Exp Cell Res* 252:363–75.
- Dikovskaya D, Schiffmann D, Newton IP, et al (2007) Loss of APC induces polyploidy as a result of a combination of defects in mitosis and apoptosis. *J Cell Biol* 176:183–95.
- Dixon JR, Selvaraj S, Yue F, et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–80.
- Dulak AM, Schumacher SE, van Lieshout J, et al (2012) Gastrointestinal Adenocarcinomas of the Esophagus, Stomach, and Colon Exhibit Distinct Patterns of Genome Instability and Oncogenesis. *Cancer Res* 72:4383–4393.
- Dunlop MG, British Society for Gastroenterology, Association of Coloproctology

- for Great Britain and Ireland (2002) Guidance on gastrointestinal surveillance for hereditary non-polyposis colorectal cancer, familial adenomatous polyposis, juvenile polyposis, and Peutz-Jeghers syndrome. *Gut* 51 Suppl 5:V21-7.
- Eldai H, Periyasamy S, Al Qarni S, et al (2013) Novel genes associated with colorectal cancer are revealed by high resolution cytogenetic analysis in a patient specific manner. *PLoS One* 8:e76251.
- Ellis NA, German J (1996) Molecular genetics of Bloom's syndrome. *Hum Mol Genet* 5 Spec No:1457-63.
- Engel E (1980a) A new genetic concept: uniparental disomy and its potential effect, isodisomy. *Am J Med Genet* 6:137-43.
- Engel E (1980b) A new genetic concept: Uniparental disomy and its potential effect, isodisomy. *Am J Med Genet* 6:137-143.
- Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. *Cell* 61:759-67.
- Federico C, Scavo C, Cantarella CD, et al (2006) Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates. *Chromosoma* 115:123-8.
- Fenaux P, Kelaidi C (2006) Treatment of the 5q- Syndrome. *Hematology* 2006:192-198.
- Ferlay J, Soerjomataram I, Dikshit R, et al (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J cancer* 136:E359-86.
- Fitzgibbon J, Iqbal S, Davies A, et al (2007) Genome-wide detection of recurring sites of uniparental disomy in follicular and transformed follicular lymphoma. *Leukemia* 21:1514-20.
- Fitzgibbon J, Smith L-L, Raghavan M, et al (2005) Association between Acquired Uniparental Disomy and Homozygous Gene Mutation in Acute Myeloid Leukemias. *Cancer Res* 65:9152-9154.
- Fleming ID, Phillips JL, Menck HR, et al (1997) The National Cancer Data Base report on recent hospital cancer program progress toward complete American Joint Committee on Cancer/TNM staging. *Cancer* 80:2305-10.
- Flotho C, Steinemann D, Mullighan CG, et al (2007) Genome-wide single-nucleotide polymorphism analysis in juvenile myelomonocytic leukemia identifies uniparental disomy surrounding the NF1 locus in cases associated with neurofibromatosis but not in cases with mutant RAS or PTPN11. *Oncogene* 26:5816-21.
- Forconi F, Poretti G, Kwee I, et al (2008) High density genome-wide DNA profiling reveals a remarkably stable profile in hairy cell leukaemia. *Br J Haematol* 141:622-30.

- Frankel A, Armour N, Nancarrow D, et al (2014) Genome-wide analysis of esophageal adenocarcinoma yields specific copy number aberrations that correlate with prognosis. *Genes Chromosomes Cancer* 53:324–38.
- Fraser P, Bickmore W (2007a) Nuclear organization of the genome and the potential for gene regulation. *Nature* 447:413–7.
- Fraser P, Bickmore W (2007b) Nuclear organization of the genome and the potential for gene regulation. *Nature* 447:413–7.
- Fu XJ, Morisada N, Hashimoto F, et al (2014) A patient with autosomal recessive Alport syndrome due to segmental maternal isodisomy. *Hum genome Var* 1:14006.
- Gaasenbeek M, Howarth K, Rowan AJ, et al (2006) Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. *Cancer Res* 66:3471–9.
- Galipeau PC, Cowan DS, Sanchez CA, et al (1996) 17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. *Proc Natl Acad Sci U S A* 93:7081–4.
- Ganem NJ, Godinho SA, Pellman D (2009) A mechanism linking extra centrosomes to chromosomal instability. *Nature* 460:278–82.
- Garcia-Linares C, Fernández-Rodríguez J, Terribas E, et al (2011) Dissecting loss of heterozygosity (LOH) in neurofibromatosis type 1-associated neurofibromas: Importance of copy neutral LOH. *Hum Mutat* 32:78–90.
- Gaymes TJ, Mohamedali A, Eiliazadeh AL, et al (2017) FLT3 and JAK2 Mutations in Acute Myeloid Leukemia Promote Interchromosomal Homologous Recombination and the Potential for Copy Neutral Loss of Heterozygosity. *Cancer Res* 77:1697–1708.
- Gerlinger M, Rowan AJ, Horswell S, et al (2012) Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N Engl J Med* 366:883–892.
- Goel A, Xicola RM, Nguyen T-P, et al (2010) Aberrant DNA methylation in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology* 138:1854–62.
- Goetze S, Mateos-Langerak J, Gierman HJ, et al (2007) The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol Cell Biol* 27:4475–87.
- Gondek LP, Tiu R, O'Keefe CL, et al (2008) Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. *Blood* 111:1534–42.
- González JR, Rodríguez-Santiago B, Cáceres A, et al (2011) A fast and accurate method to detect allelic genomic imbalances underlying mosaic

- rearrangements using SNP array data. *BMC Bioinformatics* 12:166.
- Grade M, Becker H, Liersch T, et al (2006) Molecular cytogenetics: genomic imbalances in colorectal cancer and their clinical impact. *Cell Oncol* 28:71–84.
- Grasser F, Neusser M, Fiegler H, et al (2008) Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. *J Cell Sci* 121:1876–86.
- Gray JW, Dean PN, Fuscoe JC, et al (1987) High-speed chromosome sorting. *Science* 238:323–9.
- Gümüş G, Sunguroğlu A, Tükün A, et al (2002) Common fragile sites associated with the breakpoints of chromosomal aberrations in hematologic neoplasms. *Cancer Genet Cytogenet* 133:168–71.
- Hagstrom SA, Dryja TP (1999) Mitotic recombination map of 13cen-13q14 derived from an investigation of loss of heterozygosity in retinoblastomas. *Proc Natl Acad Sci U S A* 96:2952–7.
- Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* 100:57–70.
- Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144:646–74.
- Hansermann D (1890) Über asymmetrische zelltheilung in epithelkrebsen und deren biologische bedeutung. *Arch Pathol Anat Physiol Klin Med* 119:299–326.
- Harada T, Chelala C, Bhakta V, et al (2008) Genome-wide DNA copy number analysis in pancreatic cancer using high-density single nucleotide polymorphism arrays. *Oncogene* 27:1951–60.
- Hatzis P, van der Flier LG, van Driel MA, et al (2008) Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. *Mol Cell Biol* 28:2732–44.
- Hazan I, Hofmann TG, Aqeilan RI (2016) Tumor Suppressor Genes within Common Fragile Sites Are Active Players in the DNA Damage Response. *PLOS Genet* 12:e1006436.
- Hepperger C, Mannes A, Merz J, et al (2008) Three-dimensional positioning of genes in mouse cell nuclei. *Chromosoma* 117:535–51.
- Heselmeyer-Haddad K, Berroa Garcia LY, Bradley A, et al (2012) Single-Cell Genetic Analysis of Ductal Carcinoma in Situ and Invasive Breast Cancer Reveals Enormous Tumor Heterogeneity yet Conserved Genomic Imbalances and Gain of MYC during Progression. *Am J Pathol* 181:1807–1822.
- Höglund M, Gisselsson D, Hansen GB, et al (2002) Dissecting karyotypic patterns in colorectal tumors: two distinct but overlapping pathways in the adenoma-carcinoma transition. *Cancer Res* 62:5939–46.

- Holland AJ, Cleveland DW (2009) Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol* 10:478–87.
- Howarth K, Ranta S, Winter E, et al (2009) A mitotic recombination map proximal to the APC locus on chromosome 5q and assessment of influences on colorectal cancer risk. *BMC Med Genet* 10:54.
- Hu N, Clifford RJ, Yang HH, et al (2010) Genome wide analysis of DNA copy number neutral loss of heterozygosity (CNNLOH) and its relation to gene expression in esophageal squamous cell carcinoma. *BMC Genomics* 11:576.
- Ioannou D, Kandukuri L, Quadri A, et al (2015a) Spatial positioning of all 24 chromosomes in the lymphocytes of six subjects: evidence of reproducible positioning and spatial repositioning following DNA damage with hydrogen peroxide and ultraviolet B. *PLoS One* 10:e0118886.
- Ioannou D, Kandukuri L, Simpson JL, Tempest HG (2015b) Chromosome territory repositioning induced by PHA-activation of lymphocytes: A 2D and 3D appraisal. *Mol Cytogenet* 8:47.
- Jacobs KB, Yeager M, Zhou W, et al (2012) Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet* 44:651–658.
- Jasek M, Gondek LP, Bejanyan N, et al (2010) TP53 mutations in myeloid malignancies are either homozygous or hemizygous due to copy number-neutral loss of heterozygosity or deletion of 17p. *Leukemia* 24:216–9.
- Jasmine F, Rahaman R, Dodsworth C, et al (2012) A genome-wide study of cytogenetic changes in colorectal cancer using SNP microarrays: opportunities for future personalized treatment. *PLoS One* 7:e31968.
- Johnson N, Speirs V, Curtin NJ, Hall AG (2008) A comparative study of genome-wide SNP, CGH microarray and protein expression analysis to explore genotypic and phenotypic mechanisms of acquired antiestrogen resistance in breast cancer. *Breast Cancer Res Treat* 111:55–63.
- Jones AV, Kreil S, Zoi K, et al (2005) Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders. *Blood* 106:2162–8.
- Kadyk LC, Hartwell LH (1992) Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae*. *Genetics* 132:387–402.
- Kalhor R, Tjong H, Jayathilaka N, et al (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol* 30:90–8.
- Kallioniemi A, Kallioniemi OP, Sudar D, et al (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258:818–21.
- Kawamata N, Ogawa S, Yamamoto G, et al (2008a) Genetic profiling of myeloproliferative disorders by single-nucleotide polymorphism

- oligonucleotide microarray. *Exp Hematol* 36:1471–1479.
- Kawamata N, Ogawa S, Zimmermann M, et al (2008b) Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* 111:776–84.
- Knudson a G (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68:820–3.
- Kölbl AC, Weigl D, Mulaw M, et al (2012) The radial nuclear positioning of genes correlates with features of megabase-sized chromatin domains. *Chromosome Res* 20:735–52.
- Korn WM, Yasutake T, Kuo WL, et al (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.
- Kosak ST, Groudine M (2004a) Form follows function: the genomic organization of cellular differentiation. *Genes Dev* 18:1371–1384.
- Kosak ST, Groudine M (2004b) Form follows function: the genomic organization of cellular differentiation. *Genes Dev* 18:1371–1384.
- Kotzot D (2008) Complex and segmental uniparental disomy updated. *J Med Genet* 45:545–56.
- Kralovics R, Guan Y, Prchal JT (2002) Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera. *Exp Hematol* 30:229–36.
- Kralovics R, Passamonti F, Buser AS, et al (2005) A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 352:1779–90.
- Kuiper RP, Schoenmakers EFPM, van Reijmersdal S V, et al (2007) High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. *Leukemia* 21:1258–66.
- Küpper K, Kölbl A, Biener D, et al (2007) Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma* 116:285–306.
- Kurashina K, Yamashita Y, Ueno T, et al (2008) Chromosome copy number analysis in screening for prognosis-related genomic regions in colorectal carcinoma. *Cancer Sci* 99:1835–40.
- Kurz A, Lampel S, Nickolenko JE, et al (1996) Active and inactive genes localize preferentially in the periphery of chromosome territories. *J Cell Biol* 135:1195–205.
- Lapunzina P, Monk D (2011) The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biol Cell* 103:303–317.

- Lehmann S, Ogawa S, Raynaud SD, et al (2008) Molecular allelokaryotyping of early-stage, untreated chronic lymphocytic leukemia. *Cancer* 112:1296–305.
- Lemaître C, Grabarz A, Tsouroula K, et al (2014) Nuclear position dictates DNA repair pathway choice. *Genes Dev* 28:2450–2463.
- Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* 396:643–9.
- Li A, Liu Y, Zhao Q, et al (2014) Genome-wide identification of somatic aberrations from paired normal-tumor samples. *PLoS One* 9:e87212.
- Lieberman-Aiden E, van Berkum NL, Williams L, et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–93.
- Liehr T (2010) Cytogenetic contribution to uniparental disomy (UPD). *Mol Cytogenet* 3:8.
- Lips EH, de Graaf EJ, Tollenaar RAEM, et al (2007) Single nucleotide polymorphism array analysis of chromosomal instability patterns discriminates rectal adenomas from carcinomas. *J Pathol* 212:269–77.
- Llorente B, Smith CE, Symington LS (2008) Break-induced replication: What is it and what is it for? *Cell Cycle* 7:859–864.
- Lo KC, Bailey D, Burkhardt T, et al (2008) Comprehensive analysis of loss of heterozygosity events in glioblastoma using the 100K SNP mapping arrays and comparison with copy number abnormalities defined by BAC array comparative genomic hybridization. *Genes Chromosomes Cancer* 47:221–37.
- Loh ML, Sakai DS, Flotho C, et al (2009) Mutations in CBL occur frequently in juvenile myelomonocytic leukemia. *Blood* 114:1859–63.
- Lothe RA, Peltomäki P, Meling GI, et al (1993) Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. *Cancer Res* 53:5849–52.
- Lupski JR (2015) Structural variation mutagenesis of the human genome: Impact on disease and evolution. *Environ Mol Mutagen* 56:419–436.
- Lynch HT, de la Chapelle A, Lynch HT, de la Chapelle A (2003) Hereditary colorectal cancer. *N Engl J Med* 348:919–32.
- Machiela MJ, Zhou W, Sampson JN, et al (2015) Characterization of Large Structural Genetic Mosaicism in Human Autosomes. *Am J Hum Genet* 96:487–497.
- Mahy NL, Perry PE, Bickmore WA (2002) Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol* 159:753–63.

- Makishima H, Cazzolli H, Szpurka H, et al (2009) Mutations of e3 ubiquitin ligase cbl family members constitute a novel common pathogenic lesion in myeloid malignancies. *J Clin Oncol* 27:6109–16.
- Makishima H, Maciejewski JP (2011) Pathogenesis and Consequences of Uniparental Disomy in Cancer. *Clin Cancer Res* 17:3913–3923.
- Mateos-Langerak J, Bohn M, de Leeuw W, et al (2009) Spatially confined folding of chromatin in the interphase nucleus. *Proc Natl Acad Sci U S A* 106:3812–7.
- Mayer R, Brero A, von Hase J, et al (2005) Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biol* 6:44.
- Mehta IS, Kulashreshtha M, Chakraborty S, et al (2013) Chromosome territories reposition during DNA damage-repair response. *Genome Biol* 14:R135.
- Meijer GA, Hermsen MA, Baak JP, et al (1998) Progression from colorectal adenoma to carcinoma is associated with non-random chromosomal gains as detected by comparative genomic hybridisation. *J Clin Pathol* 51:901–9.
- Melcher R, Al-Taie O, Kudlich T, et al (2007) SNP-Array genotyping and spectral karyotyping reveal uniparental disomy as early mutational event in MSS- and MSI-colorectal cancer cell lines. *Cytogenet Genome Res* 118:214–21.
- Melcher R, Hartmann E, Zopf W, et al (2011) LOH and copy neutral LOH (cnLOH) act as alternative mechanism in sporadic colorectal cancers with chromosomal and microsatellite instability. *Carcinogenesis* 32:636–42.
- Middeldorp A, van Puijenbroek M, Nielsen M, et al (2008) High frequency of copy-neutral LOH in MUTYH-associated polyposis carcinomas. *J Pathol* 216:25–31.
- Mohamedali A, Gäken J, Twine NA, et al (2007) Prevalence and prognostic significance of allelic imbalance by single-nucleotide polymorphism analysis in low-risk myelodysplastic syndromes. *Blood* 110:3365–73.
- Moore JK, Haber JE (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol* 16:2164–73.
- Morita T, Uzawa N, Mogushi K, et al (2016) Characterizing Genetic Transitions of Copy Number Alterations and Allelic Imbalances in Oral Tongue Carcinoma Metastasis. *Genes, Chromosom Cancer* 55:975–986.
- Mosimann C, Hausmann G, Basler K (2009) Beta-catenin hits chromatin: regulation of Wnt target gene activation. *Nat Rev Mol Cell Biol* 10:276–86.
- Moynahan ME, Jasin M (1997) Loss of heterozygosity induced by a chromosomal double-strand break. *Proc Natl Acad Sci U S A* 94:8988–93.
- Moynahan ME, Jasin M (2010) Mitotic homologous recombination maintains

- genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol* 11:196–207.
- Müller I, Boyle S, Singer RH, et al (2010) Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells. *PLoS One* 5:e11560.
- Mullighan CG, Goorha S, Radtke I, et al (2007) Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446:758–64.
- Murthy SK, DiFrancesco LM, Ogilvie RT, Demetrick DJ (2002) Loss of Heterozygosity Associated with Uniparental Disomy in Breast Carcinoma. *Mod Pathol* 15:1241–1250.
- Nagano T, Lubling Y, Stevens TJ, et al (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502:59–64.
- Nakao K, Mehta KR, Fridlyand J, et al (2004) High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis* 25:1345–57.
- Nancarrow DJ, Handoko HY, Smithers BM, et al (2008) Genome-wide copy number analysis in esophageal adenocarcinoma using high-density single-nucleotide polymorphism arrays. *Cancer Res* 68:4163–72.
- Neuwirth EAH, Honma M, Grosovsky AJ (2007) Interchromosomal crossover in human cells is associated with long gene conversion tracts. *Mol Cell Biol* 27:5261–74.
- Nicholson JM, Cimini D (2011) How mitotic errors contribute to karyotypic diversity in cancer. *Adv Cancer Res* 112:43–75.
- O’Keefe C, McDevitt MA, Maciejewski JP (2010) Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. *Blood* 115:2731–9.
- Olaharski AJ, Sotelo R, Solorza-Luna G, et al (2006) Tetraploidy and chromosomal instability are early events during cervical carcinogenesis. *Carcinogenesis* 27:337–43.
- Olshen AB, Bengtsson H, Neuvial P, et al (2011) Parent-specific copy number in paired tumor–normal studies using circular binary segmentation. *Bioinformatics* 27:2038–2046.
- Palles C, Cazier J-B, Howarth KM, et al (2012) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 45:136–144.
- Peiffer DA, Le JM, Steemers FJ, et al (2006) High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* 16:1136–48.
- Pfeifer D, Pantic M, Skatulla I, et al (2007) Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. *Blood*

- 109:1202–10.
- Pino MS, Chung DC (2010) The Chromosomal Instability Pathway in Colon Cancer. *Gastroenterology* 138:2059–2072.
- Platzer P, Upender MB, Wilson K, et al (2002) Silence of chromosomal amplifications in colon cancer. *Cancer Res* 62:1134–8.
- Pole JCM, Courtay-Cahen C, Garcia MJ, et al (2006) High-resolution analysis of chromosome rearrangements on 8p in breast, colon and pancreatic cancer reveals a complex pattern of loss, gain and translocation. *Oncogene* 25:5693–706.
- Rabl C (1885) Über Zelltheilung. *Morph Jb* 10:214–330.
- R Core Team (2017) R: A Language and Environment for Statistical Computing.
- Raghavan M, Lillington DM, Skoulakis S, et al (2005) Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res* 65:375–8.
- Rajaram M, Zhang J, Wang T, et al (2013a) Two Distinct Categories of Focal Deletions in Cancer Genomes. *PLoS One* 8:e66264.
- Rajaram M, Zhang J, Wang T, et al (2013b) Two Distinct Categories of Focal Deletions in Cancer Genomes. *PLoS One* 8:e66264.
- Ranade D, Koul S, Thompson J, et al (2017) Chromosomal aneuploidies induced upon Lamin B2 depletion are mislocalized in the interphase nucleus. *Chromosoma* 126:223–244.
- Ried T, Knutzen R, Steinbeck R, et al (1996) Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* 15:234–45.
- Rodriguez-Bigas MA, Boland CR, Hamilton SR, et al (1997) A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. *J Natl Cancer Inst* 89:1758–62.
- Rodríguez-Santiago B, Malats N, Rothman N, et al (2010) Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am J Hum Genet* 87:129–38.
- Rouquette J, Genoud C, Vazquez-Nin GH, et al (2009) Revealing the high-resolution three-dimensional network of chromatin and interchromatin space: a novel electron-microscopic approach to reconstructing nuclear architecture. *Chromosome Res* 17:801–10.
- Sakamoto K (2016) Notch signaling in oral squamous neoplasia. *Pathol Int* 66:609–617.

- Schreibman IR, Baker M, Amos C, McGarrity TJ (2005) The Hamartomatous Polyposis Syndromes: A Clinical and Molecular Review. *Am J Gastroenterol* 100:476–490.
- Segditsas S, Rowan AJ, Howarth K, et al (2009) APC and the three-hit hypothesis. *Oncogene* 28:146–55.
- Sehgal N, Seifert B, Ding H, et al (2016) Reorganization of the interchromosomal network during keratinocyte differentiation. *Chromosoma* 125:389–403.
- Serra E, Puig S, Otero D, et al (1997) Confirmation of a double-hit model for the NF1 gene in benign neurofibromas. *Am J Hum Genet* 61:512–9.
- Shackney SE, Smith CA, Miller BW, et al (1989) Model for the genetic evolution of human solid tumors. *Cancer Res* 49:3344–54.
- Shi Z-Z, Zhang Y-M, Shang L, et al (2012) Genomic profiling of rectal adenoma and carcinoma by array-based comparative genomic hybridization. *BMC Med Genomics* 5:52.
- Shuman C, Smith AC, Steele L, et al (2006) Constitutional UPD for chromosome 11p15 in individuals with isolated hemihyperplasia is associated with high tumor risk and occurs following assisted reproductive technologies. *Am J Med Genet A* 140:1497–503.
- Siegel RL, Miller KD, Jemal A (2016) Cancer statistics, 2016. *CA Cancer J Clin* 66:7–30.
- Siegel RL, Ward EM, Jemal A (2012) Trends in colorectal cancer incidence rates in the United States by tumor location and stage, 1992-2008. *Cancer Epidemiol Biomarkers Prev* 21:411–6.
- Silkworth WT, Nardi IK, Scholl LM, Cimini D (2009) Multipolar spindle pole coalescence is a major source of kinetochore mis-attachment and chromosome mis-segregation in cancer cells. *PLoS One* 4:e6564.
- Smith DI, Zhu Y, McAvoy S, Kuhn R (2006) Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett* 232:48–57.
- Stack SM, Brown DB, Dewey WC (1977) Visualization of interphase chromosomes. *J Cell Sci* 26:281–99.
- Steinemann D, Arning L, Praulich I, et al (2010) Mitotic recombination and compound-heterozygous mutations are predominant NF1-inactivating mechanisms in children with juvenile myelomonocytic leukemia and neurofibromatosis type 1. *Haematologica* 95:320–3.
- Stephens K, Weaver M, Leppig KA, et al (2006) Interstitial uniparental isodisomy at clustered breakpoint intervals is a frequent mechanism of NF1 inactivation in myeloid malignancies. *Blood* 108:1684–9.
- Storchova Z, Kuffer C (2008) The consequences of tetraploidy and aneuploidy. *J Cell Sci* 121:3859–3866.

- Storchova Z, Pellman D (2004) From polyploidy to aneuploidy, genome instability and cancer. *Nat Rev Mol Cell Biol* 5:45–54.
- Strickfaden H, Zunhammer A, van Koningsbruggen S, et al (2010) 4D Chromatin dynamics in cycling cells: Theodor Boveri's hypotheses revisited. *Nucleus* 1:284–297.
- Sulong S, Moorman A V, Irving JAE, et al (2009) A comprehensive analysis of the CDKN2A gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups. *Blood* 113:100–7.
- Sun HB, Shen J, Yokota H (2000) Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys J* 79:184–90.
- Suzuki M, Kato M, Yuyan C, et al (2008) Whole-genome profiling of chromosomal aberrations in hepatoblastoma using high-density single-nucleotide polymorphism genotyping microarrays. *Cancer Sci* 99:564–70.
- Tanabe H, Müller S, Neusser M, et al (2002) Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A* 99:4424–9.
- Teh M-T, Blaydon D, Chaplin T, et al (2005) Genomewide single nucleotide polymorphism microarray mapping in basal cell carcinomas unveils uniparental disomy as a key somatic event. *Cancer Res* 65:8597–603.
- Thompson DA, Desai MM, Murray AW (2006) Ploidy controls the success of mutators and nature of mutations during budding yeast evolution. *Curr Biol* 16:1581–90.
- Thompson SL, Bakhoun SF, Compton DA (2010) Mechanisms of chromosomal instability. *Curr Biol* 20:R285-95.
- Toma MI, Grosser M, Herr A, et al (2008) Loss of heterozygosity and copy number abnormality in clear cell renal cell carcinoma discovered by high-density affymetrix 10K single nucleotide polymorphism mapping array. *Neoplasia* 10:634–42.
- Torabi K, Miró R, Fernández-Jiménez N, et al (2015) Patterns of somatic uniparental disomy identify novel tumor suppressor genes in colorectal cancer. *Carcinogenesis* 36:1103–1110.
- Torre LA, Bray F, Siegel RL, et al (2015) Global cancer statistics, 2012. *CA Cancer J Clin* 65:87–108.
- Tuna M, Knuutila S, Mills GB (2009) Uniparental disomy in cancer. *Trends Mol Med* 15:120–8.
- Tyybäkinöja A, Elonen E, Vauhkonen H, et al (2008) Single nucleotide polymorphism microarray analysis of karyotypically normal acute myeloid leukemia reveals frequent copy number neutral loss of heterozygosity. *Haematologica* 93:631–2.

- Van Loo P, Nordgard SH, Lingjaerde OC, et al (2010) Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci* 107:16910–16915.
- van Puijenbroek M, Middeldorp A, Tops CMJ, et al (2008) Genome-wide copy neutral LOH is infrequent in familial and sporadic microsatellite unstable carcinomas. *Fam Cancer* 7:319–30.
- Vasen HF, Mecklin JP, Khan PM, Lynch HT (1991) The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum* 34:424–5.
- Vasen HF, Watson P, Mecklin JP, Lynch HT (1999) New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* 116:1453–6.
- Volpi E V, Chevret E, Jones T, et al (2000) Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci* 1565–76.
- Walker BA, Leone PE, Jenner MW, et al (2006) Integration of global SNP-based mapping and expression arrays reveals key regions, mechanisms, and genes important in the pathogenesis of multiple myeloma. *Blood* 108:1733–43.
- Walsh CS, Ogawa S, Scoles DR, et al (2008) Genome-wide loss of heterozygosity and uniparental disomy in BRCA1/2-associated ovarian carcinomas. *Clin Cancer Res* 14:7645–51.
- Wang K, Kan J, Yuen ST, et al (2011) Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* 43:1219–1223.
- Wang L, Fidler C, Nadig N, et al (2008) Genome-wide analysis of copy number changes and loss of heterozygosity in myelodysplastic syndrome with del(5q) using high-density single nucleotide polymorphism arrays. *Haematologica* 93:994–1000.
- Wang L, Wheeler DA, Prchal JT (2016) Acquired uniparental disomy of chromosome 9p in hematologic malignancies. *Exp Hematol* 44:644–652.
- Williams MJ, Werner B, Barnes CP, et al (2016) Identification of neutral tumor evolution across cancer types. *Nat Genet* 48:238–244.
- Williams RRE, Azuara V, Perry P, et al (2006) Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *J Cell Sci* 119:132–40.
- Williams RRE, Broad S, Sheer D, Ragoussis J (2002) Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp Cell Res* 272:163–75.
- Wu R-C, Wang T-L, Shih I-M (2014) The emerging roles of ARID1A in tumor suppression. *Cancer Biol Ther* 15:655–64.

- Yamamoto G, Nannya Y, Kato M, et al (2007) Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am J Hum Genet* 81:114–26.
- Zack TI, Schumacher SE, Carter SL, et al (2013a) Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 45:1134–1140.
- Zack TI, Schumacher SE, Carter SL, et al (2013b) Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 45:1134–1140.
- Zarzour P, Boelen L, Luciani F, et al (2015) Single nucleotide polymorphism array profiling identifies distinct chromosomal aberration patterns across colorectal adenomas and carcinomas. *Genes Chromosomes Cancer* 54:303–14.
- Zhang F, Gu W, Hurles ME, Lupski JR (2009) Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* 10:451–81.
- Zhao M, Kim P, Mitra R, et al (2016) TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res* 44:D1023–31.
- Zink D, Amaral MD, Englmann A, et al (2004) Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *J Cell Biol* 166:815–25.
- Zirbel RM, Mathieu UR, Kurz A, et al (1993) Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome Res* 1:93–106.
- Zorn C, Cremer C, Cremer T, Zimmer J (1979) Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. *Exp Cell Res* 124:111–9.

7 Annex

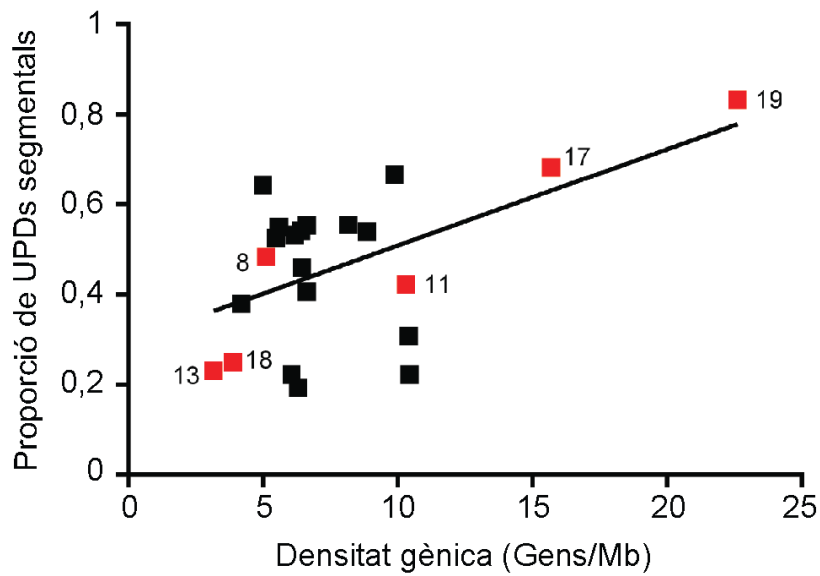


Figura 16. Proporció de UPDs segmentals per cromosoma. En vermell s'indiquen els cromosomes amb més densitat gènica (17 i 19) i amb menys densitat gènica (13 i 18). A més, també s'indiquen els cromosomes 8 i 11, els quals s'ha estudiat la seva distribució radial dins el nucli interfàsic en el tercer capítol d'aquesta tesi.

8 Agraïments

Abans de sentir el tret de sortida, sóc conscient que és una cursa dura, exigent i amb molt de desnivell. Una cursa que tot i haver-la entrenat molt, mai estàs del tot preparat per fer-la. De fet, molts amics em pregunten sovint "Per què corres curses tant llargues?" Els dubtes apareixen a la línia de sortida. Respiro profundament i penso en la preparació que he fet per arribar aquí...

Per començar m'agradaria agrair el suport dels meus primers companys d'entrenaments; Javi, Alba, Jordi R, Agustín, Sergio, Gemma D, Laia, Rita, Marta, Paula, Kristin, Pi, Anna D, Vicenç, Àngels, M^a Àngels, Ana, Quima, Jordi B, Carme, Cristina, Alberto i tothom que ha passat per la Unitat en aquest temps. Però, per córrer aquesta cursa és necessari ser més complert i entrenar en tot tipus de terrenys i condicions. Per això en els meus viatges entre el Vallès i Barcelona vull agrair als meus companys d'entrenament de l'altra banda de la serra de Collserola; Eva, Isa, Clàudia, Elena A, Martina G, Queralt, Marcos, Sebas, Paula, Clara E, Jenny, Esther, Irene, Manuel, Elena V, María, Saray, Lorena, Mireia, Coral, Laia, Maria V, Guillaume, Renata, Pau, Juanjo, Sergi, Txell, Antoni C, Maribel, Anna B i Maria C.

Però tard o d'hora, sé que l'entrenament recurrent en territori conegut porta a la confiança excessiva. Conèixer el camí és bo, però conèixer on estan posades cadascuna de les pedres pot afectar negativament la progressió d'un corredor. En aquest sentit només puc agrair a tota la gent que treballa al laboratori del Dr. Thomas Ried perquè córrer en territori desconegut mai havia sigut tant familiar. Gràcies especialment a tu Thomas per ser l'estàndard d'aquesta manera d'adoptar corredors foranis. Gràcies Dara per ensenyar-me tant. I moltes gràcies David, Rüdiger i Laura per córrer amb mi durant aquells mesos. Danke!

Capítol a part mereixen els meus tres entrenadors oficials. Gràcies Jordi, Rosa i Imma per tot l'esforç dedicat durant aquest temps a ensenyar-me a córrer amb bona tècnica. Sobretot pel tipus de corredor que acostumo a ser, poc precís i amb altes probabilitats de lesionar-me per córrer com no és degut. Em reitero, gràcies!

Però per córrer una cursa com aquesta és necessari molt més que físic i bona tècnica. Els constants canvis de desnivell, el córrer de nit, les relliscades, el dolor... Per sobreposar-te a tots aquests factors cal una ment forta i positiva. En aquest punt haig d'agrair a tota la gent que en algun moment de la cursa m'ha avituallat, animat, o fins i tot ha corregut amb mi una estona. Gràcies especialment al grup de seguidors incondicionals de l'esplai, al grup de biòlegs de la UAB, al Superhullahop Team i als amics del barri de Gràcia.

Especial menció d'agraïment mereixen els organitzadors, aquells que m'han proporcionat tot el necessari per realitzar la cursa en bones condicions. Gràcies especialment als meus pares però també a la resta de família.

I finalment vull agrair a les meves llebres, les que fan que el camí sigui més fàcil, més planer i més emocionant. Gràcies Clara, Albert i Guifré. I a tu Martina per córrer amb mi els trams més difícils de la cursa.

Tot just arribo a meta i és en aquest precís instant d'extrema realització que sé perfectament com respondre a la pregunta.

- Corro perquè fer-ho tan ben acompanyat és un plaer.

Moltes gràcies a tots!