

Treball Final de Màster
Màster de Biomedicina - Universitat de Barcelona

Anàlisi quantitativa de la disomia uniparental en 1179 càncers gastrointestinals

Autor:	Pau Erola Cañellas pauerola@gmail.com +34 652 530 578
Director:	Dr. Juan José Lozano Salvatella juanjo.lozano@ciberehd.org
Grup de recerca:	Plataforma de Bioinformàtica CIBEREHD Centre Esther Koplovitz C/ Rosselló 153 08036 Barcelona

Barcelona, 9 de juliol de 2015

Taula de continguts

1. Introducció	1
2. Objectius	3
3. Metodologia i materials	4
3.1. Mostres de càncer gastrointestinal	4
3.2. Anàlisi <i>genome-wide</i> de CNA i LOH	4
3.3. Detecció de segments significatius i TSG	5
4. Resultats	6
5. Discussió i conclusions	16
Bibliografia	18

1. Introducció

Els càncers gastrointestinals són aquell grup de càncers que afecten el sistema digestiu, incloent el càncer d'esòfag, vesícula biliar, fetge, pàncrees, estómac, intestí prim, còlon, recte i anus. A Catalunya, la incidència d'aquest grup de càncers els situa com els tumors més freqüents en homes, i els segons en freqüència en dones després del càncer de mama [23], tenint una supervivència a 5 anys que va del 53% en el càncer colorectal, al 27% en estómac, 13% esòfag, 10% en càncer hepàtic i només del 5% en pàncrees [8]. S'espera que aquest últim sigui el segon tumor maligne més mortal als EUA el 2020 [14].

L'oncogènesi és un procés evolutiu guiat per l'adquisició somàtica d'alteracions genètiques, incloent substitucions de parells de bases (SNP), translocacions, infeccions i alteracions en el número de còpies (*copy-number alterations*, CNA). La majoria dels tumors gastrointestinals són d'origen epitelial i tenen com a característica comuna patrons d'alta freqüència de CNA, que poden amplificar regions amb oncogens o produir delecions de gens supressors de tumors (*tumor supressor genes*, TSG) [7]. Identificar quines d'aquestes alteracions del DNA són funcionals, en el sentit de contribuir a l'avantatge selectiu de les cèl·lules tumorals, de les que son adquirides de forma aleatòria per un ràtio d'alteracions de *background* o degut a una major inestabilitat del locus, és una dels questions clau en *genome-wide analysis*.

En aquest sentit, Beroukhim et al. [7] van estudiar en 3131 mostres de 26 tipus de càncer diferents regions de CNA freqüents, trobant 158 regions focals recurrents. A [36], l'estudi de 4934 va identificar un total de 140 regions de *somatic CNA* (SCNA) freqüents, i va observar una duplicació de tot el genoma en un 37% dels casos. Dulak et al. [11] també van buscar recurrències de SCNA en 486 mostres de càncer esofàgic, gàstric i de colon, trobant 64 regions recurrents d'amplificació o delecio.

A [9] van caracteritzar les alteracions somàtiques en *copy number* (CN), metilació de promotors, mRNA i microRNA, de càncers de colon i recte. L'estudi va trobar 32 gens mutats de forma recurrent, incloent APC, TP53, KRAS, PIK3CA i SMAD4. Hoadley et al. [16] van presentar un estudi integratiu per a 12 tipus de càncer, fent servir dades de *whole-exome sequencing*, metilació, mRNA, microRNA i nivells de proteïna. Van trobar freqüències de CNA superiors al 10% per a TP53, PIK3CA i PTEN, i l'anàlisi de *clustering* va establir una correlació entre els perfils de CNA i el teixit d'origen.

Ciriello et al. [10] van integrar dades d'alteracions en CN, metilació del DNA i mutacions

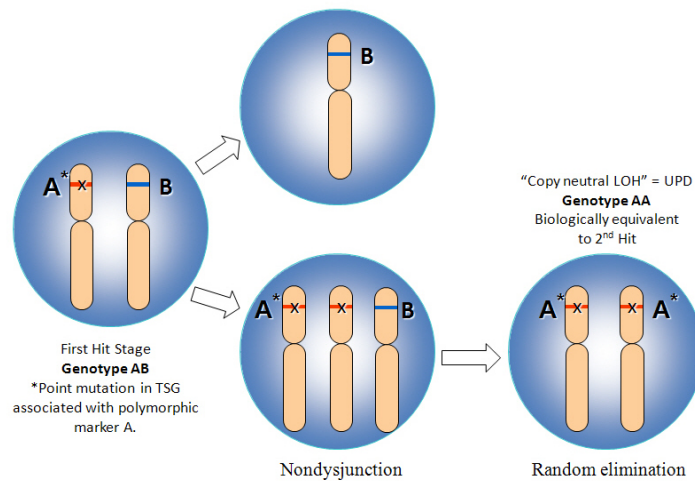


Figura 1. Esquema d'un possible mecanisme de formació somàtica d'UPD. La disomia uniparental és una forma de pèrdua d'heterozigositat (*loss of heterozygosity*, LOH) neutral en número de còpies. Els efectes biològics de la UPD són importants ja que les mutacions recessives deixen d'estar emmascarades per les dominants normals, i aquest esdeveniment pot actuar som a segon *hit* segons el model de Knudson. Reproduït de [35].

somàtiques de 3299 mostres de 12 tipus de càncer. Amb aquestes dades van proposar una classificació independent del teixit d'origen dels tumors en base a les alteracions genètiques i epigenètiques, que sorprenentment troba semblances en els perfils de mutació i CNA.

En el cas de TSG, un mecanisme alternatiu a la deleció que pot provocar la inactivació bial·lèlica del gen és la disomia uniparental (*uniparental disomy*, UPD). La UPD es produeix quan les dues còpies d'un cromosoma provenen del mateix pare, per errors en la meiosi, o pot ser somàtica, adquirida per errors mitòtics en la divisió cel·lular, com és descrit a la Figura 1. Al tractar-se d'alteracions neutrals en número de còpies, per detectar-les cal estudiar conjuntament CNA amb genotipatge de SNP.

Alguns estudis han intentat identificar les regions recurrents d'UPD. Lindbjerg et al. [3] van estudiar el perfil d'UPD en càncer colorectal trobant uns patrons homogenis en les regions de LOH (incloent 5q, 8p, 18p, 18q i 20p), fet que ho atribueixen a que possiblement aquestes alteracions són primerenques en el desenvolupament del tumor. Més recentment, Torabi et al. [32] van identificar patrons d'UPD en dues cohorts independents de càncer colorectal, trobant perfils similars entre les deleccions de CN i les regions d'UPD. Fins al moment però, no hem s'ha fet cap ampli estudi *genome-wide* com en el cas de CNA. Per això, en aquest treball proposem estudiar els perfils d'UPD en càncers gastrointestinals.

2. Objectius

En aquest treball volem construir el perfil de segments *genome-wide* de CNA i UPD d'un conjunts ampli de mostres de càncer gastrointestinal amb els següents dos objectius:

- Caracteritzar els perfils de CNA i UPD per teixit i de forma global per identificar regions genòmiques on hi ha una alta freqüència d'alteracions.
- Identificar les regions amb UPD que tinguin significança estadística i, per tant, que puguin ser susceptibles de contenir gens supressors de tumors. La identificació de TSG coneguts ens permetrà validar el nostre *pipeline*.

3. Metodologia i materials

3.1. Mostres de càncer gastrointestinal

Per aquest estudi es va recollir un total de 1461 parells de mostres de teixit tumorós i normal (*Level 1 Affymetrix SNP6 arrays*) del portal de dades de The Cancer Genome Atlas (TCGA, <https://tcga-data.nci.nih.gov/tcga/>) per a càncer de colon, recte, esòfag, estómac, fetge i pàncrees. Després del *pipeline* de detecció de CN i UPD vam obtenir 1179 perfils genòmics. La Taula 1 resumeix els detalls de la mostra.

3.2. Anàlisi *genome-wide* de CNA i LOH

Per a fer l'anàlisi de d'alteracions en CN i LOH vam fer servir el llenguatge de programació R [29] (versió 3.2.1).

El preprocessament dels microarrays d'Affymetrix es va fer amb la llibreria *aroma.affymetrix* (versió 2.13.2) [5] que implementa el mètode CRMAv2 per a l'estimació a màxima resolució, calibrada i normalitzada, del nombre de còpies *raw* del genoma.

Després vam fer servir el paquet *PSCBS* [4,25] (versió 0.44.0) per a trobar aquelles regions que presentaven una variació significativa en el número de còpies i en pèrdua d'heterozigositat. *PSCBS* va ser dissenyat per a treballar amb dades de microarrays de SNP, i implementa un algorisme de segmentació circular binària per a mostres aparellades que particiona el genoma en regions amb un mateix nombre de CN. També identifica les regions amb pèrdua d'heterozigositat fent servir un algorisme de mostreig aleatori o *bootstrapping*. És important destacar que *PSCBS* ens permet quantificar el *parent-specific copy number* (PSCN), retornant uns llindars específics per a cada parell de mostres, basats en quantils, que determinen les *calls* d'amplificació o deleció en CN i LOH.

Un cop segmentades les mostres, es va aplicar un filtre seleccionant només les mostres que presentaven menys de 300 segments per descartar-ne aquelles amb alta inestabilitat genòmica, deguda l'evolució tumoral o a altres causes com una possible inestabilitat de microsatèl·lits (MMR).

Per determinar si hi havia diferències significatives en el nombre de segments de CN i UPD entre grups es va fer servir una prova t-Student.

Cancer Types	Acronym	Clinical Annot.	Available	Discarded		Final
				PSCBS	>300 seg.	
Colon adenocarcinoma	COAD		354	12	56	286
Esophageal carcinoma	ESCA		*185			
		adenoma	90	1	31	58
		squamous	91		31	60
Liver hepatocellular carcinoma	LIHC		325	131	33	161
Pancreatic adenocarcinoma	PAAD		184	5	23	156
Rectum adenocarcinoma	READ		156	9	28	119
Stomach adenocarcinoma	STAD		442	11	92	339

Taula 1. Resum de les mostres utilitzades de càncers gastrointestinals. De totes les mostres descarregades del portal de TCGA, es van descartar els parells de mostres que no van finalitzar la detecció de segments PSCBS en menys de 8 hores (degut al procés de *bootstrapping*), i aquelles que presentaven més de 300 segments.

* No es van incloure 4 mostres per falta d'informació clínica.

3.3. Detecció de segments significatius i TSG

Per detectar els segments amb UPD més significatius dins de cada tipologia de càncer, vam implementar un test de permutacions al llarg de cada genoma per descartar aquelles alteracions que podien ser fruit de l'atzar o de l'acumulació de *background* (hipòtesi nul·la). Es van realitzar 10^9 permutacions per mostra.

Amb aquest test vam obtenir un *p-value* que a l'estar realitzant múltiples contrastos era susceptible de falsos positius. Per a corregir aquest efecte vam establir el llindar de significança amb la correcció per FDR (*false discovery rate*) de Hochberg [6].

Per identificar TSG dins de les regions significatives d'UPD vam considerar els gens obtinguts de la base de dades TSGene [37] (<http://bioinfo.mc.vanderbilt.edu/TSGene/>).

4. Resultats

Hem analitzat el número de còpies i la presència de disomies uniparentals adquirides en 1179 càncers gastrointestinals. La Figura 2 presenta els perfils d'aquestes alteracions genòmiques per a cada mostra, agrupades en 7 tipologies diferents (Taula 1), i les seves freqüències.

Els grups ESAD, ESSQ, LIHC i READ presenten amplificacions o delecions en un 93,1–98,3% dels casos, i UPD entre un 88,8–93,3%. COAD i STAD mostren guanys o pèrdues en número de còpies en un 88,2–91,1% del casos, i UPD en un 89,8% i 82,5% respectivament. PAAD presenta una freqüència d'alteracions significativament més baixa ($P < 0,001$) amb un 80,1% per a amplificacions, 76,2% per a delecions, i 68,5% per a UPD. El nombre de mostres sense cap aberració detectada és de 2 per al grup de COAD, 3 per a LIHC, 7 en STAD i 16 casos en PAAD.

Pel que fa a la mida de les regions afectades per CNA i UPD podem diferenciar entre esdeveniments que afecten tot un braç cromosòmic o alteracions focals. La Figura 3 mostra el nombre d'esdeveniments a nivell de braç en cada grup separat per guanys, pèrdues i UPD. La comparació entre alteracions mostra que hi ha diferències significatives ($P < 0,01$) entre guanys (LIHC $3,08 \pm 2,66$; COAD $3,14 \pm 2,86$, READ $4,28 \pm 3,42$; ESSQ $4,6 \pm 3,68$; STAD $3,80 \pm 3,68$; PAAD $1,65 \pm 2,09$) i UPD (LIHC $1,06 \pm 2,08$; COAD $2,10 \pm 2,57$, READ $4,77 \pm 2,47$; ESSQ $3,74 \pm 2,80$; STAD $1,85 \pm 2,75$; PAAD $0,79 \pm 2,12$), i pèrdues (LIHC $4,04 \pm 2,94$; COAD $2,79 \pm 3,03$, READ $3,67 \pm 3,21$; ESSQ $5,83 \pm 4,35$; STAD $3,79 \pm 4,37$; PAAD $4,01 \pm 3,84$) i UPD, per a tots els grups excepte ESAD ($5,07 \pm 4,19$, $5,26 \pm 4,52$, $3,74 \pm 2,80$ per a guanys, pèrdues i UPD respectivament). Així, la freqüència d'UPD es mostra menor en casi tots els casos en comparació amb les alteracions en número de còpia.

Si comparem el nombre d'esdeveniments d'UPD a nivell de braç entre grups, no trobem diferències estadísticament significatives entre LIHC i PAAD ($P = 0,1314$), COAD i ESSQ ($P = 0,0729$), COAD i STAD ($P = 0,1915$), READ i ESAD/ESSQ ($P = 0,582$, $P = 0,8948$), ni ESAD i ESSQ ($P = 0,1192$).

En les alteracions focals, a PAAD ($1,78 \pm 2,09$) el nombre d'esdeveniment és més petit que a la resta de grups ($P < 0,001$), i a ESAD ($4,32 \pm 4,19$) el nombre d'esdeveniments és major. READ ($3,36 \pm 3,34$) presenta una freqüència d'UPD més alta ($P < 0,05$) que LIHC ($2,65 \pm 2,66$) i COAD ($2,78 \pm 2,86$). Aquests resultats ens indiquen que la UPD és més prevalent al càncer esofàgic adenomatós en consonància a les majors pèrdues ja conegudes [11].

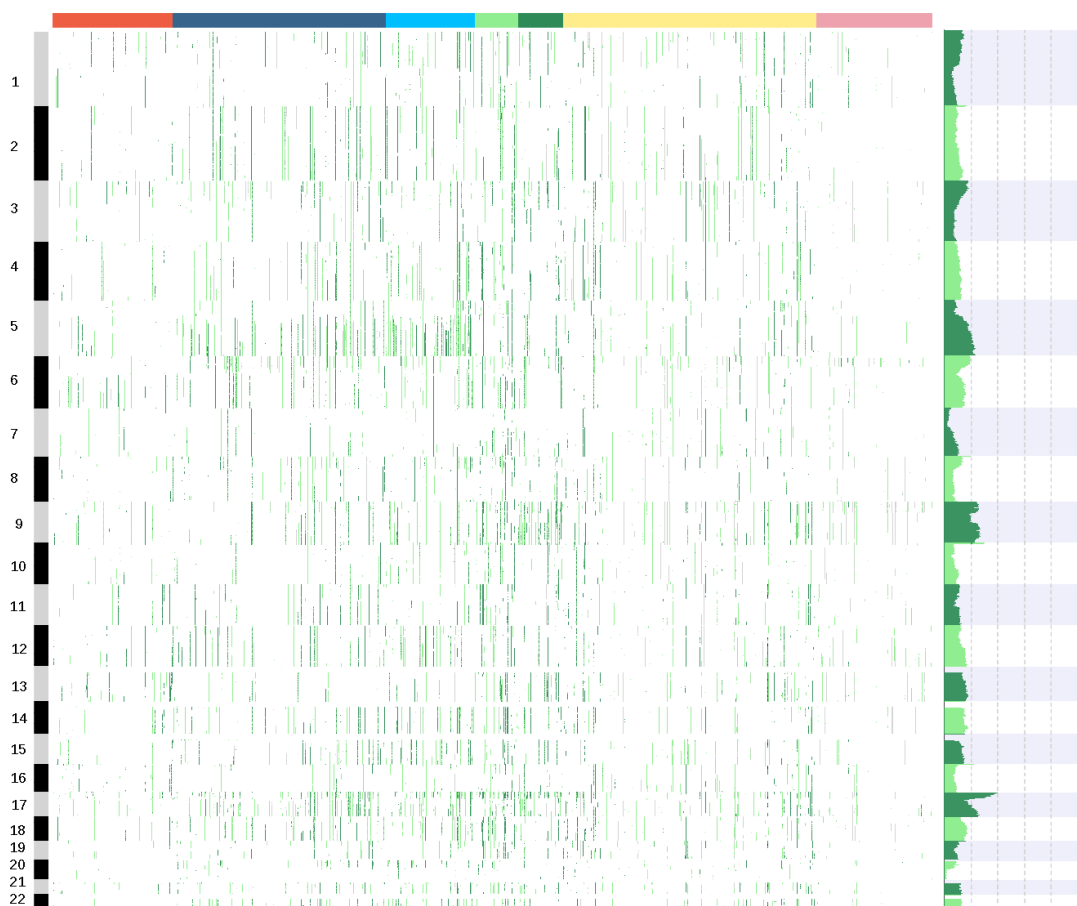
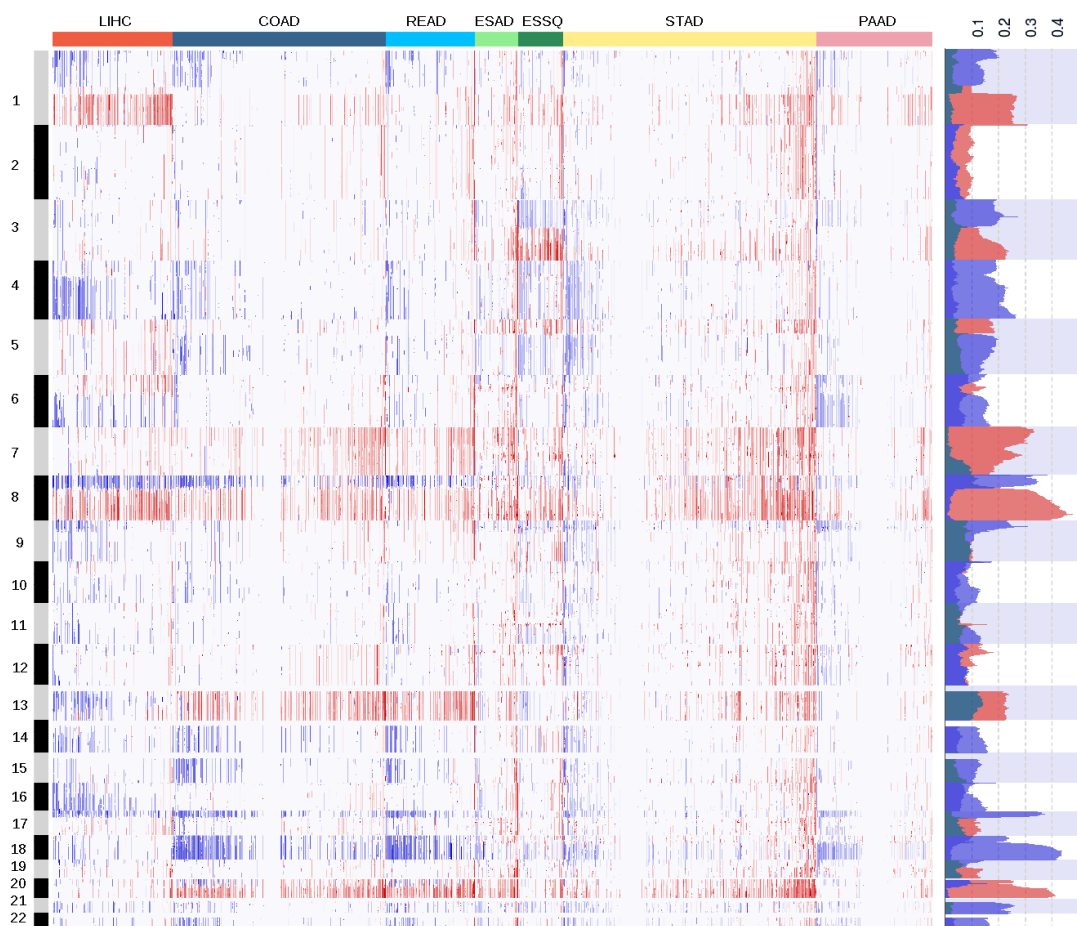


Figura 2. Perfil d'alteracions per a cada mostra (eix x) en funció de la seva posició genòmica (eix y). A dalt, mapa de CNA (vermell per a amplificacions, blau per a delecions) i freqüència mitjana de cada tipus d'alteració. Les mostres s'han ordenat dins de cada grup en funció del valor de CN ($\log_2 \text{ratio}$). A baix, mapa d'UPD i freqüència mitjana. Colors més intensos indiquen major alteració del nombre de còpies o una pèrdua heterozigositat més alta.

Les regions genòmiques que mostren una freqüència mitjana d'UPD superior al 10% es troben a 5q i els cromosomes 9 i 17. En aquestes mateixes regions també podem observar freqüències de pèrdua de CN que arriben al 20% a 5q, i al 30% a 9p i 17p. És conegut que aquestes zones contenen TSG com APC a 5q [32], CDKN2 a 9p [15] i TP53 a 17p. A la Taula 2 mostrem una representació de coneguts TSG [37] ordenats per major freqüència global d'UPD. Aquesta relació entre les regions amb delecions i regions UPD la podem observar també a la Figura 4 que mostra la correlació de les freqüències globals de pèrdua i UPD per tots els gens de *HG19* per cromosoma. Podem observar casos en que les correlacions són molt altes com en els cromosomes 1, 7, 8, 10, 11 i 20. En altres casos amb valors de r més baixos, podem observar definits subgrups que mostrarien una correlació més alta per separat, com en els cromosomes 17 i 19. Aquest resultat podria apuntar a que ambdues aberracions donen fruit a avantatges evolutius funcionalment semblants.

Tot i així, és optimista esperar trobar uns oncogens o TSG comuns i funcionalment rellevants dins de l'heterogeneïtat del càncer [11]. Per a estudiar amb més detall les aberracions genòmiques, la Figura 5 integra els perfils de guany i pèrdua de CN i UPD en gràfiques *circos* per grup. A més, mostra els segments d'UPD estadísticament significatius trobats amb l'anàlisi de permutacions i senyala alguns TSG coneguts dins d'aquestes regions.

A COAD s'ha trobat regions d'UPD significatives a nivell de braç amb freqüències superiors al 15% a 5q i el cromosoma 17, i al voltant del 10% a 6p, 18q i 22q. També presenta alteracions focals a 2q, 3p (FHIT), 6q (PARK2), 12q, 14q. Dos candidats no identificats a la llista de TSG [37] es troben a 16p, A2BP1 (12,9%), i a 20p, C20orf133 (19,9%), ambdós casos amb una freqüència de pèrdues de CN superior al 20%. No obstant, tot i haver-se observat amb anterioritat la seva alta freqüència de delecions, no s'ha pogut establir el seu rol de supressor tumoral [30].

El perfil de READ mostra una altra freqüència d'UPD. Trobem alteracions en un 30% dels casos a 5q, d'un 22% a 17p i entre un 10–15% a 4q, 8p, 9q, 14q, 15q, 20p, 22q i als cromosomes 6, 12 i 18. En alguns casos, com 5q, la freqüència d'UPD és superior a la de pèrdua de CN. L'anàlisi de permutacions ens destaca com a regions significatives 4q i 5q,

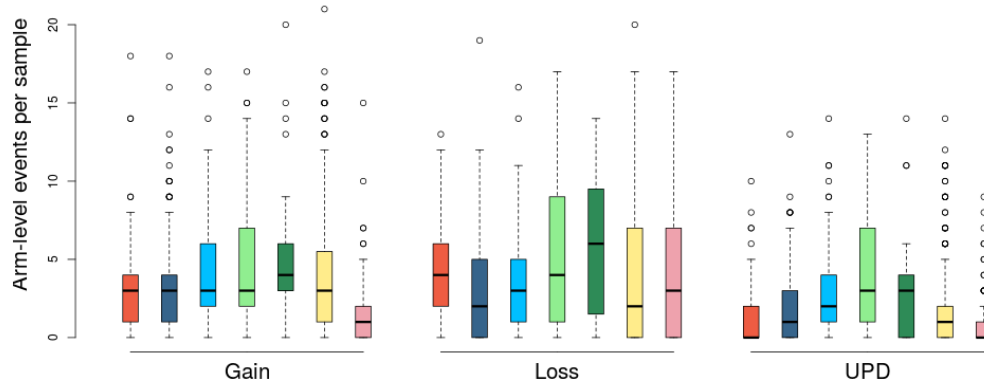


Figura 3. Nombre de CNA i UPD a nivell de braç cromosòmic per mostra. Els colors de cada grup corresponen als de la Figura 2.

Gene	Chr. arm	Position HG19		UPD	Loss	Gain	UPD per cancer						
		Start	End				LI	CO	RE	EA	ES	ST	PA
HIC1	17p	1.96	1.96	.204	.345	.022	.068	.171	.244	.431	.367	.118	.032
DPH1	17p	1.93	1.95	.204	.345	.022	.068	.171	.244	.431	.367	.118	.032
DAPK1	9q	90.11	90.32	.139	.094	.092	.043	.052	.126	.207	.400	.080	.064
GAS1	9q	89.56	89.56	.135	.095	.091	.043	.052	.118	.207	.383	.080	.064
SOCS3	17q	76.35	76.36	.131	.075	.114	.031	.164	.160	.138	.267	.094	.064
CYGB	17q	74.52	74.53	.131	.078	.112	.025	.164	.160	.155	.250	.097	.064
PTPRD	9p	8.32	9.02	.128	.248	.083	.031	.056	.092	.207	.250	.103	.154
TOPORS	9p	32.54	32.55	.124	.180	.085	.037	.063	.092	.190	.267	.094	.122
SCGB3A1	5q	180.02	180.02	.114	.126	.075	.031	.171	.319	.086	.117	.056	.019
FHIT	3p	59.74	61.24	.114	.397	.040	.043	.115	.160	.241	.100	.118	.019
DOK3	5q	176.93	176.94	.113	.127	.074	.031	.171	.311	.086	.117	.056	.019
PARK2	6q	161.77	163.15	.100	.213	.072	.075	.115	.160	.190	.100	.047	.013
TFAP2A	6p	10.40	10.42	.096	.141	.088	.012	.136	.101	.155	.167	.088	.013
FOXC1	6p	1.61	1.61	.096	.155	.088	.019	.136	.109	.155	.150	.088	.013
RARB	3p	25.47	25.64	.091	.184	.041	.062	.094	.084	.172	.117	.083	.026
WWOX	16p	78.13	79.25	.089	.252	.075	.037	.059	.084	.276	.050	.100	.013
DCC	18q	49.87	51.06	.086	.440	.021	.043	.105	.126	.172	.067	.071	.019
TCF4	18q	52.89	53.26	.086	.435	.021	.050	.105	.126	.172	.067	.062	.019
CHFR	12q	133.42	133.46	.085	.084	.084	.037	.098	.118	.138	.067	.059	.077
SPRY2	13q	80.91	80.92	.084	.122	.229	.062	.035	.050	.172	.150	.062	.058
GPR68	14q	91.70	91.72	.084	.151	.043	.037	.098	.134	.172	.067	.053	.026
RB1	13q	48.88	49.06	.084	.139	.228	.056	.031	.059	.172	.167	.050	.051
CDK2AP1	12q	123.75	123.76	.082	.075	.086	.043	.091	.118	.121	.067	.059	.077
PTPRK	6q	128.29	128.84	.080	.150	.069	.106	.084	.109	.121	.083	.041	.019
ARID1A	1p	27.02	27.11	.078	.213	.018	.037	.063	.101	.121	.117	.088	.019
MEG3	14q	101.29	101.33	.077	.157	.046	.037	.098	.109	.155	.067	.050	.026
SFN	1p	27.19	27.19	.076	.211	.020	.043	.063	.101	.121	.100	.086	.019
ERBB4	2q	212.25	213.40	.076	.062	.088	.037	.084	.076	.138	.117	.071	.006
FES	15q	91.43	91.44	.074	.097	.081	.025	.070	.168	.069	.100	.059	.026
BLM	15q	91.26	91.36	.074	.097	.082	.025	.070	.168	.069	.100	.059	.026

Taula 2. Llista de TSG en regions UPD significatives. S'ha limitat la llista a 2 gens per braç cromosòmic, escollint sempre els de major freqüència mitjana d'UPD.

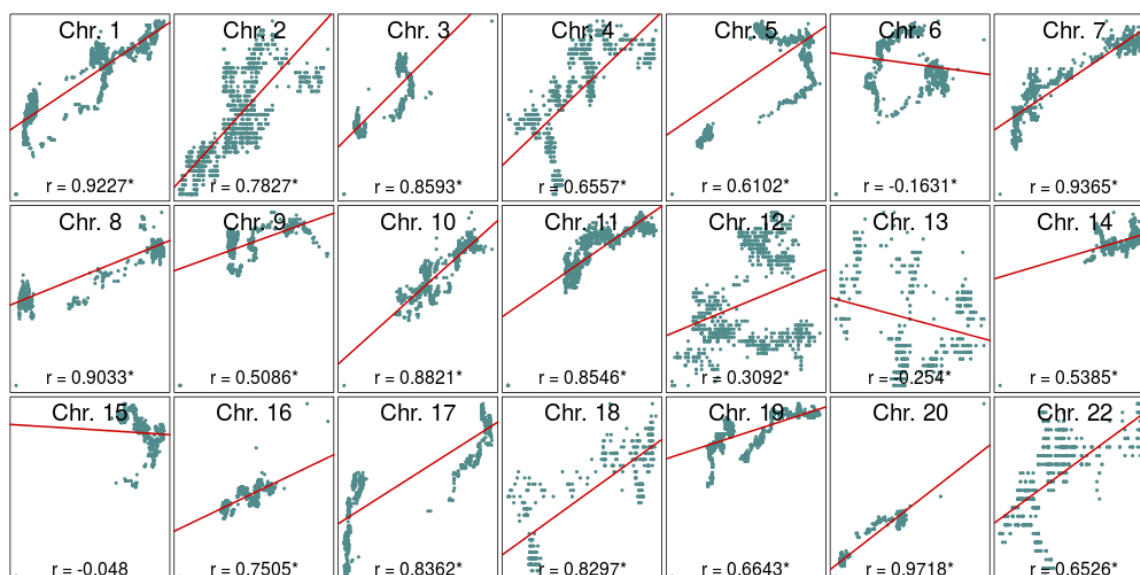
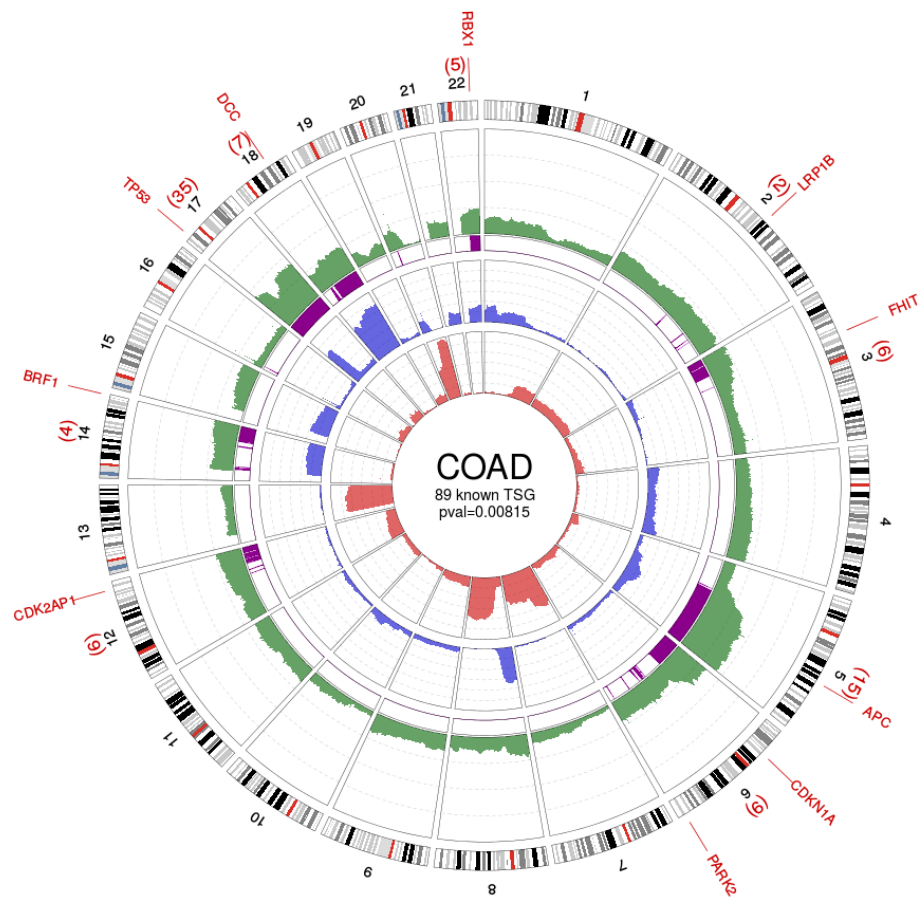


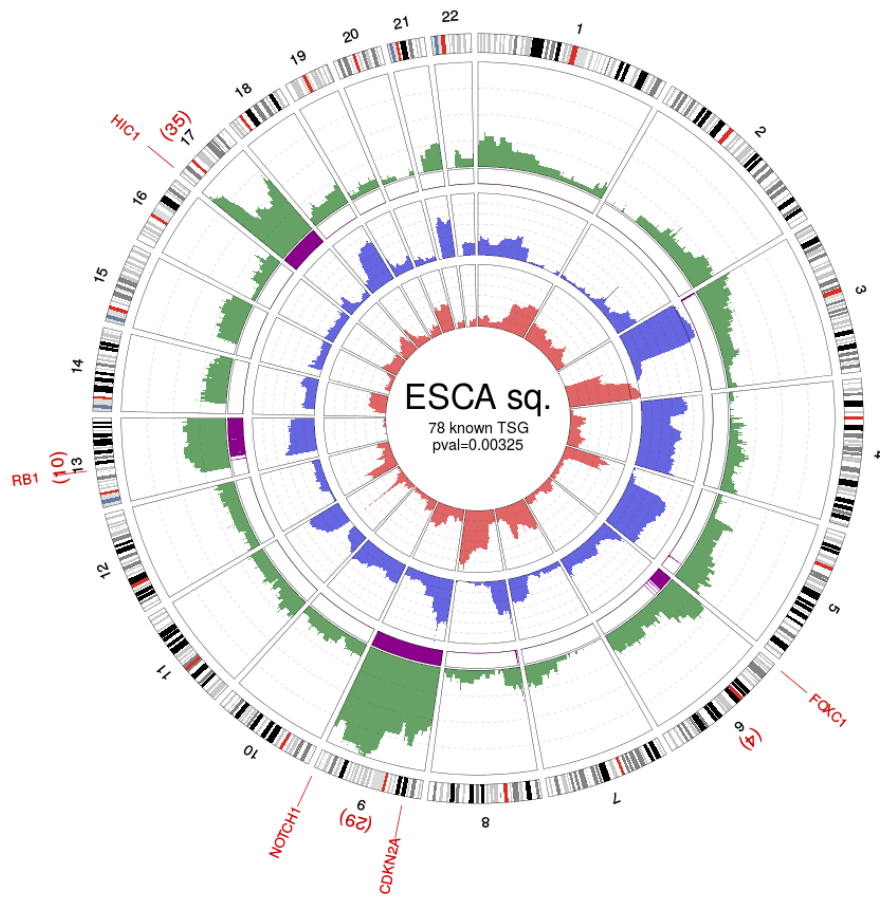
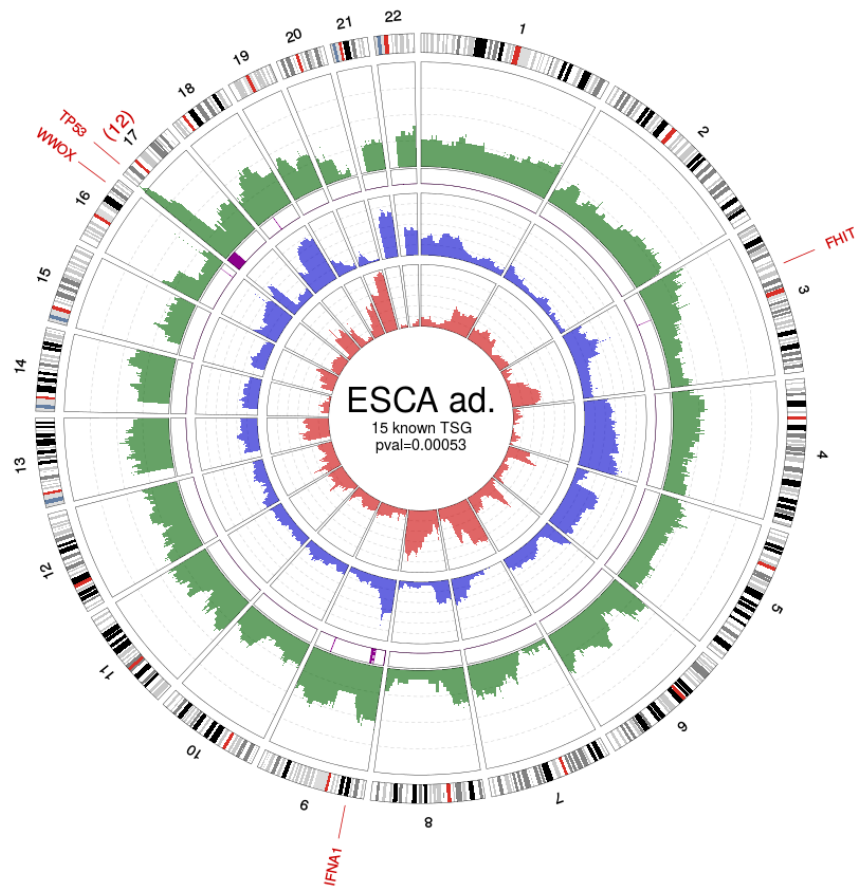
Figura 4. Correlació entre les freqüències de deleccions de CN (eix x) i d'UPD (eix y) per cada cromosoma (excepte el 21). * $P < 0.001$

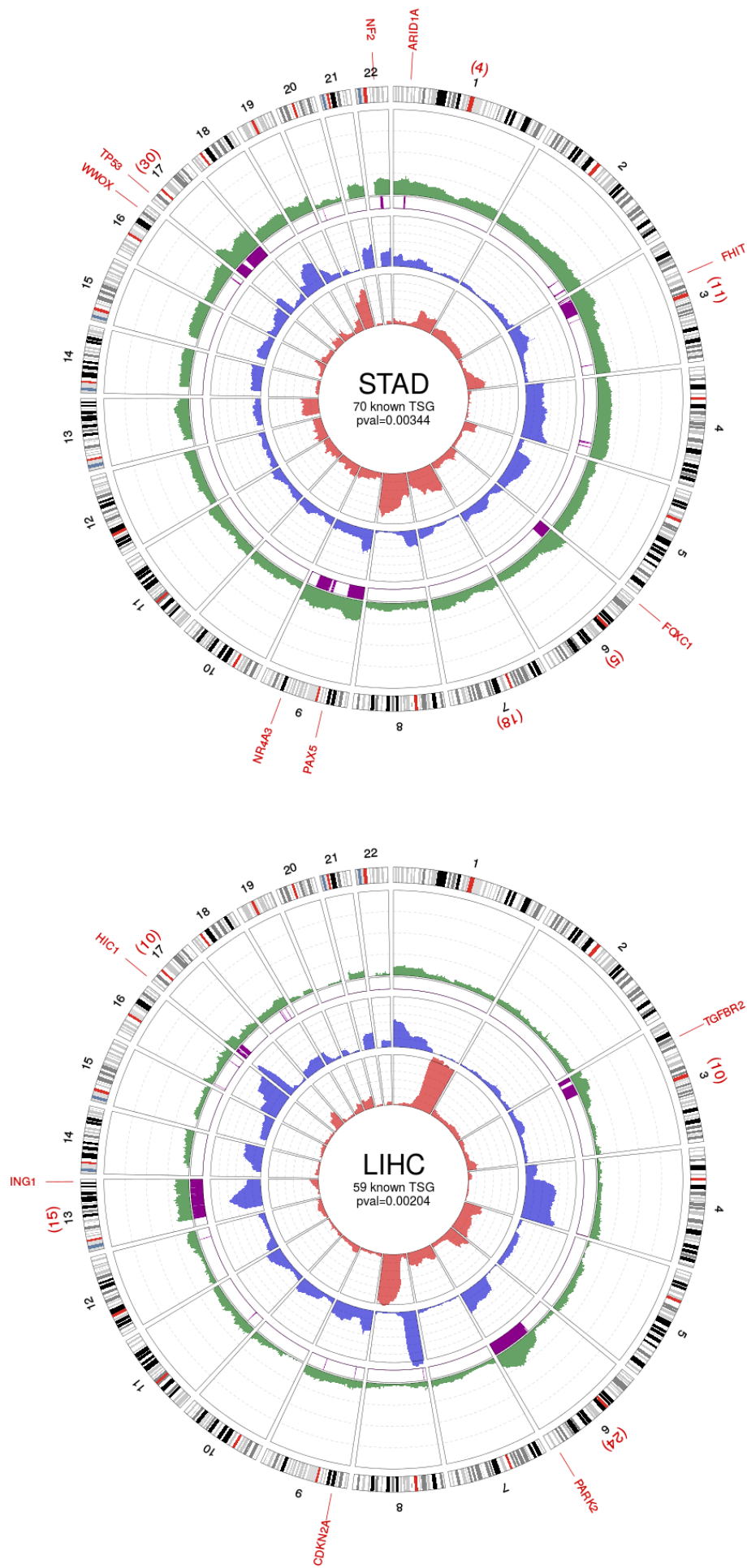
15q i el cromosoma 17, a part de regions de menor longitud. Com apunta Torabi et al. [32], és d'esperar trobar alteracions semblants a COAD al cromosoma 17 (que conté varis TSG coneguts com TP53) i a 5q (que conté APC). Alguns TSG focals com FHIT (15,9%), PARK2 (15,9%), A2BP1 i C20orf133 també son recurrents. A 8p trobem una petit *peak* significatiu de 3 gens que conté DLC1, que ha demostrat el seu paper de supressor tumoral en càncer colorectal [26] i hepàtic [38]. A 18q destaquem CABLES1, que s'ha trobat en delecció en càncers de colon i té un rol important en la progressió tumoral [21].

Els resultats d'ESAD mostren una alta freqüència d'UPD, amb una mitjana al voltant del 10% de freqüència a tot el genome, que arriba al 20% al cromosoma 9 i a un 40% a 17q. Amb aquesta alta variabilitat, l'anàlisi de permutacions només ens destaca un event a nivell de braç a 17q, on trobem una petita regió amb 13 gens amb una freqüència d'UPD del 43,1%, entre ells HIC1. Trobem també TSG significatius a 3p FHIT (24,1%), com en COAD i READ, a 9p IFNA1 (24,1%) i a 16q WWOX (27,8%). No identificats dins de la nostra base de dades de TSG, és significatiu DBC1 a 9q (20,1%) que recentment s'ha demostrat que regula l'estabilitat de p53 en càncer de mama [28]. També observem a 18q MAPRE2 (22,4%) que ha sigut associat a la invasivitat en càncer pancreàtic [1].

ESSQ també presenta un ràtio d'UPD important al cromosoma 9 (20–40%) i al cromosoma 17 (18–35%), i per sobre del 10% a 6p i 13q. A l'igual que en ESAD, la major freqüència al cromosoma 17 és al gen HIC1 (36,7%). A 9p trobem CDKN2A (15,0%) que s'ha demostrat que té una alta freqüència de pèrdua en carcinoma escatós de bufeta [15]. D'altra banda,







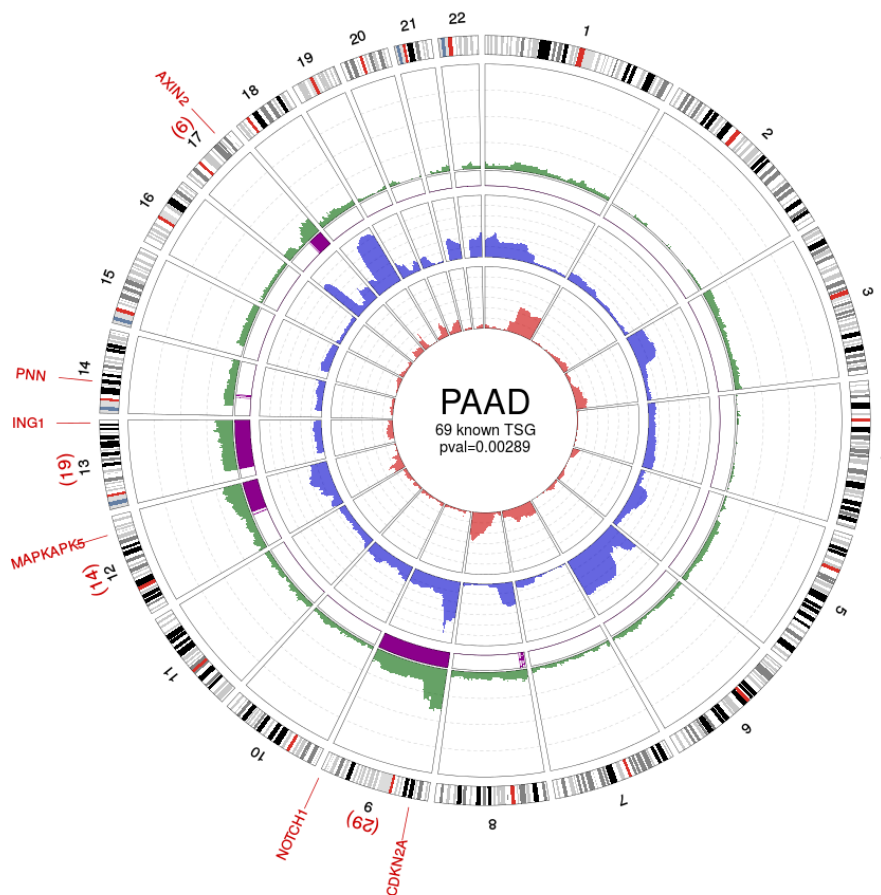


Figura 5. Gràfiques *circos* amb el perfil d'alteracions genòmiques de CN i UPD per cada tipologia de càncer. En els radis, de dins a fora, es mostra la freqüència de guanys de CN (vermell), la freqüència de pèrdues de CN (blau), les regions significatives d'UPD (morat), i la freqüència d'UPD (verd). Cada franja dins de les regions de CNA i UPD correspon a una freqüència del 10%. A l'interior, s'indica l'acrònim del tipus de càncer, el nombre de TSG detectats a les regions d'UPD, i el llindar de significança *p-value* després d'aplicar l'anàlisi de permutacions i la correcció per múltiples comparacions. A l'exterior hi ha els ideogrames de cada cromosoma, i s'indica entre parèntesis en vermell el nombre de TSG coneguts trobats a les regions UPD significatives i es senyala algun d'aquests gens a nivell de mostra.

a 9q hem de destacar que la freqüència de pèrdues de CN és només d'un 10% quan aquest braç presenta freqüències d'UPD del 30–40%. Observem aquest mateix comportament al cromosoma 17.

El perfil de STAD es mostra molt semblant a ESAD en quant a CNA, com s'ha reportat amb anterioritat [13,12], però les freqüències d'UPD són notablement inferiors, per sota del 10%

excepte a 17p. Com a ESAD, trobem un *peak* a 16q amb WWOX (10,0%), i a 3p FHIT (11,8%). Interessantment, a 1p trobem en aquest cas ARID1A (8,8%) que la seva inactivació està associada a la metastasi en càncer gàstric [17], i NF2 a 22p [27] que codifica la proteïna supressora de tumors *Merlin*. També troben una alteració focal significativa a 4q ANXA10 (7,9%), tot i que segons [19] no sembla tractar-se d'un TSG.

A LIHC la freqüència d'UPD és inferior al 10%. Trobem dues regions significatives a nivell de braç, a 6q i 13q, i diversos UPD focals. A 3p trobem el TSG TGFBR2 (6,2%), i a 9p tenim un *peak* amb CDKN2A (6,2%). En el mateix cromosoma, a 9q, hi ha una petita regió significativa (5,9%) amb 5 gens no reconeguts com a TSG, però podem destacar-ne CDC14B que té un rol en la divisió cel·lular i s'ha trobat disminuït en càncer renal [20]. A 11p trobem un *peak* amb LUZP2 (5,6%), tot i que no s'ha observat cap anomalia fenotípica en la seva deleció en models animals [34]. Al cromosoma 18 destaca freqüència d'UPD del gen SOCS6 (6,2%), que s'ha trobat en deleció en càncers colorectals [22]. Altres gens significatius sense funcionalitat supressora de tumors coneguda són HYDIN (12.4%), a 16q, que està relacionat amb el desenvolupament de les cèl·lules epitelials, i ACACB i FOXN4 (5.9%), a 12q.

PAAD també presenta un perfil d'UPD amb unes freqüències molt baixes, inferiors al 10% excepte a 9p que arriba al 15%, que no correlacionen amb les freqüències de pèrdua que arriben al 50%. Les regions significatives a nivell de braç són 13q i el cromosoma 9. Dins de 13q trobem una regió *peak* amb 20 gens, entre ells el TSG ING1 (7,1%). També trobem regions més focals amb TSG com FALZ (7,1%) a 7q, PNN (4,5%) a 14q i MAPKAPK5 (7,1%) a 12q. A l'inici del cromosoma 8 trobem una regió amb alguns esdeveniments focals, amb 100 gens, que en aquest cas si correlacionen amb el perfil de pèrdues de CN.

5. Discussió i conclusions

En aquest treball hem presentat una ampla anàlisi de perfils de CNA i UPD de diferents tipus de càncers gastrointestinals. El nostre *pipeline* ha demostrat ser vàlid per a la identificació de TSG i nous candidats a TSG amb un nombre d'ocurrències significatives d'UPD adquirida.

En la caracterització de les mostres, hem trobat una notable presència de SCNA i UPD a nivell de braç cromosòmic que reflecteix, d'acord amb [2], la facilitat en que aquestes mutacions poden ocórrer respecte els esdeveniments focals. Com apunta [7], hem observat que moltes d'aquestes alteracions tendeixen a ser compartides per varis tipus de càncer com la freqüent deleció del braç 17p, que conté l'important TSG TP53, en tots els grups menys PAAD. Tot i així, aquests llargs segments fan difícil la identificació de gens controladors específics.

També hem observat en molts casos una concordança entre els perfils de pèrdua al genoma i la presència d'UPD. Hem confirmat aquesta observació amb l'anàlisi de correlacions entre les freqüències d'aquestes alteracions. Com ha apuntat recentment [32], aquesta correlació pot indicar una mateixa finalitat, la inactivació de TSG.

Pel que fa a les aberracions focals, hi ha una tendència a que els peaks identificats en un tipus de càncer es trobin en altres grups. És el cas per exemple de FHIT, que presenta una alta freqüència de pèrdua i UPD, a COAD, READ, ESAD i STAD. Hem de tenir en compte però, que com que el nostre objectiu principal és l'estudi de la UPD, el concepte d'esdeveniment focal pot ser diferent que en estudis previs de CNA [7,9,11,16,36]. Aquests detecten les alteracions focals multinivell que s'inclouen dins de regions més grans [24] fent servir el valor de CN per determinar l'inici i final de cada regió. Però aquesta magnitud no té sentit en el cas de *copy-neutral* LOH. La nostra aproximació és més senzilla: un esdeveniment focal és aquell amb una longitud inferior al braç del cromosoma.

Els perfils d'alteracions genòmiques semblen també dependre de la proximitat dels òrgans afectats. Així, COAD i READ presenten alteracions d'UPD semblants al cromosoma 17 i als braços 5q, 12q i 14q, on trobem APC (a 5q31.2) que ha demostrat ser un important TSG en càncer de colon [32]. Fins i tot em trobat a ambdós grups dos gens candidats no identificats fins ara com a TSG [30,37] a 16p i 20p, A2BP1 i C20orf133.

ESAD i STAD presenten uns perfils de CNA molt semblants, tot i que STAD amb

freqüències inferiors. En UPD, comparteixen regions significatives a 17p i 9p, i un esdeveniment focal al gen WWOX, a 16q. A LIHC i PAAD també trobem un *peak* comú a CDKN2A (9p). Dulak et al. [11] fins i tot apunten a que hi ha uns nivells diferents d'inestabilitat cromosòmica entre els òrgans superiors i inferiors de l'aparell digestiu. Podem confirmar aquest resultat a la nostra mostra, on veiem unes freqüències d'alteracions significativament més altes a ESAD que en la resta de tumors.

Tot i que fins ara hem apuntat a que moltes regions i TSG són alterats freqüentment en diversos tipus de càncer, no és segur que aquestes alteracions tinguin una funcionalitat rellevant en tots els teixits. Així, l'estudi de les alteracions per ocurrència suposaria un *bias* important en la investigació, obviant aquells esdeveniments petits en magnitud però que siguin funcionalment importants, de forma singular o en combinació amb altres alteracions. De fet, anant en contra del que descriu [3], fins i tot dins d'un mateix tipus de càncer podem observar una heterogeneïtat en els perfils de CNA i UPD, com veiem a la Figura 2. Mentre que l'anàlisi de moltes mostres ens donarà una major potència estadística, ajuntar perfils diferents ens pot emmascarar alteracions importants fent-les difícils de diferenciar de les aberracions del *background*.

Els nostres esforços futurs en aquest sentit, van enfocats a l'estudi de les alteracions genòmiques amb una aproximació basada en *network medicine*. La representació de les interaccions entre gens i productes gènics en forma de xarxa, tot i ser una simplificació que perd la informació de cada node, ha demostrat ser útil per a la comprensió de molts processos biològics [31,33] i ens pot permetre fer inferències sobre la causalitat de les alteracions en els processos cel·lulars.

Bibliografia

- [1] Abiatari, Ivane, et al. "The microtubule-associated protein MAPRE2 is involved in perineural invasion of pancreatic cancer cells." *International journal of oncology* 35.5 (2009): 1111-1116.
- [2] Albertson, Donna G. "Gene amplification in cancer." *TRENDS in Genetics* 22.8 (2006): 447-455.
- [3] Andersen, Claus Lindbjerg, et al. "Frequent occurrence of uniparental disomy in colorectal cancer." *Carcinogenesis* 28.1 (2006): 38-48.
- [4] Bengtsson, Henrik, Pierre Neuvial, and Terence P. Speed. "TumorBoost: Normalization of allele-specific tumor copy numbers from a single pair of tumor-normal genotyping microarrays." *BMC bioinformatics* 11.1 (2010): 245.
- [5] Bengtsson, Henrik, Pratyaksha Wirapati, and Terence P. Speed. "A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6." *Bioinformatics* 25.17 (2009): 2149-2156.
- [6] Benjamini, Yoav, and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society. Series B (Methodological)* (1995): 289-300.
- [7] Beroukhi, Rameen, et al. "The landscape of somatic copy-number alteration across human cancers." *Nature* 463.7283 (2010): 899-905.
- [8] Berrino, Franco, et al. "Survival of cancer patients in Europe: the EURO CARE-3 study [Journal supplement issue]." *Annals of Oncology* 14.Supplement 5 (2003): v1-v155.
- [9] Cancer Genome Atlas Network. "Comprehensive molecular characterization of human colon and rectal cancer." *Nature* 487.7407 (2012): 330-337.
- [10] Ciriello, Giovanni, et al. "Emerging landscape of oncogenic signatures across human cancers." *Nature genetics* 45.10 (2013): 1127-1133.
- [11] Dulak, Austin M., et al. "Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis." *Cancer research* 72.17 (2012): 4383-4393.
- [12] Dulak, Austin M., et al. "Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity." *Nature genetics* 45.5 (2013): 478-486.
- [13] Frankel, Adam, et al. "Genome-wide analysis of esophageal adenocarcinoma yields specific copy number aberrations that correlate with prognosis." *Genes, Chromosomes and Cancer* 53.4 (2014): 324-338.

- [14] Garrido-Laguna, Ignacio, and Manuel Hidalgo. "Pancreatic cancer: from state-of-the-art treatments to promising novel therapies." *Nature Reviews Clinical Oncology* 12.6 (2015): 319-334.
- [15] Gonzalez-Zulueta, Mirella, et al. "High frequency of chromosome 9p allelic loss and CDKN2 tumor suppressor gene alterations in squamous cell carcinoma of the bladder." *Journal of the National Cancer Institute* 87.18 (1995): 1383-1393.
- [16] Hoadley, Katherine A., et al. "Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin." *Cell* 158.4 (2014): 929-944.
- [17] Inada, Ryo, et al. "ARID1A expression in gastric adenocarcinoma: Clinicopathological significance and correlation with DNA mismatch repair status." *World journal of gastroenterology*: WJG 21.7 (2015): 2159.
- [18] Jiang, Peng, et al. "Inference of transcriptional regulation in cancers." *Proceedings of the National Academy of Sciences* 112.25 (2015): 7731-7736.
- [19] Kim, Jung Ho, et al. "Gastric-type expression signature in serrated pathway-associated colorectal tumors." *Human pathology* 46.5 (2015): 643-656.
- [20] Kim, Younghye, et al. "Loss of CDC14B Expression in Clear Cell Renal Cell Carcinoma Meta-Analysis of Microarray Data Sets." *American journal of clinical pathology* 141.4 (2014): 551-558.
- [21] Kirley, Sandra D., et al. "The Cables gene on chromosome 18Q regulates colon cancer progression in vivo." *Cancer biology & therapy* 4.8 (2005): 861-863.
- [22] Letellier, Elisabeth, et al. "Identification of SOCS2 and SOCS6 as biomarkers in human colorectal cancer." *British journal of cancer* 111.4 (2014): 726-735.
- [23] Marcos-Gragera, R., et al. "Incidencia del cáncer en Cataluña, 1998-2002." *Medicina clínica* 131 (2008): 4-10.
- [24] Mermel, Craig H., et al. "GISTIC 2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers." *Genome Biol* 12.4 (2011): R41.
- [25] Olshen, Adam B., et al. "Parent-specific copy number in paired tumor-normal studies using circular binary segmentation." *Bioinformatics* 27.15 (2011): 2038-2046.
- [26] Peng, Haixia, et al. "Downregulation of DLC-1 gene by promoter methylation during primary colorectal cancer progression." *BioMed research international* 2013 (2012).
- [27] Petrilli, A. M., and C. Fernández-Valle. "Role of Merlin/NF2 inactivation in tumor biology." *Oncogene* (2015).
- [28] Qin, Bo, et al. "DBC1 Functions as a Tumor Suppressor by Regulating p53 Stability." *Cell reports* 10.8 (2015): 1324-1334.
- [29] R Core Team. "R: A language and environment for statistical computing." (2015).
- [30] Rajaram, Megha, et al. "Two distinct categories of focal deletions in cancer genomes." *PloS one* 8.6 (2013): e66264.

- [31] Rolland, Thomas, et al. "A proteome-scale map of the human interactome network." *Cell* 159.5 (2014): 1212-1226.
- [32] Torabi, Keyvan, et al. "Patterns of somatic uniparental disomy identify novel tumor suppressor genes in colorectal cancer", *unpublished work*.
- [33] Vidal, Marc, Michael E. Cusick, and Albert-Laszlo Barabasi. "Interactome networks and human disease." *Cell* 144.6 (2011): 986-998.
- [34] Wu, Min, Edward J. Michaud, and Dabney K. Johnson. "Cloning, functional study and comparative mapping of Luzp2 to mouse chromosome 7 and human chromosome 11p13–11p14." *Mammalian genome* 14.5 (2003): 323-334.
- [35] Wikipedia contributors. "Loss of heterozygosity." *Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia*, 27 Jan. 2015. Web. 9 Jul. 2015.
- [36] Zack, Travis I., et al. "Pan-cancer patterns of somatic copy number alteration." *Nature genetics* 45.10 (2013): 1134-1140.
- [37] Zhao, Min, Jingchun Sun, and Zhongming Zhao. "TSGene: a web resource for tumor suppressor genes." *Nucleic acids research* 41.D1 (2013): D970-D976.
- [38] Zimonjic, Drazen B., and Nicholas C. Popescu. "Role of DLC1 tumor suppressor gene and MYC oncogene in pathogenesis of human hepatocellular carcinoma: Potential prospects for combined targeted therapeutics (Review)." *International journal of oncology* 41.2 (2012): 393-406.