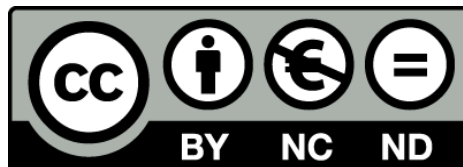


# Caracterització epigenètica dels elements Alu i el seu paper en la regulació transcripcional

Berta Martín Abad



Aquesta tesi doctoral està subjecta a la llicència Reconeixement- NoComercial – SenseObraDerivada 3.0. Espanya de Creative Commons.

Esta tesis doctoral está sujeta a la licencia Reconocimiento - NoComercial – SinObraDerivada 3.0. España de Creative Commons.

This doctoral thesis is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0. Spain License.

# Caracterització epigenètica dels elements Alu i el seu paper en la regulació transcripcional

**Berta Martín Abad**

Tesi doctoral 2013



## Caracterització epigenètica dels elements Alu i el seu paper en la regulació transcripcional

Memòria presentada per  
**Berta Martín Abad**

Per optar al títol de  
**Doctora per la Universitat de Barcelona**

Tesi dirigida sota la direcció del Dr. Miquel Àngel Peinado i la Dra. Mireia Jordà a l'IMPPC (Institut de Medicina Predictiva i Personalitzada del Càncer)

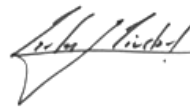
Facultat de Farmàcia  
Universitat de Barcelona  
Programa de Biomedicina

Directors



Miquel Àngel Peinado  
Morales i Mireia Jordà  
Ramos

Tutor



Carles Ciudad  
Gómez

Doctoranda



Berta Martín  
Abad

*Barcelona, setembre 2013*

*"El camí del progrés no és ni ràpid ni fàcil"*

Marie curie

*"Un país sense investigació és un país sense desenvolupament"*

Severo Ochoa

A tots als que he tingut  
a prop durant tot aquest temps

Sobre el 2001 vaig entrar al que era l'antic IRO, actual IDIBELL. Allà vaig passar 7 anys i haig agrair a tota la gent que em va obrir les portes a la recerca ja que gràcies a ells ara em trobo en aquest punt. Gent amb la que vaig compartir grans moments i que han format part de la meua vida des d'aquell moment:

A la Ruth, a la petarda que, tot i estar lluny, sempre està present i amb la que he compartit moltíssimes coses: els inicis al lab, les confidències, les penes i les alegries.

A l'Eli, l'Eva, la Regi, l'Anna, l'Elena i la Mire amb les que he disfrutat de viatges, festes, sopars i un llarg etc dels que s'han quedat moltes anècdotes i records. Han anat passant els anys i, entre ells, nous projectes i il·lusions, canvis de feina, de país i algunes bodes, naixements.

A la Txiki, la Rebeca, la Verena amb les que vam formar un gran equip en l'AS lab. I a l'Antonia, tota la vida recordaré la meua entrada triomfal a l'IRO.

I com no a la Magüi, la Bego i l'Ernest. Les meves flors. Ja fa un munt d'anys que ens coneixem i aquí continuem, fotuts en aquest món de bojós. M'encanta passar tardes senceres amb vosaltres intentant arreglar el món. La Wiki amb la seva gran personalitat, la Bego *con su caràcter manchego* i el Pitu que, tot i la seva cresta, és un gran amic.

Van anar passant els anys i com que aquest món de la recerca és més aviat petit i tots acabem coneixent-nos i entrecruant-los, el Miquel Àngel i la Mireia em van brindar la possibilitat de formar part del MAP lab i van passar de companys de feina i amics, a ser els meus directors de tesi. Un canvi en la nostra relació que m'ha fet veure que a part de grans persones són grans científics, cadascú amb la seva idea ben formada de com entendre la vida i la ciència. El Miquel Àngel, que mai perd la il·lusió o intenta no perdre-la i que sempre té la porta del seu despatx oberta per parlar de ciència, de la vida o de la cosa més absurda. Miquel Àngel, no perdis mai la teua essència. I la Mireia, que s'ha transformat en una gran investigadora i que ha superat de llarg la dificultat de ser *la jefa* i codirectora de tesi a part que amiga. Gràcies a tots dos per donar-me aquesta possibilitat i per ensenyar-me a fer ciència rigorosa i a gaudir d'ella.

Així doncs, al 2007 vaig començar a l'IMPPC. Entrar en un centre que comença de zero et brinda la possibilitat de conèixer cada racó i anar veient com es va omplint de grups que entren amb la il·lusió d'iniciar un nou projecte. Tinc que agrair a moltíssima gent que ha compartit amb mi aquesta nova etapa:

Al Marcus lab, un grup que a part de fer bona ciència sap gaudir de la vida. Gràcies a la Cath, la Mel, el Julien, el Roberto, l'Anna, la Neus i com no a la Vane per fer-me passar grans moments dintre i fora del lab.

Als Fumis, que ens han brindat uns quants Sant Valentine's dinners com a excusa per trobar-nos tots fora de la feina i que, fent honor als seus orígens, han immortalitzat els moments més èpics de l'IMPPC. I a l'Emili, un pou de sabiduria i el *comodin de la llamada* en moments crítics.

A les Jeris per la seva alegria i riure contagiós. A la Bea i al Sergio, i a la Sònia que, ara, arrel del rocambolesc món del Comitè d'Empresa he conegut més a fons. Al grup del

Edu Serra lab (a l'Edu, a l'Imma, al Bernat, a l'Eli i al Josep) per ser tan bona gent. Al departament de bioinformàtica, als que estan i als ja no, en especial a l'Anna i al Alonso, sempre tan *cañeros*.

Agrair també a la Patri, per tenir qualsevol reactiu preparat al minut de demanar-li i a la Bárbara per preocupar-se de què tot funcioni correctament. Gràcies a les dos per fer-nos la vida més fàcil al centre.

I sens dubte, al MAP lab. Crec que hem fet un gran equip dins i fora del lab. Treballar amb gent com vosaltres ho fa tot molt més fàcil.

Agrair als que ja no estan: A la Regi, sempre tant positiva, m'encanta la teva forma d'afrontar la vida. A la Inés, la portuguesa - catalana - londinenca, amb la que hem compartit uns quants Vinils. A la Marta, la pituriua que vaig conèixer de pràctiques entre WB i WB, preocupada per prendre la millor decisió i, mira-la ara. Al Llorenç, sempre amb les seves conyes, jocs de paraules i indirectes. Al Sergi, el doble del Miquel Noguera, mira que m'has arribat a fer riure. Al Víctor, l'enciclopèdia del lab. I a tots els que heu anat passant de pràctiques pel lab (Sara, Berta, Dèlia, Ruth,...). Molta sort a tots amb els vostres nous projectes.

I als que continuen aquí: A la Mar, la MasterMix del lab, sempre amb ganes de tirar les coses *pa'lante*, gràcies per endinsar-me en el món del MAP lab, de les PCRs i dels bisulfits. Al Quim, sempre tan clar i sincer, molts ànims, ja estàs en la recta final. A la Raquel, eufòrica i motivada amb tot allò que fa. I al *muscle team*: A la Mònica, tot un nervi, corrent a munt i avall per conciliar vida laboral i familiar. A l'Elvira, la perleta i alegria del lab, gràcies per buscar sempre el cantó positiu de les coses. A la Yaiza, la guapi entre les guapis, i sí, *tú más*. A l'Anna, a la que portem de cul entre gràfics, tests estadístics i que més d'un cop és posa les mans al cap amb les nostres "n". I a la Izaskun, *por fin el mundo de los primates encontrará su camino*.

Finalment tinc que agrair a la meva família, als meus pares i al meu germà per estar sempre amb mi, en els bons i mals moments, per donar-me suport i animar-me a perseguir els meus somnis.

I al Luis. Moltíssimes gràcies per la teva paciència i per aguantar els meus rotllos sobre temes tan esotèrics com "los Alos". Gràcies per posar-m'ho fàcil, per entendre els meus horaris i per tindre sempre un plat a la taula a les tantes de la nit.





Abreviatures	15
Índex de figures	17
Índex de taules	21
<b>INTRODUCCIÓ</b>	<b>23</b>
<b>1. EPIGENÈTICA</b>	<b>25</b>
1.1. Estructura de la cromatina	25
1.2. Mecanismes epigenètics	28
1.2.1. La metilació de l'ADN	29
1.2.1.1. Les ADN metiltransferases (DNMTs)	31
1.2.1.2. Metilació en les illes CpG	33
1.2.2. Funcions de la metilació de l'ADN	34
1.2.2.1. Imprompta genètica	34
1.2.2.2. Inactivació del cromosoma X	35
1.2.2.3. Silenciament dels elements repetitius	35
1.2.2.4. Altres implicacions de la metilació de l'ADN	36
1.2.3. Modificacions d'histones	36
1.2.3.1. Acetilació	38
1.2.3.2. Fosforilació	39
1.2.3.3. Metilació	39
1.2.3.4. El codi d'histones i la conformació de la cromatina	40
1.3. Organització del genoma en el nucli	41
<b>2. ELEMENTS REPETITIUS</b>	<b>42</b>
2.1. Metilació dels elements repetitius	42
2.2. Classificació dels elements repetitius	43
2.2.1. ADN repetit en tàndem	43
2.2.2. ADN repetit dispers	45
<b>3. ELEMENTS ALU</b>	<b>49</b>
3.1. Origen i estructura dels elements Alu	49
3.2. Transposició dels elements Alu	51
3.3. Evolució i classificació dels elements Alu	53
3.4. Implicació evolutiva dels elements Alu	55
3.5. Implicació genòmica dels elements Alu	57
3.6. Canvis epigenètics en els elements Alu	58
<b>4. CÀNCER</b>	<b>60</b>
4.1. Bases moleculars del càncer	62
4.2. Inestabilitat genètica	63
4.2.1. Inestabilitat de microsatèl·lits	63
4.2.2. Inestabilitat cromosòmica	64
<b>5. CÀNCER I EPIGENÈTICA</b>	<b>65</b>
5.1. Hipermetilació	66
5.2. Hipometilació	67

5.2.1.	Hipometilació local	69
5.2.2.	Hipometilació global	69
<b>6.</b>	<b>DIFERENCIACIÓ I EPIGENÈTICA</b>	<b>70</b>
6.1.	Metilació de l'ADN	70
6.2.	Modificació d' histones	71
<b>7.</b>	<b>RESULTATS PREVIS: QUANTIFICACIÓ I IDENTIFICACIÓ D'ELEMENTS ALU</b>	<b>72</b>
<b>HIPÒTESI I OBJECTIUS</b>		<b>73</b>
<b>MATERIAL I MÈTODES</b>		<b>77</b>
<b>1.</b>	<b>MATERIAL</b>	<b>79</b>
1.1.	Línies cel·lulars	79
1.2.	Mostres humanes	80
1.3.	Mostres de primats no humans	80
<b>2.</b>	<b>MÈTODES</b>	<b>81</b>
2.1.	Tractaments/drogues	81
2.2.	Expressió a nivell d'ARNm	81
2.2.1.	Extracció de l'ARN total	81
2.2.2.	Obtenció de l'ADN complementari	82
2.2.3.	PCR quantitativa en temps real	82
2.2.4.	Northern Blot	84
2.3.	RACE	87
2.4.	Western Blot	90
2.5.	siRNA (ARN de silenciament)	92
2.6.	Estudi de la metilació	93
2.6.1.	Tractament per bisulfit	93
2.6.2.	QUMA i AUMA	98
2.6.3.	NSUMA	100
2.7.	Estudi de les marques d'histona: Immunoprecipitació de Cromatina	103
<b>RESULTATS</b>		<b>107</b>
<b>1.</b>	<b>CARACTERITZACIÓ A NIVELL DE SEQÜÈNCIA DELS ELEMENTS ALU DESMETILATS</b>	<b>109</b>
<b>2.</b>	<b>CARACTERITZACIÓ EPIGENÈTICA DELS ELEMENTS ALU</b>	<b>113</b>
2.1	Estudi de la metilació dels elements Alu	114
2.1.1	Estudi de la metilació dels elements Alu durant la tumorigènesi	114
2.1.1.1.	Estudi de la metilació dels elements Alu en tumors primaris i línies de càncer de còlon	114
2.1.1.2.	Efecte de <i>drogues epigenètiques</i> en la metilació dels elements Alu	117
2.1.1.3.	Estudi de la metilació de les regions flanquejants als elements Alu	118
2.1.2.	Estudi de la metilació dels elements Alu durant la diferenciació cel·lular	121
2.1.3.	Conservació evolutiva de la metilació dels elements Alu en primats	123
2.2	Estudi de l'estructura de la cromatina associada als elements Alu	129
2.2.1	Estudi de la cromatina associada als elements Alu durant la tumorigènesi	130

2.2.1.1.	Estudi de la cromatina associada als elements Alu en línies de càncer de còlon	130
2.2.1.2.	Estudi de la cromatina associada a les regions flanquejants als elements Alu	135
2.2.2.	Estudi de la cromatina associada als elements Alu en la diferenciació	140
<b>3.</b>	<b>ELEMENTS ALU COM A POTENCIALS REGULADORS EPIGENÈTICS DE L'EXPRESSIÓ GÈNICA</b>	<b>142</b>
3.1.	Alu AJ2C1	143
3.1.1.	Estudi epigenètic de l'element AJ2C1	143
3.1.2.	Estudi epigenètic de la regió flanquejant a AJ2C1	145
3.1.3.	Estudi de l'expressió de <i>DIEXF</i> a nivell d'ARNm	147
3.1.4.	Estudi de l'expressió de <i>DIEXF</i> a nivell de proteïna	161
3.1.5.	Paper de l'element AJ2C1 en càncer	167
3.2.	Alu ZALU3	172
3.2.1.	Estudi epigenètic de l'element ZALU3 i la regió flanquejant	172
3.2.2.	Estudi de l'expressió de <i>GLDC</i> a nivell d'ARNm	175
3.2.3.	Estudi de l'expressió de <i>GLDC</i> a nivell de proteïna	176
3.2.4.	Regulació de l'element ZALU3 en la diferenciació	177
<b>DISCUSSIÓ</b>		<b>183</b>
<b>1.</b>	<b>ELS ELEMENTS ALU DESMETILATS CONSTITUEIXEN UNA SUBPOBLACIÓ ESPECÍFICA DINS L'ALUOMA</b>	<b>185</b>
<b>2.</b>	<b>ELS ELEMENTS ALU DESMETILATS PRESENTEN UN MAJOR GRAU DE CONSERVACIÓ EVOLUTIVA</b>	<b>188</b>
<b>3.</b>	<b>ESTAT EPIGENÈTIC DELS ELEMENTS ALU I EL SEU PAPER COM A BARRERA DE METILACIÓ</b>	<b>190</b>
<b>4.</b>	<b>HIPOMETILACIÓ DELS ELEMENTS ALU EN CÀNCER</b>	<b>194</b>
4.1.	Resistència a la desmetilació de certs elements Alu	195
4.2.	Paper de l'element Aj2c1 en la regulació de <i>DIEXF</i> i la seva implicació en càncer	196
4.2.1.	<i>DIEXF</i> i el seu paper en càncer	196
4.2.2.	Aj2c1 actua com a barrera de metilació, regulant la conformació de la cromatina i afectant l'expressió de <i>DIEXF</i>	200
<b>5.</b>	<b>PAPER DELS ELEMENTS ALU DURANT LA DIFERENCIACIÓ</b>	<b>202</b>
5.1.	Paper de l'element ZALU3 en la regulació de <i>GLDC</i> i la seva implicació en el procés de diferenciació	203
5.1.1.	<i>GLDC</i> i el seu paper en el metabolisme de la glicina	203
5.1.2.	Alteracions en <i>GLDC</i>	204
5.1.3.	<i>GLDC</i> i el seu paper en càncer	205
5.1.4.	ZALU3 actua com a aïllador genòmic i regula l'expressió de <i>GLDC</i>	206
<b>6.</b>	<b>IMPLICACIÓ DELS ELEMENTS ALU EN LA REGULACIÓ GÈNICA</b>	<b>208</b>
<b>CONCLUSIONS</b>		<b>211</b>
<b>BIBLIOGRAFIA</b>		<b>215</b>
<b>ANNEXOS</b>		<b>233</b>
I.	Encebadors i condicions de PCR	235
II.	Estat de metilació dels elements Alu	251
III.	Posició dels encebadors de ChIP	263

5-AzaC	5-aza-2' deoxicitidina
5-mC	5-metil Citosina
5-hmC	5-hidroximetil Citosina
18S	ARN ribosomal 18
28S	ARN ribosomal 28
AcH3K14	Acetilació de la lisina 14 en la histona 3
AcH3K9	Acetilació de la lisina 9 en la histona 3
Ac4c1	Element Alu trobat amb la tècnica AUMA
ADN	Àcid desoxiribonucleic
ADNc	ADN complementari o de cadena senzilla
Alu	Element de la família dels SINEs
Alu S	Família d'elements Alu més antiga
Alu Y	Família d'elements Alu més jove
Alu J	Família d'elements Alu intermèdia
ARN	Àcid ribonucleic
ARNm	ARN missatger
AUMA	<i>Amplification of Unmethylated Alu's</i>
Aza	5-aza-2' deoxicitidina
C	Citosina
C1orf107 o DIEXF	<i>Digestive Organ Expansion factor homolog</i>
CaCo2	Línia de càncer colorectal
CCR	Càncer de còlon i recte o càncer colorectal
ChIP	Immunoprecipitació de Cromatina
Cp	<i>Crossing Point</i>
CpG	Dinucleòtid CG (citocina-guanina)
CpGi	Illa CpG
DIEXF o C1orf107	<i>Digestive Organ Expansion factor homolog</i>
DKO	Línia de càncer colorectal, HCT116 deficient en DNMT1 i DNMT3b
DNMTs	ADN metiltransferases
DNMT1	ADN metiltransferasa de manteniment1
DNMT3a	ADN metiltransferasa de novo 3a
DNMT3b	ADN metiltransferasa de novo 3b
DTT	Dithiothreitol
E	Eficiència
EB	Cos Embrioide o <i>Embryoid bodies</i>
F	Fibroblast
GLDC	Glicina deshidrogenasa (descarboxilant)
H3	Histona 3
H3K4me3	Trimetilació de la lisina 4 en la histona 3
H3K9me3	Trimetilació de la lisina 9 en la histona 3
H3K27me3	Trimetilació de la lisina 27 en la histona 3
HAT	<i>Histone acetyltransferase</i>
HDAC	<i>Histone deacetylase</i>
HCT116	Línia de càncer colorectal
HCT116-DKO	Línia de càncer colorectal deficient en DNMT1 i DNMT3b
hESC 2	Human Embryonic Stem cells 2
hESC 4	Human Embryonic Stem cells 4
hESC 6	Human Embryonic Stem cells 6
HFF	Fibroblast de prepuci
HT29	Línia de càncer colorectal
IgG	Immunoglobulina
IP	Immunoprecipitat
INPUT	Cromatina total
K3	Línia cel·lular de queratinòcits 3
K18	Línia cel·lular de queratinòcits 18

K137	Queratinòcits 137
KiPS4F1	Cèl·lules pluripotents induïdes diferenciades de queratinòcits de prepuci F1
KiPS4F8	Cèl·lules pluripotents induïdes diferenciades de queratinòcits de prepuci F8
[H]KiPS4F1	Cèl·lules pluripotents induïdes diferenciades de queratinòcits de bulb pilós
LINE	<i>Long Interspersed Nuclear Elements</i>
LoVo	Línia de càncer colorectal
LTR	<i>Long terminal repeat</i>
MMTA	Línia cel·lular de queratinòcits
MSI	Microsatèl·lits
NB	Northern Blot
NSUMA	<i>Next Generation Sequencing of Unmethylated Alu's.</i>
ORF	<i>Open Reading Frame</i>
pb	Parell de bases
PCR	Reacció en Cadena de la Polimerasa
qALU	Quantificació d'elements Alu totals
QUMA	<i>Quantification of Unmethylated Alu's</i>
RACE	<i>5' and 3' Rapid Amplification of cDNA Ends</i>
RT-qPCR	Transcripció Reversa – PCR quantitativa en temps real
SINE	<i>Short Interspersed Nuclear Elements</i>
siRNA	ARN de silenciació o <i>small interfering RNA</i>
SmaI	Enzim de restricció sensible a la metilació
Sw480	Línia de càncer colorectal
T	Timina
TCGA	<i>The Cancer Genome Atlas</i>
T <sub>m</sub>	Temperatura de <i>melting</i> (temperatura d'hibridació)
TSA	Trichostatina A
TSS	<i>Transcription start site</i> (Inici de transcripció)
U	Uracil
UTR	<i>Untranslated region</i>

- Fig. I.1. Condensació de l'ADN.
- Fig. I.2. CTCF s'uneix a regions de l'ADN anomenades *insulators* actuant com a barrera i evitant així que l'heterocromatina s'expandeixi cap a l'eucromatina.
- Fig. I.3. Metilació de la citosina.
- Fig. I.4. Hidroximetilació de la 5 metilcitosina.
- Fig. I.5. Procés de silenciament d'una regió transcripcionalment activa per mitjà de la metilació de l'ADN.
- Fig. I.6. Patró de metilació en illes CpG.
- Fig. I.7. CTCF regula la imprompta genètica del locus *IGF2/H19*.
- Fig. I.8. Modificacions d'histones.
- Fig. I.9. Tipus d'ADN repetit dispers en el genoma humà.
- Fig. I.10. Estructura de 7SL ARN del qual segurament deriven els elements Alu.
- Fig. I.11. Formació de l'element Alu a partir de la fusió de les dos seqüències originàries de 7SL ARN amb el domini S deletionat.
- Fig. I.12. Estructura de l'element Alu.
- Fig. I.13. Mecanisme de retrotransposició de l'element Alu.
- Fig. I.14. Alineament de les seqüències consens de les subfamílies Alu.
- Fig. I.15. Expansió dels elements Alu en primats.
- Fig. I.16. Aparició dels elements Alu fa 65 milions d'anys, coincidint amb l'origen i expansió dels primats.
- Fig. I.17. Etapes de progressió del CCR.
- Fig. I.18. Patrons de hipermetilació en cèl·lules canceroses.
- Fig. I.19. Patrons de hipometilació en cèl·lules canceroses.
- Fig. I.20. Quantificació dels elements Alu desmetilats en 17 parelles de mucosa normal i carcinoma colorectal per QUMA.
- Fig. M.1. Programa del Light Cycler®, permet programar les temperatures i temps de desnaturalització, amplificació i fusió de la mostra.
- Fig. M.2. Valors obtinguts amb el software del Light Cycler®.
- Fig. M.3 . Procés de transformació de l'ADN per mitjà del tractament per bisulfit sòdic.
- Fig. M.4. Representació d'un fragment seqüenciat amb l'encebador directe.
- Fig. M.5. Representació d'un fragment seqüenciat amb l'encebador indirecte.
- Fig. M.6. Esquema de les tècniques QUMA i AUMA. Metodologies aplicades per quantificar i identificar respectivament.

- Fig. M.7. Esquema de la tècnica ZUMA.
- Fig. M.8. Esquema representatiu de la immunoprecipitació de cromatina pas a pas.
- Fig. M.9. Esquema del muntatge d'un Northern Blot en el moment de la transferència.
- Fig. M.10. RACE. Disseny dels encebadors GSP, específics de la regió d'interès.
- Fig. R.1. Trets genòmics específics dels elements Alu desmetilats.
- Fig. R.2. EN1 i CPLX2 utilitzats com a controls del correcte funcionament del tractament amb 5-AzadC, TSA i una combinació d'ambdós en la línia HCT116.
- Fig. R.3. Estat de metilació de les regions flanquejants als elements Alu seleccionats en diferents línies de CCR.
- Fig. R.4. Comparació entre l'espècie i l'estat de conservació dels elements Alu.
- Fig. R.5. Comparació entre la família i l'estat de conservació dels elements Alu.
- Fig. R.6. Comparació entre l'estat de metilació i l'estat de conservació dels elements Alu en ximpanzé, goril·la, orangutan i macaco.
- Fig. R.7. Estat de metilació dels elements Alu seleccionats per l'estudi en mostres de sangs de diferents espècies de primats: humana, ximpanzé, goril·la i orangutan.
- Fig. R.8. Anàlisi per ChIP dels elements Alu en la línia HCT116.
- Fig. R.9. Anàlisi per ChIP dels elements Alu en les línies HCT116, Sw480 i LoVo.
- Fig. R.10. Anàlisi per ChIP dels elements Alu en les línies HCT116 i HCT116-DKO.
- Fig. R.11. Anàlisi per ChIP de les regions flanquejants als elements Alu desmetilats (Ar3c7, Au104, Lv165) de la línia HCT116.
- Fig. R.12. Anàlisi per ChIP de les regions flanquejants als elements Alu parcialment metilats (Aj2c1, Lv183, Au60) de la línia HCT116.
- Fig. R.13. Anàlisi per ChIP dels elements Alu en les línies hESC4 i HFF.
- Fig. R.14. Esquema de la regió Aj2c1 d'estudi obtinguda del *Genome Browser versió Mar. 2006 (NCBI36/hg18)*.
- Fig. R.15. Estat de metilació de l'element Alu Aj2c1 en 12 parelles Normal-Tumor.
- Fig. R.16. Estat de metilació de l'element Alu Aj2c1 en mostres normals (cèl·lules mare embrionàries, cèl·lules pluripotents induïdes, fibroblasts, queratinòcits, pell, leucòcits i teixit de còlon normal) i mostres tumorals (teixit tumoral de còlon i línies de CCR).
- Fig. R.17. Estat de metilació de la regió flanquejant a l'element Alu seleccionat Aj2c1 en cèl·lules mare embrionàries, cèl·lules pluripotents induïdes, fibroblasts, queratinòcits, una parella normal-tumor i diferents línies de CCR.
- Fig. R.18. Anàlisi per ChIP de les regions flanquejants a l'element Alu parcialment metilat Aj2c1 en les línies hESC2, HFF, Sw480, LoVo, HCT116 i DKO.
- Fig. R.19. Transcrits alternatius del gen *DIEXF*.

- Fig. R.20. Nivells d'expressió de *DIEXF* en diferents línies de CCR (HCT116, HT29, LoVo i Sw480), cèl·lules mare humanes (hES2, hES4, hES6), cèl·lules pluripotents induïdes ([H]KiPS4F1, KiPS4F1 i KIPS4F8), queratinòcits (K1), fibroblasts (HFF).
- Fig. R.21. Localització de les sondes dissenyades pel Northern Blot.
- Fig. R.22. Estudi de *DIEXF* a nivell d'ARN per Northern Blot de diferents línies cel·lulars.
- Fig. R.23. Estudi de diferents sondes dissenyades arreu del gen *DIEXF* hibridades en la línia SW480.
- Fig. R.24. Posició dels encebadors de la tècnica RACE.
- Fig. R.25. 3'-RACE de les línies Sw480, hESC4 i HFF.
- Fig. R.26. UTR3'-RACE.
- Fig. R.27. Llocs poli(A) que estan descrits a en la Poly(A)\_DB del *Genome Browser versió GRCh37/hg19*.
- Fig. R.28. Seqüència UTR3' del gen *DIEXF*.
- Fig. R.29. 5'-RACE de les línies Sw480, hESC4 i HFF.
- Fig. R.30. 5'-RACE.
- Fig. R.31. Seqüència UTR5' del gen *DIEXF*.
- Fig. R.32. Imatge de l'Ensembl dels tres transcrits de *DIEXF*.
- Fig. R.33. Western Blot de les línies de CCR (DKO, HCT116, HT29, LoVo i SW48), de les cèl·lules mare humanes (hESC4) i la línia de fibroblast (HFF).
- Fig. R.34. Silenciament de *DIEXF* en la línia HCT116 utilitzant siRNA3 i siRNA4 a una concentració de 20µM.
- Fig. R.35. Western Blot de la línia HCT116 silenciada amb siRNA3 i siRNA4 a 48 hores.
- Fig. R.36. Silenciament de *DIEXF* en la línia Sw480 utilitzant siRNA3 a una concentració de 20µM.
- Fig. R.37. Western Blot de la línia Sw480 silenciada amb siRNA3 a 48 hores.
- Fig. R.38. Silenciament de *DIEXF* en la línia HCT116 utilitzant siRNA1 i siRNA2 a una concentració de 20µM.
- Fig. R.39. Western Blot de la línia HCT116 silenciada amb siRNA1 a 48 hores.
- Fig. R.40. Mapa del *Genome Browser* amb la localització de les sondes d'estudi segons el track *CpG Methylation by Methyl 450K Bead Arrays from ENCODE/HAIB*.
- Fig. R.41. Estat de metilació de les 3 regions en diferents tipus de càncer segons les dades del TCGA.
- Fig. R.42. Esquema dels diferents grups N-T segons l'estat de metilació de les sondes consultades en el TCGA.
- Fig. R.43. Posició de les sondes específiques per *DIEXF*, tant en l'array d'*Affymetrix HG-U133 Plus 2.0* com d'*Agilent Whole human genome 44k*.
- Fig. R.44. Estadística de l'expressió d'ARNm de càncer colorectal obtinguts de l'*Oncomine*, amb 237 mostres del TCGA utilitzant les sondes d'*Agilent Whole human 44K*.



- Fig. R.45. Estat de metilació de la regió flanquejant a l'element Alu ZALU3 seleccionat (en vermell) en diferents línies de CCR, queratinòcits (K18 i K37), cèl·lules mare embrionàries (hESC4 i hESC2), cèl·lules pluripotents induïdes (KiPS4F8), fibroblasts (F, HFF) i una parella normal-tumor (541N i 541T).
- Fig. R.46. Anàlisi per ChIP de les regions flanquejants a l'element Alu ZALU3 en les línies Sw480, HCT116, DKO i CaCo2.
- Fig. R.47. Anàlisi per ChIP de les regions flanquejants a l'element Alu ZALU3 en les línies HFF(fibroblasts) i hESC2 (cèl·lules mare embrionàries).
- Fig. R.48. Gen GLDC. *Genome Browser versió Feb. 2009 (GRCh37/hg19)*.
- Fig. R.49. Nivells d'expressió de *GLDC* en diferents línies de CCR (HCT116, DKO, Sw480, LoVo i HT29), una parella Normal-Tumor, fibroblasts (HFF), limfòcits (GM12878), leucòcits, hepatòcits, queratinòcits (K1), cèl·lules pluripotents induïdes derivades de K1(KiPS4F1 i KiPS4F8) i cèl·lules mare embrionàries (hESC2, hESC4).
- Fig. R.50. Imatge de l'*Ensembl* del transcrit de *GLDC*.
- Fig. R.51. Expressió de *GLDC* en les diferents línies cel·lulars per Western Blot.
- Fig. R.52. *OCT4* i *GATA6* utilitzats com a controls del correcte funcionament del estudi de diferenciació a diferents temps de cèl·lules embrionàries cap a EBs.
- Fig. R.53. Estat de metilació de la regió flanquejant a l'element Alu ZALU3 seleccionat en diferents línies de diferenciació.
- Fig. R.54. Anàlisi per ChIP de les regions flanquejants a l'element ZALU3 en els diferents temps de diferenciació.
- Fig. R.55. Expressió de *GLDC* dels diferents temps de diferenciació.
- Fig. R.56. Expressió de *GLDC* a diferents temps després d'induir la diferenciació en la línia CaCo2.
- Fig. D.1. Mapa de la regió dels diferents elements Alu analitzats obtingut a partir del *Genome Browser versió Mar. 2006 (NCBI36/hg18)*.
- Fig. D.2. Tipus de promotors múltiples establerts segons la distribució dels TSS.
- Fig. D.3. Possibles TSS segons el teixit.
- Fig. D.4. Metabolisme de la Glicina.
- Fig. D.5. Model resum del paper dels elements Alu en la regulació transcripcional segons els nostres resultats.

- TAULA I.1. Modificacions d'histones i el seu paper en la transcripció.
- TAULA I.2. Fracció del genoma de varis organismes ocupat per diferents tipus de repeticions disperses.
- TAULA I.3. Nombre de còpies i fracció del genoma humà que forma part de l'ADN repetit dispers.
- TAULA M.1. Programa de la PCR nested.
- TAULA M.2. Programa de PCR per la tècnica QUMA.
- TAULA M.3. Programa per la PCR de ZUMA.
- TAULA M.4. Anticossos utilitzats per la immunoprecipitació.
- TAULA M.5. Programa de PCR per Northern Blot.
- TAULA M.6. Components de la pcr d'amplificació del RACE ADNc 3' i 5'.
- TAULA M.7. Programa de PCR per RACE .
- TAULA M.8. Programa de PCR caracteritzar el producte del RACE.
- TAULA M.9. Anticossos emprats.
- TAULA M.10. Disseny dels siRNA.
- TAULA R.1. Comparació dels elements Alu random i virtuals.
- TAULA R.2. Elements Alu seleccionats.
- TAULA R.3. Estat de metilació dels elements Alu seleccionats en parelles de còlon normal (N) / Tumor (T).
- TAULA R.4. Estat de metilació dels elements Alu en línies cel·lulars de CCR.
- TAULA R.5. Estat de metilació dels elements Alu seleccionats en la línia HCT116 i HCT116 tractada amb 5-AzadC, TSA i AZA+TSA i la línia variant DKO.
- TAULA R.6. Estat de metilació dels elements Alu seleccionats en diferents línies cel·lulars.
- TAULA R.7. Estat de conservació dels elements Alu en diferents espècies de primats.
- TAULA R.8. Dissenys dels encebadors per l'estudi d'expressió dels diferents transcrits del *DIEXF*.
- TAULA R.9. Sondes per l'estudi d'expressió dels diferents transcrits de *DIEXF* per Northern Blot.
- TAULA R.10. Mida esperada de les amplificacions amb els encebadors del RACE.
- TAULA R.11. Expressió dels transcrits del gen *DIEXF*.
- TAULA R.12. siRNAs de silenciament i la seva posició dins del gen.
- TAULA R.13. Relació entre l'estat epigenètic de l'element Alu AJ2C1 i l'expressió de *DIEXF*.
- TAULA R.14. Relació entre l'estat epigenètic de l'element Alu ZALU3 i l'expressió de *GLDC*.

# INTRODUCCIÓ

## 1. EPIGENÈTICA

*“La diferència entre genètica y epigenètica probablement pot comparar-se amb la diferència que existeix entre escriure i llegir un llibre. Un cop el llibre ha estat escrit, el text (els gens o la informació emmagatzemada en l'ADN) serà el mateix en totes les còpies que es distribueixin entre els lectors. No obstant, cada lector podria interpretar la història del llibre d'una forma lleugerament diferent, amb les seves emocions i projeccions que poden anar canviant a mida que es desenvolupen els episodis. D'una forma molt similar, l'epigenètica permetria diferents interpretacions d'un motlle fix (el llibre o codi genètic) i resultaria en diferents lectures, depenent de les condicions variables en les que s'interpreti el motlle”.*

Thomas Jenuwein (Viena, Àustria).

Sovint s'atribueix la definició d'**epigenètica** a Conrad Hal Waddington, qui l'any 1942 la va definir com la branca de la biologia que estudia les interaccions causals entre els gens i els seus productes que donen lloc al fenotip. Les primeres anotacions en la literatura sobre epigenètica, però, daten de mitjans del segle XIX, encara que podríem dir que el concepte d'epigenètica es remunta al 384-322 ac., quan Aristòtil defensava l'epigènesi com el desenvolupament de la forma orgànica de l'individu a partir de la matèria amorfa.

Avui en dia encara no està clar fins a quin punt estem preprogramats o l'ambient juga un paper molt important. L'epigenètica apareix com a pont entre els factors genètics i ambientals a l'hora de determinar el fenotip (Aguilera et al., 2010; Fraga et al., 2005; Jaenisch and Bird, 2003).

Epigenètica (del grec *epi-*, en o sobre, i *-genètica*) fa referència, en sentit ampli, a l'estudi de tots aquells factors no genètics que intervenen en la determinació de l'ontogènia. Actualment hi ha un gran nombre de definicions per explicar que és l'epigenètica, una d'elles defineix l'epigenètica com l'estudi de canvis heretables en la funció gènica que es produeixen sense un canvi en la seqüència d'ADN (Feinberg, 2007; Jirtle and Skinner, 2007). Per contra, existeixen altres definicions en què el terme heretable queda exclòs, buscant una visió més unificadora dels esdeveniments epigenètics i definint l'epigenètica com l'adaptació estructural de les regions cromosòmiques amb la finalitat de registrar, marcar o perpetuar estats d'activitat alterats (Bird, 2007). Aquesta definició inclou la possibilitat de què els processos epigenètics actuïn com a reguladors de la variació gènica.

### 1.1. ESTRUCTURA DE LA CROMATINA

El material genètic de la cèl·lula eucariota està organitzat en una estructura complexa formada per ADN i proteïnes, localitzada en un compartiment especialitzat, el nucli. Aquesta estructura s'anomena **cromatina** (de la paraula grega "khroma", que significa colorat, i "soma" cos).

La cromatina manté l'estabilitat de la seva estructura gràcies a un seguit de proteïnes que interaccionen amb l'ADN. En qualsevol cèl·lula hi ha més de 1.000 tipus diferents de proteïnes unides a l'ADN. Dins d'aquestes proteïnes trobem un grup que inclou factors de transcripció, polimerases, receptors hormonals i altres enzims nuclears. No obstant, les histones són la principal classe de proteïnes involucrades en mantenir l'estructura compacta de la cromatina. Hi ha 5 tipus diferents d'histones identificades com H1, H2A, H2B, H3 y H4 (Khorasanizadeh, 2004).

La unitat fonamental de l'organització estructural de la cromatina es denomina **nucleosoma**. Aquesta estructura constitueix la base del primer nivell de compactació de l'ADN en la cromatina, dins del nucli interfàsic (Figura I.1).

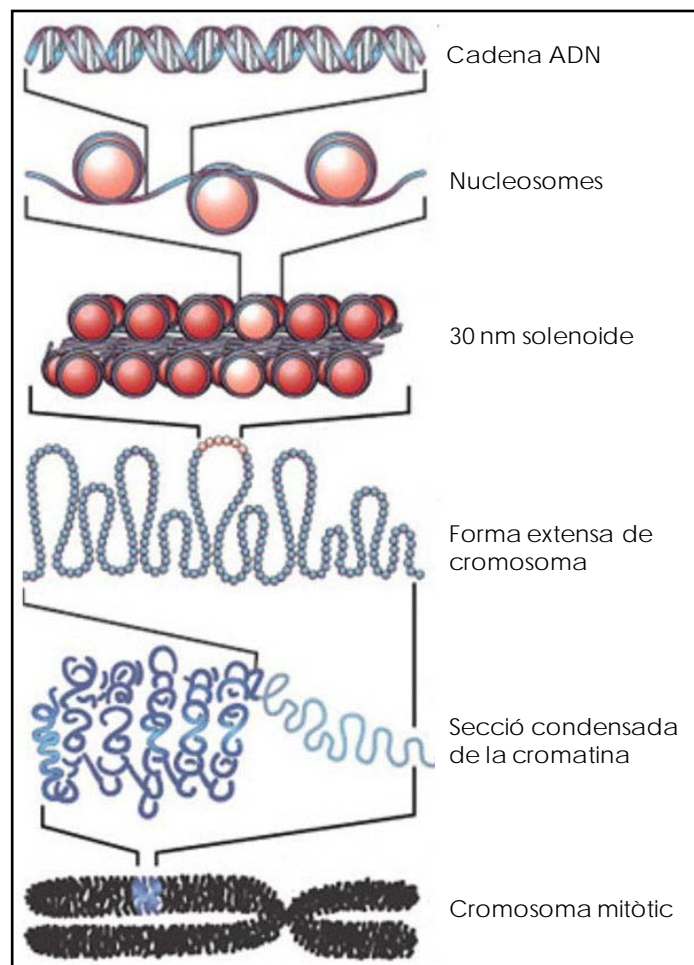


Fig. I.1. Condensació de l'ADN. L'ADN es troba compactat en el nucli. Un primer nivell de condensació es forma gràcies a la unió de les histones. L'ADN rodeja un nucli format per vuit histones. Aquest nucli més la regió d'enllaç que uneix els nuclis adjacents s'anomena nucleosoma. Cada nucleosoma conté un fragment d'ADN de 146 nucleòtids més vuit histones. Els nucleosomes es compacten un sobre l'altre formant una estructura anomenada cromatina. D'acord al grau de compactació, la cromatina es classifica en eucromatina y heterocromatina. Figura adaptada de l'article de Felsenfeld G (Felsenfeld and Groudine, 2003).

El nucleosoma consta de dues parts, una central o nucli i una regió d'enllaç o internucleosomal que uneix els nuclis adjacents (Felsenfeld and Groudine, 2003). El

nucli és una estructura nucleoproteica formada per 146 parells de bases d'ADN enrotllada en una superhèlix levogira de tres quarts de volta al voltant d'un octàmer de proteïnes. Aquest octàmer està format per dues còpies de cadascuna de les histones H3, H4, H2A i H2B, més concretament es tracta d'un tetràmer de (H3)<sub>2</sub>(H4)<sub>2</sub> i un parell de dímers H2AH2B.

Les **histones** són petites proteïnes amb carrega bàsica degut al seu alt contingut en aminoàcids d'arginina i lisina, carregats positivament. Aquesta característica permet a les histones unir-se fortament a les molècules d'ADN, carregades negativament.

Els nucleosomes estan units per la regió d'enllaç, la qual varia de 10 a 80 parells de bases de longitud depenent de l'espècie o teixit. En aquesta regió s'uneix la histona d'enllaç H1. Aquesta unió forma una subunitat de la cromatina coneguda com a cromatosoma. El **cromatosoma** està format per 166 parells de bases (dues voltes de la superhèlix) envoltant el centre octamèric d'histones i fixat per la histona H1. L'empaquetament de l'ADN amb histones en els nucleosomes produeix una fibra de cromatina o nucleofilament d'aproximadament 10 nm de diàmetre.

La cromatina es condensa encara més enrotllant-se en una fibra de 30 nm en forma de solenoide, amb uns sis nucleosomes per volta (Henikoff, 2008). Les interaccions amb la histona H1 semblen portar a terme un rol important en aquesta etapa de condensació de la cromatina per estabilitzar l'interior d'aquesta fibra. La fibra de 30 nm pot formar bucles d'ADN, cadascun contenint aproximadament 60.000 parells de bases, els quals estan fixats en la base de la matriu nuclear (Kornberg and Lorch, 1999; Lorch et al., 1999) (Figura I.1).

Dins d'un nucli amb uns pocs micròmetres de diàmetre trobem gairebé dos metres d'ADN. Malgrat aquest elevat grau de compactació, l'ADN ha de ser ràpidament accessible i permetre la seva interacció amb les proteïnes que regulen les funcions de la cromatina en la replicació, la reparació i la recombinació. Per tant, l'organització dinàmica de la cromatina té influència, de manera potencial, sobre totes les funcions del genoma (Felsenfeld and Groudine, 2003; Probst et al., 2009).

La cromatina es divideix en heterocromatina i eucromatina. L'**heterocromatina** es defineix com una estructura que no altera el seu nivell de condensació o compactació al llarg del cicle cel·lular i es localitza principalment en la perifèria del nucli. Podem diferenciar entre heterocromatina constitutiva i heterocromatina facultativa, l'heterocromatina constitutiva conté pocs gens i està formada principalment per seqüències repetitives localitzades en grans regions coincidents amb centròmers i telòmers, i l'heterocromatina facultativa comparteix la mateixa estructura compacta que l'heterocromatina constitutiva però pot esdevenir transcripcionalment activa en resposta a senyals del desenvolupament o ambientals específiques com el cromosoma X inactiu de mamífers (Grewal and Jia, 2007). Per altra banda trobem l'**eucromatina**, la qual és rica en gens i sovint és transcripcionalment activa. Es descondensa durant la interfase i es localitza en l'interior del nucleoplasma (Bannister and Kouzarides, 2011).

En mamífers, la demarcació entre aquests diferents estats de cromatina ve delimitat per elements barrera (Gaszner and Felsenfeld, 2006). Per exemple, la proteïna CTCF

s'uneix a elements aïllants de l'ADN o "insulators" impedit que la heterocromatina s'expandeixi cap a regions d'euromatina (Herold et al., 2012) (Figura 1.2).

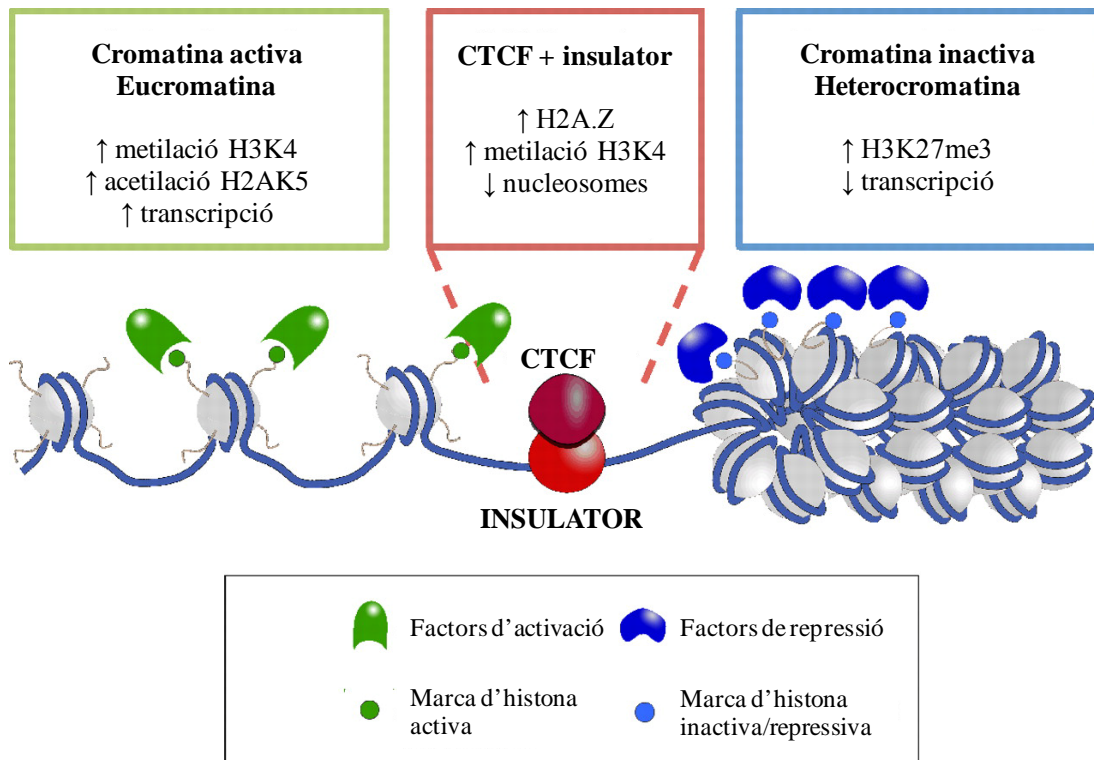


Fig. 1.2. CTCF s'uneix a regions de l'ADN anomenades insulators actuant com a barrera i evitant així que l'heterocromatina s'expandeixi cap a l'euromatina. Figura adaptada de l'article (Herold et al., 2012).

## 1.2. MECANISMES EPIGENÈTICS

La informació epigenètica modula l'expressió dels gens sense alterar la seqüència d'ADN (Feinberg, 2007; Jirtle and Skinner, 2007). L'**epigenoma** és la informació epigenètica global d'un organisme. I els estats epigenètics estan dirigits per un gran nombre de senyals entre les que participen factors de transcripció, ARN no codificants, metilació de l'ADN i modificacions d'histones (Bannister and Kouzarides, 2011; Bernstein et al., 2006; Bonasio et al., 2010; Feinberg; Kouzarides, 2007). Els canvis epigenètics succeeixen de manera distinta entre diferents teixits i individus (Bernstein et al., 2010; Feinberg, 2007).

L'epigenètica ha descobert grans incògnites de la biologia; per exemple la capacitat d'un mateix genoma de regular més de 200 tipus cel·lulars diferents; el fet que una cèl·lula es pugui reprogramar en una cèl·lula tumoral i certs funcionaments cerebrals (Bonasio et al., 2010); la relació entre bessons monozigòtics, els quals són considerats genèticament idèntics ja que provenen d'un mateix zigot però en canvi presenten diferències epigenètiques que es manifesten fenotípicament, presentant diferent

susceptibilitat a patir certes malalties així com característiques antropomòrfiques diferents (Fraga et al., 2005; Wong and Akerley, 2005).

En els últims anys s'han destinat grans esforços en l'obtenció de mapes epigenòmics de diferents tipus cel·lulars. Tècniques com ChiP-seq (assajos d'immunoprecipitació de cromatina seguits d'ultraseqüenciació) i la seqüenciació massiva de ADN transformat per bisulfit han permès crear mapes d'alta resolució, especialment en cèl·lules mare embrionàries, i diferents tipus cel·lulars en humans (Feinberg, 2010b; Li et al., 2010; Lister et al., 2009). En l'actualitat han sorgit nous projectes (com el *NIH Roadmap Epigenomics Mapping Consortium* i el *Human Epigenome Project*) (Satterlee et al., 2010) o l'*ENCODE Project Consortium (Encyclopedia of DNA elements)* (Duggan et al., 1994) per crear mapes de metilació de l'ADN, modificacions d'histones i altres components de la cromatina en diferents grups cel·lulars i teixits.

Ens centrarem en els dos tipus principals d'informació epigenètica, que són metilació de l'ADN i les modificacions d'histones.

### 1.2.1. LA METILACIÓ DE L'ADN

La **metilació de l'ADN** consisteix en l'addició d'un grup metil al carboni de la posició 5 de l'anell de pirimidina de la citosina. Les ADN metiltransferases, DNMTs (de l'anglès *DNA methyltransferases*) són els enzims responsables de catalitzar aquesta reacció, on la S-adenosil-metionina (SAM) és la molècula donadora dels grups metil (Goll and Bestor, 2005; Herman and Baylin, 2003; Ji et al., 2010) (Figura 1.3). Les DNMTs s'explicaran en més detall en el següent apartat.

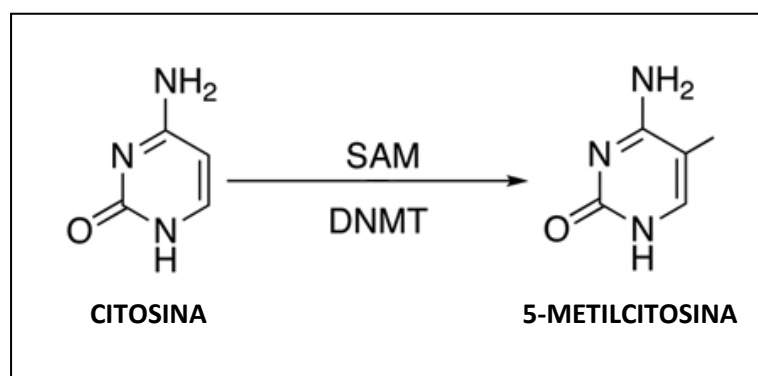


Fig. 1.3. Metilació de la citosina. Les DNMTs catalitzen la reacció i SAM es la molècula donadora del grup metil. El producte final és la 5-metilcitosina (5-mC).

Aquesta modificació pot ser heretada a través de la divisió cel·lular. La metilació de l'ADN sol tenir lloc en la formació del zigot i es restableix a través de les successives divisions cel·lulars durant el desenvolupament. Es tracta d'un procés fonamental pel desenvolupament normal dels organismes i la diferenciació cel·lular en organismes superiors. La metilació de l'ADN estable altera el patró d'expressió gènica en les cèl·lules, programant-les durant el desenvolupament embrionari i fixant la seva funció al llarg de tota la vida de l'organisme (Bird, 2002; Meissner et al., 2008).



Per altra banda la metilació de l'ADN suprimeix l'expressió de gens virals i altres elements nocius que es van incorporant en el genoma al llarg del temps. A més, la metilació també constitueix la base de l'estructura de la cromatina i permet que les cèl·lules es diferenciïn funcionalment les unes de les altres partint d'una mateixa seqüència d'ADN. La metilació de l'ADN és essencial pel desenvolupament normal i s'associa amb un seguit de processos clau com la imprompta genètica, la inactivació del cromosoma X i la supressió d'elements repetitius (Smith and Meissner, 2013). Com és d'esperar la metilació de l'ADN també juga un paper fonamental en la tumorigènesi (Esteller, 2008; Jones and Baylin, 2002; Portela and Esteller, 2010).

En humans i altres mamífers aquesta modificació epigenètica es troba principalment en les citosines que precedeixen una guanina, les citosines del dinucleòtid CpG (Clark, 2007). Pel què fa a les cèl·lules mare embrionàries, una quarta part de la seva metilació es troba en un context de no-CpG, indicant que les cèl·lules mare embrionàries potser segueixen uns mecanismes de metilació diferents que afecten a la regulació gènica. La metilació en el context no-CpG desapareix durant la diferenciació i reapareix en les cèl·lules mare pluripotents induïdes (Lister et al., 2009). Normalment trobem metilació en no-CpG dins de gens i no en llocs d'unió a proteïna ni *enhancers*.

Els patrons de metilació en context CpG i no-CpG en els gens i la seva diferent correlació amb l'expressió marquen un rol diferent de metilació de l'ADN per a cada context. Així doncs el fet pel qual el context no-CpG sigui exclusiu de les cèl·lules mare degut a l'activitat constant de la metiltransferasa de novo, suggereix el seu possible rol en l'origen i conservació del seu estat pluripotent (Laurent et al., 2010).

Per altra banda, un conjunt d'evidències han suggerit que la metilació de l'ADN en cèl·lules de mamífer pot ser reversible. Recentment s'ha vist que hi ha certs enzims capaços d'intervenir en la desmetilació de l'ADN i coincidint amb això s'ha vist que les citosines poden hidroximetil·lar-se a 5-hmC. L'oxidació de la 5-mC, catalitzada per TET1, dona com a producte la 5-hidroximetilcitosina (5-hmC) (Figura I.4).

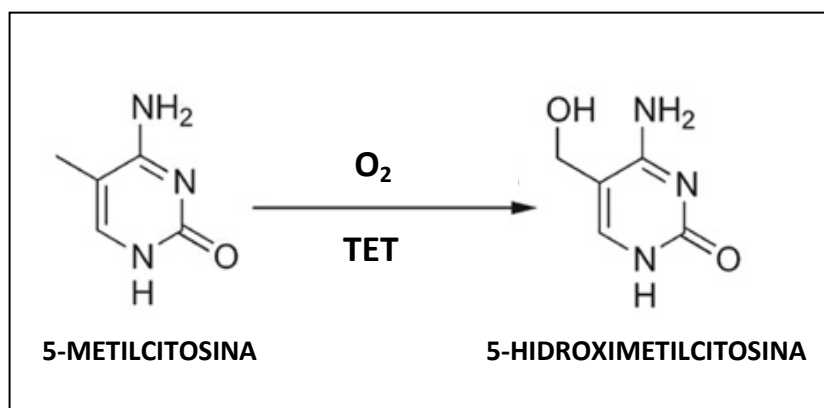


Fig. I.4. Hidroximetilació de la 5 metilcitosina. Els enzims TET catalitzen la reacció i  $\text{O}_2$  és la molècula donadora del grup hidroxil. El producte final és la 5-hidroximetilcitosina.

Inicialment la hidroxilació semblava una possible alternativa per l'activació dels gens silenciats per la metilació, però estudis recents han refutat aquesta teoria. Així doncs sembla que la metilació i desmetilació de l'ADN és un procés de doble sentit en la que intervenen múltiples vies. Aquesta nova modificació podria formar part d'un mecanisme regulador de la desmetilació de l'ADN; per tant, la presència de 5-hmC evitaria l'acció de les DNMTs sobre les illes CpGs (Tahiliani et al., 2009; Valinluck and Sowers, 2007). Sembla que 5-hmC actuaria com un intermediari per l'eliminació de citosines metilades ja sigui passivament donat que la presència de 5-hmC afecta a la remetilació per DNMTs quan les cèl·lules es divideixen o bé de manera activa a través de la reparació de l'ADN en absència de divisió cel·lular (Bhutani et al., 2011).

#### 1.2.1.1. Les ADN metiltransferases (DNMTs)

Els enzims responsables de metilar el ADN són les **ADN metiltransferases** (DNMTs). La DNMT1 és la metiltransferasa de manteniment, s'encarrega de copiar els patrons de metilació de l'ADN a les cadenes filles durant la replicació de l'ADN.

Per altra banda, la DNMT3a i DNMT3b són metiltransferases de *novo* ja que afegeixen grups metil en posicions totalment noves, canviant el patró de metilació en una regió localitzada del genoma, per exemple estableixen patrons de metilació en les primeres etapes del desenvolupament. En general es considera que la metilació és un procés unidireccional, d'aquesta manera, quan una CpG adquireix una metilació de *novo*, aquesta modificació es fa estable i és heretada com un patró de metilació clonal. La DNMT3L és una proteïna homòloga a la DNMT3a, però no té activitat catalítica. En canvi quan la DNMT3L s'uneix a les metiltransferases de *novo* i augmenta la capacitat d'unió a l'ADN.

Per últim, la DNMT2 (TRDMT1) s'ha descrit com un homòleg que conté tots els motius de les metiltransferases però que no metila la citosina-5 de l'ADN sinó que metila la citosina-38 del llaç anticodó de transferència de l'àcid aspàrtic de l'ARN (Robertson and Wolffe, 2000).

La metilació de l'ADN pot afectar a la transcripció dels gens de dues maneres. En primer lloc, la metilació de l'ADN pot impedir la unió de les proteïnes responsables de la transcripció al promotor i en segon lloc, i probablement la més important, l'ADN metilat pot reclutar proteïnes amb un domini d'unió a CpG metilades, MBD (de l'anglès *Methyl-CpG binding domain*). Les proteïnes d'unió a grups metil-CpG són proteïnes que, juntament amb les HDAC i altres proteïnes de la remodelació de la cromatina, poden modificar les histones i compactar la cromatina, inactivant-la i per tant silenciant-la (Figura 1.5).

Aquest vincle entre la metilació de l'ADN i l'estructura de la cromatina és molt important. L'activitat de manteniment de la metilació és necessària per preservar la metilació de l'ADN després de cada cicle de replicació cel·lular. Sense l'ADN metiltransferasa (DNMT), la maquinària de replicació produiria ADN metilat però amb el temps s'aniria produint una desmetilació passiva.

Degut a que molts gens supressors de tumors són silenciats per la metilació de l'ADN durant la carcinogènesi s'han fet molts estudis per revertir aquest pas mitjançant la

inhibició de les DNMTs. La **5-aza-2'-desoxicitidina** és un anàleg que inhibeix les DNMTs arrestant-les en un complex covalent amb l'ADN, impedit l'eliminació de la catàlisi i donant lloc a enzims de degradació. No obstant aquest anàleg, per ser actiu, cal que s'incorpori en el genoma de la cèl·lula i això pot causar mutacions en les cèl·lules filles. Per altra banda la 5-aza-2'-desoxicitidina és tòxica per la medul·la òssia fent que el seu avantatge terapèutic es vegi minvat (Liu et al., 2003; Robertson et al., 2000).

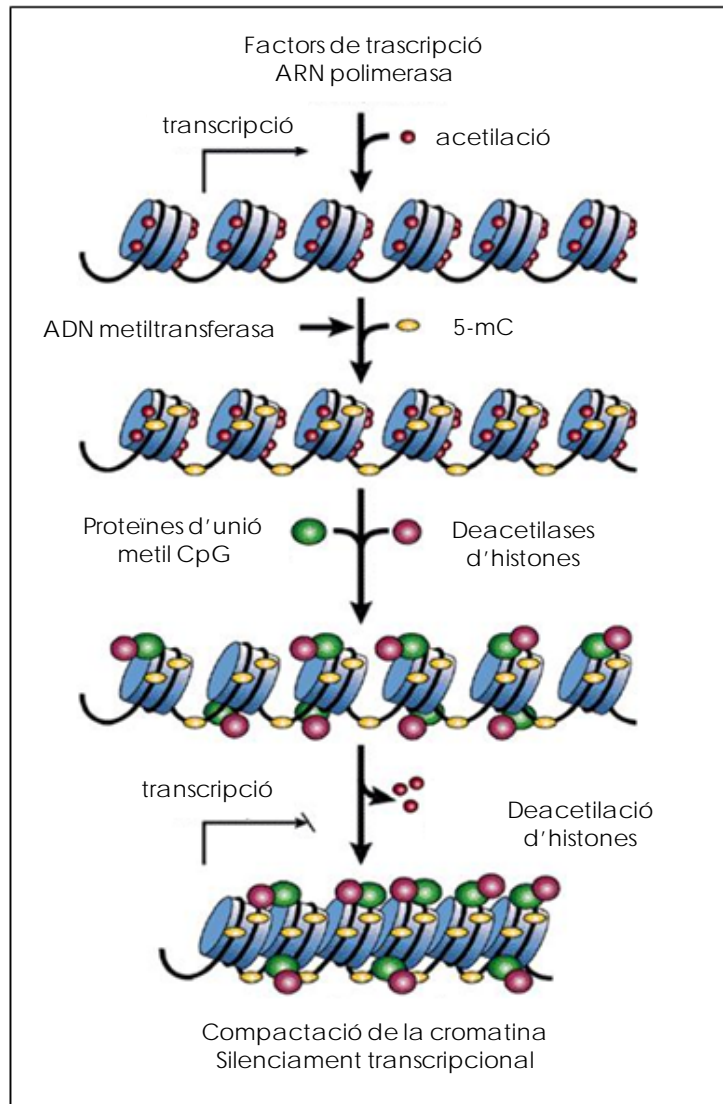


Fig. 1.5. Procés de silenciament d'una regió transcripcionalment activa per mitjà de la metil·lació de l'ADN. Primer l'ADN metiltransferasa metila l'ADN, després s'hi uneixen les deacetilases d'histones (HDAC) i les proteïnes d'unió metil-CpG. Després la DNMT1 interacciona directament amb la HDAC i té lloc l'empaquetament de la cromatina, la seva compactació i el conseqüent silenciament de la transcripció. Figura adaptada de l'article de Robertson (Robertson and Wolffe, 2000).

### 1.2.1.2. Metilació en les illes CpG

Com ja s'ha comentat en humans i altres mamífers la metilació principalment la trobem en les citosines del dinucleòtid CpG, poc freqüent en el genoma de mamífers tret de les regions denominades illes CpG (CpGi) (Clark, 2007). En els éssers humans la freqüència calculada del dinucleòtid CpG és relativament baixa, això es deu a un fenomen referit com a supressió de CpG. Aquesta supressió és característica de tots els genomes que tenen metilació de la citosina i pot estar relacionada amb la hipermutabilitat de les citosines metilades (Polak and Arndt, 2008).

Entre el 60% i el 90% de totes les CpG està metilades en mamífers. La majoria de CpG no metilades s'agrupen formant les illes CpG que estan presents en els extrems 5' de molts gens, és a dir, en les regions promotores (Goll and Bestor, 2005).

La definició d'**illa CpG** ha generat força controvèrsia. La primera va venir de la mà de Gardiner-Garden i Frommer en 1987 (Gardiner-Garden and Frommer, 1987). Posteriorment Takai i Jones van refer la definició d'illa CpG, considerant com a CpGi a aquella regió amb almenys 300 pb, amb un percentatge de Citosines i Guanines enllaçades per fosfats major de 55% i un promig de CpG observat/esperat major de 0.65 (Takai and Jones, 2002).

Les illes CpG conformen aproximadament el 76% de promotors dels gens de mamífers (Goll and Bestor, 2005). En la majoria de casos, les CpG de les illes CpG estan desmetilades si els gens estan expressats. És conegut que la metilació dels llocs CpG en les regions promotores dels gens poden inhibir l'expressió d'un gen (Figura 1.6 a).

La metilació de l'ADN no apareix exclusivament en illes CpG. Més recentment s'han definit els elements anomenats "**CpG island shores**"; aquest terme serveix per designar aquelles regions amb menys densitat de CpG i que es troben molt adjacents a una illa CpG (~2kb) (Irizarry et al., 2009) (Figura 1.6 b).

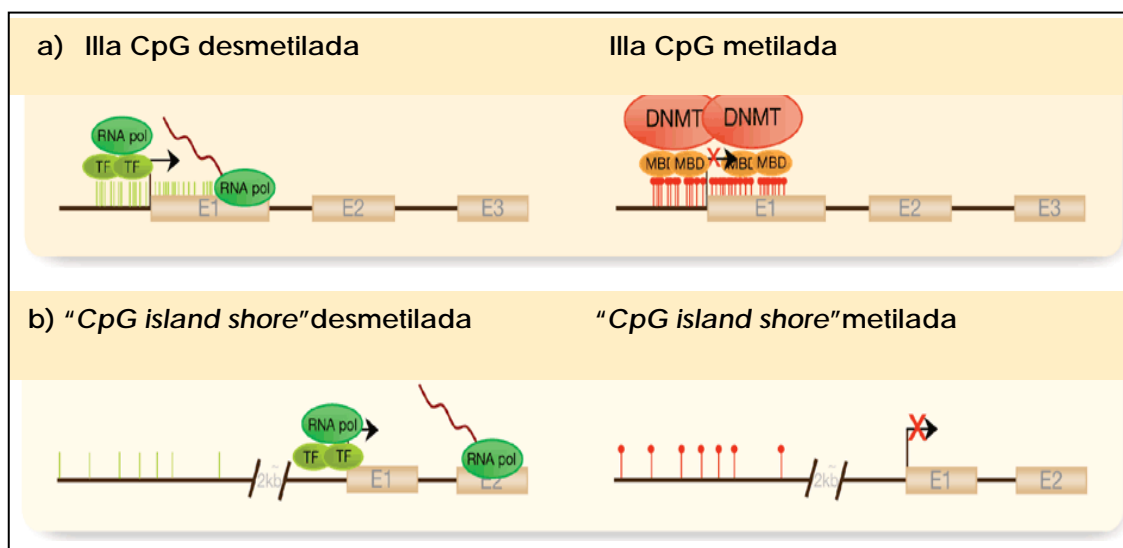


Fig. 1.6. Patró de metilació en illes CpG. a) Les illes CpG situades en regions promotores de gens estan normalment desmetilades, permetent la transcripció del gen. Una hipermetilació de l'illa s'associa a la inactivació transcripcional. b) El mateix patró s'observa en les "CpG island shores". Figura adaptada de l'article de Portela i Esteller (Portela and Esteller, 2010).

La metilació de les "CpG island shores" està estretament associada amb la inactivació transcripcional. La major part de la metilació de l'ADN específica de teixit no sembla ocórrer en les illes CpG sinó en "CpG island shores". Les "CpG island shores" diferencialment metilades són suficients per distingir entre teixits específics i es conserven entre l'ésser humà i el ratolí. D'altra banda, el 70% de les regions diferencialment metilades durant la reprogramació s'associen a "CpG island shores" (Portela and Esteller, 2010).

## 1.2.2. FUNCIONS DE LA METILACIÓ DE L'ADN

En aquest apartat comentarem els diferents processos en els que participa la metilació de l'ADN. Com ja hem comentat, la metilació de l'ADN és essencial pel desenvolupament i s'associa amb un seguit de processos clau com la imprompta genètica, la inactivació del cromosoma X i el silenciament dels elements repetitius (Smith and Meissner, 2013).

### 1.2.2.1. Imprompta genètica

La **imprompta** o *imprinting* es manifesta només en organismes superiors. Quan parlem d'*imprinting*, ens referim a gens que poden modificar el seu funcionament sense necessitat d'un canvi en la seqüència d'ADN. Aquest canvi en la seva forma de manifestar-se que tenen els gens "imprintats" es deu, generalment, al seu origen parental. Un gen imprès o "imprintat" es manifesta d'una manera quan el seu origen és patern i d'una altra quan prové de la gàmeta materna. Sembla que hi ha un mecanisme cel·lular que d'alguna manera marca o deixa una imprompta sobre tots els gens "imprintables" segons el sexe de l'individu (Surani, 2001).

En els organismes diploides, les cèl·lules somàtiques tenen dues còpies del genoma. Per tant, cada gen autosòmic està representat per dues còpies o al·lels, cadascuna d'elles heretada d'un progenitor en la fertilització. En la gran majoria dels gens autosòmics, l'expressió d'ambdós al·lels succeeix simultàniament. No obstant, una petita proporció dels gens (<1%) està "impresa", és a dir, que la seva expressió depèn de només un dels al·lels. L'expressió de l'al·lel depèn, per tant, del seu origen parental. Un gen imprintat, és funcionalment haploide, ja que té un dels dos al·lels silenciats, eliminant la protecció que confereix ser diploide contra mutacions deletèries recessives. A més, la seva expressió pot ser desregulada epigenèticament. Per tant, aquests gens representen locus susceptibles de ser alterats funcionalment, tant genèticament com epigenèticament.

Actualment, existeixen mètodes per predir l'estat d'*imprinting* del genoma, diferenciant entre gens expressats de manera monoal·lèlica o bial·lèlica. Trobem estudis de determinades regions on l'imprompta genètica és ben coneguda i ens permeten conèixer certes característiques típiques d'aquests gens, aquest és el cas de la banda 11p15.5 que presenten gens imprintats tals com H19 i IGF2 (Filippova, 2008; Wallace and Felsenfeld, 2007) (Figura 1.7).

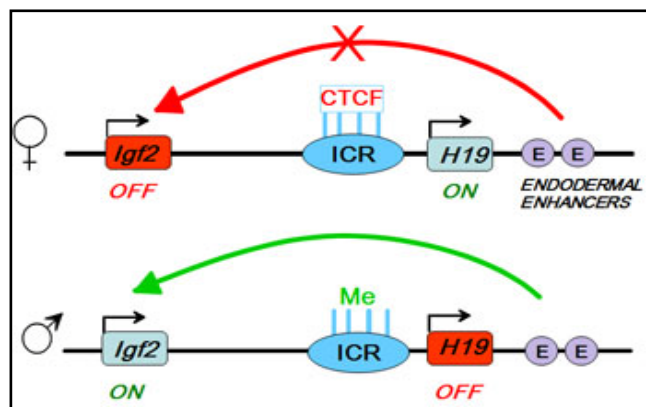


Fig. 1.7. CTCF regula la imprompta genètica del locus IGF2/H19. En l'al·lel matern CTCF s'uneix l'ADN no metil·lat de la regió control de la imprompta, el ICR (de l'anglès *Imprinting Control Region*), impedit així l'expressió de IGF2. Aquest només s'expressa a partir de l'al·lel patern, el ICR del qual està metil·lat i CTCF no s'hi pot unir, permetent l'activació de IGF2 pels enhancers (Wallace and Felsenfeld, 2007).

### 1.2.2.2. Inactivació del cromosoma X

En les cèl·lules somàtiques del sexe femení s'ha observat que només un dels dos cromosomes està actiu, l'altre roman inactiu i apareix en les cèl·lules en interfase com un cos dens en la perifèria del nucli anomenat **cos de Barr**.

La inactivació del cromosoma X és el silenciament a l'atzar d'un dels dos cromosomes X en totes les cèl·lules somàtiques de les femelles en mamífers (Park and Kuroda, 2001). Aquesta inactivació té lloc en l'estat de mòrula i es completa al final de la primera setmana del desenvolupament embrionari. La selecció del cromosoma X que s'inactivarà és un fenomen generalment aleatori tenint en compte que, al donar-se la fecundació, cada cromosoma X té un origen patern i matern, en unes cèl·lules s'inactivarà el matern i en altres el patern.

En el procés d'inactivació del cromosoma X participa el centre d'inactivació X (XIC) (de l'anglès *X inactivation complex*), la funció de XIC depent de XIST, un gen que sintetitza un ARN no codificant. XIST només s'expressa en el cromosoma X inactiu. Això desencadena una condensació de la cromatina fent que el cromosoma X inactiu adopti una conformació d'heterocromatina (Lee and Bartolomei, 2013).

Com s'ha comentat abans, CTCF juga un paper fonamental en la imprompta gènica i en la inactivació del cromosoma X (Feinberg, 2010a; van Steensel and Dekker, 2010) (Filippova, 2008).

### 1.2.2.3. Silenciament dels elements repetitius

El genoma dels mamífers és complex i a part de material codificant també conté transposons i d'altres elements repetitius, anomenats ADN no codificant o *junk DNA* (ADN de rebuig), que el genoma ha anat adquirint al llarg del temps. Aquests

elements repetitius formen gran part de les regions intergèniques i intròniques de l'ADN (Kazazian, 2004).

Aquests elements solen estar metilats i per tant silenciats, per evitar la seva expansió i poder preservar la integritat del genoma (Robertson and Wolffe, 2000). No obstant, moltes d'aquestes regions s'hipometilen durant la tumorigènesi, permetent la seva expansió. La metilació dels elements repetitius s'explicarà amb més detall en l'apartat 2.1.

#### 1.2.2.4. Altres implicacions de la metilació de l'ADN

Cal mencionar que tant l'augment com la disminució de la metilació de l'ADN estan associades al procés d'envelliment. Les evidències indiquen que aquests canvis en la metilació dependents de l'edat estan relacionats amb el desenvolupament de desordres neurològics, autoimmunitat i càncer en persones d'edat avançada (Rodenhiser and Mann, 2006).

Es sap que al llarg de la vida, l'ésser humà, acumula mutacions en la seqüència d'ADN i que aquestes mutacions es donen en una proporció d' 1 de cada 100.000 parells de bases. Aquesta freqüència de mutació no és suficient per explicar les enormes disfuncions cel·lulars i tissulars en les persones d'edat avançada. Existeixen processos epigenètics durant l'envelliment que permeten explicar aquest fenomen (Poulsen et al., 2007).

La metilació de l'ADN també té un paper important en els processos tumorigènics així com en la diferenciació. Sobre aquests dos processos parlarem més endavant.

#### 1.2.3. MODIFICACIONS D'HISTONES

Els factors que promouen l'assemblatge de la cromatina també actuen durant la fase de maduració de la mateixa per organitzar i mantenir un determinat estat. Aquests factors poden induir canvis en la conformació de la cromatina, tant a nivell de nucleosoma com de manera global sobre grans dominis (Berger, 2007). Aquest factors poden ser de dos tipus, els que formen part de la denominada maquinària de remodelació de la cromatina i que necessiten d'energia en forma d'ATP, i els que catalitzen les modificacions postraduccionals de les histones (Becker and Horz, 2002).

Un dels nivells de regulació epigenètica més importants en les cèl·lules eucariotes està constituït per les modificacions postraduccionals de les histones. En els darrers anys s'ha definit aquest conjunt de modificacions com a **codi d'histones** (Strahl and Allis, 2000). Les histones poden modificar-se per fosforilació, acetilació, metilació, ubiquitinació i sumoïlació, entre d'altres, de múltiples residus com lisines, serines o arginines (Figura 1.8).

Les marques d'histona es poden classificar en 2 grups depenent de la mida del grup químic que es modifica (Kouzarides, 2007) (Taula 1.1). Aquests residus es troben en la part amino terminal, la qual interacciona directament amb l'ADN (Felsenfeld and Groudine, 2003). En el primer grup destaquen les acetilacions de lisines(K), fosforilacions de serines(S) i metilacions de lisines i arginines (R); un segon grup es troba format per la

ubiquitinació de lisines i la sumoilació de lisines, entre d'altres. Ens centrarem més en el primer grup que engloba acetilació, fosforilació i metilació (Taula I.1) (Figura I.8).

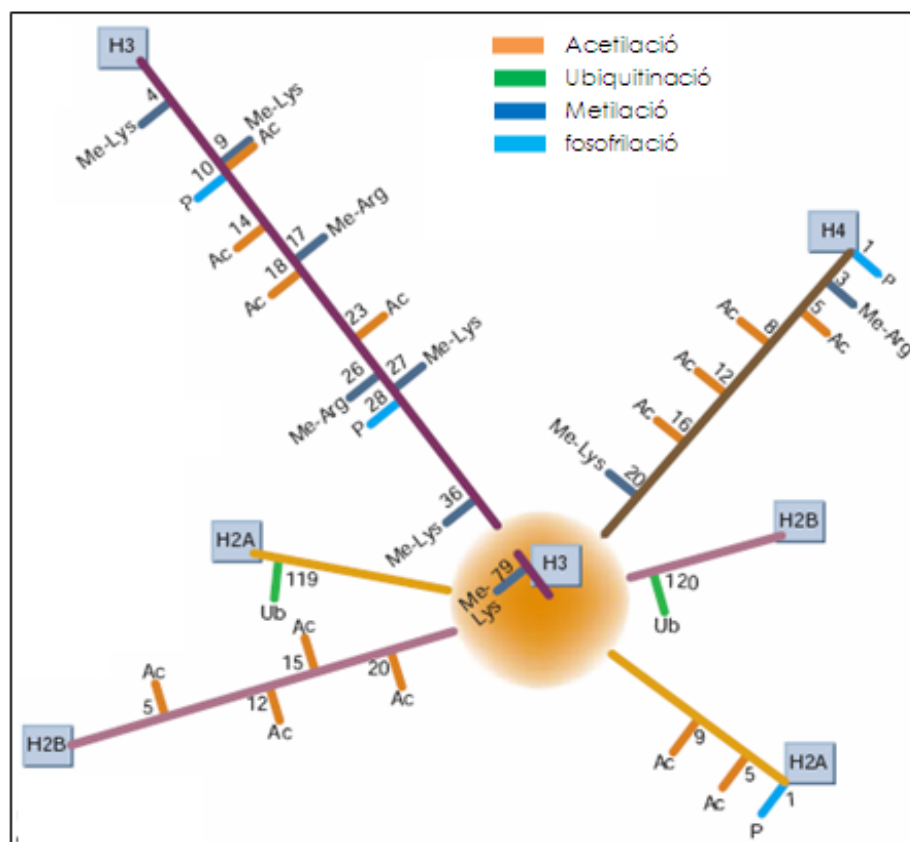


Fig. I.8. Modificacions d'histones. Cada modificació està representada per un color diferent i, a més, s'indica la posició de l'aminoàcid modificat. Figura adaptada de l'article de Felsenfeld G. (Felsenfeld and Groudine, 2003).

Les modificacions d'histones solen mantenir-se estables durant la divisió cel·lular i el seu paper en la regulació de la transcripció el poden desenvolupar en *cis*, canviant l'estructura de la cromatina, o en *trans*, dificultant o reclutant la unió de determinades proteïnes. Les modificacions en les histones són efectuades per famílies d'enzims que, generalment, s'associen a complexos activadors o repressors de la transcripció (Bonasio et al., 2010).

Algunes marques solen associar-se a l'expressió gènica, com l'acetilació de la lisina 9 en la histona 3 (AcH3K9) o la trimetilació de la lisina 4 en la histona H3 (H3K4me3), i d'altres s'associen a la repressió transcripcional, com la trimetilació de la lisina 9 en la histona 3 (H3K9me3) i la trimetilació de la lisina 27 en la histona 3 (H3K27me3). D'altra banda trobem algunes modificacions d'histones associades a diferents situacions funcionals depenent del context en el que es troben (Portela and Esteller, 2010).



TAULA 1.1. MODIFICACIONS D'HISTONES I EL SEU PAPER EN LA TRANSCRIPCIÓ

	modificació	paper	Residus modificats	enzims
GRUP I	Acetilació	activació	H3 (K9,K14,K18,K56) H4 (K5, K8, K12,K16) H2A H2B (K6,K7,K16,K17)	HAT(GNAT,MYST,CBP/p300) HDAC (tipus I, II, III I Sir-2)
	Fosforilació	activació	H3 (S10)	quinases
	Metilació	activació	H3 (K4,K36,K79)	HKMTs
			H4 (R3) H3 (R2,R17,R26)	PRMT1,PRMT4, CARM1
GRUP II	Ubiquitinació	activació	H2B (K123)	Ubp8, Ubp10
		repressió	H2A (K119)	Bim1/ring1A
	Sumoilació	repressió	H3 (?) H4 (K5, K8, K12,K16) H2A (K126) H2B (K6,K7,K16,K17)	HKMTs

### 1.2.3.1. Acetilació

L'**acetilació** neutralitza la càrrega de l'aminoàcid, disminuint la força d'unió a l'ADN i fent-lo accessible a factors de transcripció, i també origina un lloc d'unió pel bromodomini, present en moltes proteïnes associades a cromatina. L'acetilació és el resultat d'un equilibri entre dos activitats oposades, l'acetiltransferasa d'histones (HAT, de l'anglès *Histone Acetyltransferase*) i la deacetilasa d'histones (HDAC, *Histone Deacetylase*) (Kouzarides, 2007) (Taula 1.1).

Les HAT poden acetilar residus de lisina en les histones H2A, H2B, H3 i H4, disminuint així la interacció entre l'ADN i les histones, però el fet de tenir diferents dominis catalítics fa

que tinguin diferent especificitat. Això permet una mobilització i reordenament del nucleosoma facilitant l'accés de la maquinària de la transcripció a l'ADN i l'activació de l'expressió d'un gen. Moltes de les proteïnes que juguen un paper en la regulació de la transcripció tenen una activitat acetiltransferasa d'histones específica (Allis et al., 2007).

De manera oposada a les HAT, les HDAC s'encarreguen d'eliminar els grups acetil dels residus de lisina en les histones. Les deacetilases d'histones han estat descrites com components de complexos multiproteïcs associats a cromatina en estat de repressió. Les HDAC es classifiquen en base a la identitat de la seqüència i a l'organització dels dominis. Trobem doncs HDAC de tipus I, tipus II i tipus III, independents de cofactors i HDAC de tipus Sir-2 dependents del cofactor NAD com a part del seu mecanisme catalític (Kouzarides, 2007). També dins d'aquests complexos es troben membres dels factors de remodelació de la família Mi-2, enllaçant d'aquesta manera el remodelatge dels nucleosomes i la deacetilació d'histones durant la repressió duta a terme per la cromatina.

### 1.2.3.2. Fosforilació

Les diferents histones són suseptibles de ser fosforilades en residus de serina o treonina (Taula I.1). De forma similar a l'acetilació, la fosforilació, catalitzada per les quinases, també canvia la càrrega de la histona, facilitant la transcripció de l'ADN. És ben conegut que les quinases regulen les senyals de traducció de la superfície cel·lular, a través del citoplasma i dins del nucli, és el cas de la Snf1, una quinasa d'histones implicada en l'activació transcripcional.

La **fosforilació** d'histones s'ha relacionat amb diversos processos biològics com condensació de cromosomes en mitosi i meiosi, reparació de l'ADN i processos d'activació o repressió transcripcional, apoptosi o mort cel·lular programada (Peterson and Laniel, 2004). Per exemple, la fosforilació d'H1 i H3 en la serina 10 durant la mitosi i la fase S són marques de cromatina mitòtica condensada.

A més, s'ha observat que la fosforilació pot desplaçar proteïnes d'unió a la cromatina, com la HP1 quan hi ha fosforilació de la serina 10 de la histona 3. Pel que fa a reparació de l'ADN s'ha vist que la quinasa ATM es reclutada per l'ADN danyat i fosforila l'histona H2AX.

Al igual que d'altres modificacions, la fosforilació actua conjuntament amb altres modificacions, generant patrons que són reconeguts per diferents complexos a fi de dur a terme diferents processos (Zhang and Reinberg, 2001).

### 1.2.3.3. Metilació

La **metilació** d'histones juga un paper funcional molt important. La metilació es pot donar en arginines o lisines i, a més, pot haver-hi diferents estats de metilació: mono, di o tri, aquest últim només en lisines. Aquesta modificació pot representar activació o silenciament segons la posició del residu en la histona (Shi, 2007) (Taula I.1).

La metilació en la lisina 4 ha estat llargament associada a activació transcripcional i la trobem en regions eucromàtiques, sobre gens actius o activables. Els diferents estats de metilació (mono, di o trimetilació) tenen distribucions diferents i han estat associats a diferents estats de transcripció.

La metilació en la lisina 9 i el seu rol en el silenciament han estat àmpliament estudiada, es troba implicada en la formació d'heterocromatina pericentromèrica. En regions eucromàtiques s'observa un enriquiment de lisina 9 metilada en promotors, sobretot en les formes mono i dimetilació (la forma trimetilada queda més aviat restringida a regions heterocromàtiques) (Kouzarides, 2007; Martin et al., 2005).

De la metilació en l'H4K20 es coneix poc sobre el seu paper en el control de la transcripció. Tant H4K20me2 com a H4K20me3 es troben presents en la regió pericentromèrica de l'heterocromatina. A més, sembla ser que la metilació d'H3K9 requereix de la metilació d'H4K20. La monometilació d'aquest element està implicat en la mitosi.

Una modificació que actua com a repressor de gens és la metilació de l'H3K27. Aquesta modificació la podem trobar en tres llocs diferents de la cèl·lula: en la regió pericentromèrica de l'heterocromatina, en la part inactiva del cromosoma X en mamífers i en la regió d'eucromatina on hi ha elements de resposta Polycomb (PREs de l'anglès *Polycomb Response Elements*). La metilació d'H3K27 s'ha trobat vinculada al silenciament d'oncogens (Martin et al., 2005).

El procés contrari, la desmetilació, no s'ha detectat en arginines, mentre que en lisines hi ha descrits diversos dominis catalítics (LSD1, JMDM).

#### 1.2.3.4. El codi d'histones i la conformació de la cromatina

L'existència de dos modificacions possibles (acetilació i metilació) en el mateix aminoàcid (lisina 9 de la cua aminoterminal de la H3, per exemple) és un bon exemple que dóna suport a la hipòtesi del "codi d'histones" que determina el grau d'activitat i la conformació de la cromatina en funció de les modificacions que trobem en les histones (Figura 1.8).

L'eucromatina es caracteritza per l'alt grau d'acetilació d'histones així com per la trimetilació de les lisines 4, 36 i 79 en la histona 3 (H3K4me3, H3K36me3, H3K79me3), totes elles associades a la transcripció gènica. En canvi l'heterocromatina està formada per histones poc acetilades però amb alt grau de metilació de les lisines 9 i 27 en la histona 3 (H3K9me3 i H3K27me3) i de la lisina 20 en la histona 4 (H4K20me3).

Finalment, els gens actius transcripcionalment tenen histones riques en trimetilació de la lisina 4 en la histona 3 (H3K4me3), acetilació de la lisina 27 en la histona 3 (AcH3K27), acetilació de la lisina 5 en la histona H2b (AcH2BK5) en la zona promotora i marques de metilació de la lisina 79 en la histona 3 (H3K79me1) i metilació de la lisina 20 en la histona 4 (H4K20me3) al llarg del cos genètic (Portela and Esteller, 2010).

Cal tenir en compte que aquests mecanismes epigenètics participen conjuntament en la regulació de l'expressió gènica. Per tant, si el gen està silenciament per la metilació de

l'ADN de l'illa CpG del seu promotor, les histones tindran marques de repressió (H3K27me3, per exemple) i la cromatina estarà condensada (heterocromatina). D'altra banda, quan les illes CpG del promotor no estan metilades a nivell d'ADN i hi ha expressió, les histones presenten marques d'activació (acetilació, fosforilació i/o metilació dels aminoàcids esmentats abans). Així doncs, les proteïnes vinculades amb la metilació de l'ADN i la modificació d'histones estan estretament relacionades i interconnectades (Feinberg and Tycko, 2004; Kouzarides, 2007).

### 1.3. ORGANITZACIÓ DEL GENOMA EN EL NUCLI

Tal i com s'ha comentat, el nivell superior de compactació de la cromatina no es troba ben caracteritzat. Primer el nucleofilament es compacta per donar lloc a la fibra de 30 nm, que posteriorment es plega cada 150-200kpb (250nm durant la interfase) per donar lloc a un nivell màxim de compactació durant la metafase (850 nm) (Felsenfeld and Groudine, 2003).

En la interfase, l'organització del genoma recau en l'estructura dels cromosomes. Per mitjà d'estudis citogenètics es caracteritzen les diferents regions en base al seu patró de bandes. Així doncs utilitzant Giemsa, les bandes G pàl·lides (bandes R) són riques en GC, de replicació prompta i riques en gens, i les bandes fosques (bandes G) són riques en AT, de replicació tardana i amb pocs gens (Bayani and Squire, 2004).

En la perifèria nuclear es localitzen les regions de replicació tardana durant la fase S, corresponents a l'heterocromatina i als telòmers, transcripcionalment silenciats. Les regions riques en gens es localitzen principalment en l'interior, són de replicació primerenca durant la fase S i representen l'eucromatina. Aquestes regions són riques en histones acetilades i aquesta modificació es conserva durant la mitosi, fet que suggereix que l'acetilació d'histones pot servir per "marcar" la memòria de l'organització de dominis al llarg del cicle cel·lular (Xu and Price, 2011).

La localització de cromosomes en el nucli en interfase posa de manifest que cadascun d'aquests ocupa un espai determinat. En mamífers, aquesta organització varia en funció del tipus cel·lular (Bassett et al., 2009; Fraser and Bickmore, 2007). Així, encara que cadascun dels cromosomes ocupa un espai propi, certes parts de diferents cromosomes poden estar properes donant lloc a dominis funcionals. Les observacions mitjançant FISH (de l'anglès *Fluorescence In Situ Hybridization*) suggereixen que els gens tendeixen a localitzar-se en la superfície dels dominis cromosòmics. Segons el model obtingut de la localització d'alguns gens, els transcrits són alliberats en els canals intercromosòmics i després transferits als llocs adients pel seu processament, per ser, posteriorment, exportats al citoplasma després de la seva maduració.

Varis estudis han proposat que el nucli està organitzat en dominis o territoris. La localització de l'ADN en aquests dominis és, en part, una conseqüència de la pròpia activitat de la cromatina. Diversos mecanismes ajudarien a dirigir les proteïnes específiques a determinats dominis del nucli. En un model hipotètic, les proteïnes associades amb l'heterocromatina (HP1, Polycomb, Sir3p/Sir4p i ATRX), factors de transcripció (com Ikaros) i factors d'assemblatge (com CAF-1) podrien estar associats a l'establiment i manteniment dels dominis nuclears (Lamond and Earnshaw, 1998).

## 2. ELEMENTS REPETITIUS

La majoria d'estudis genètics o epigenètics es centren en les regions codificants i zones promotores o reguladores. Aquestes regions només representen un 2% del genoma.

S'ha vist que la complexitat de certs organismes, tant a nivell evolutiu com fisiològic, no pot explicar-se únicament per regions codificants. Existeixen múltiples mecanismes de regulació que ens fan veure la complexitat biològica de molts organismes. Per altra banda les modificacions postraduccionals així com els *splincings* alternatius fan incrementar aquesta complexitat donant gran variabilitat i diversitat funcional.

Les regions intergèniques o extragèniques són les que formen la major part de la seqüència del genoma humà, i la seva funció és, generalment, desconeguda (Hodgetts, 2004). Diferents estudis ens aproximen a una possible regulació duta a terme per aquest 98% del genoma anomenat *noncoding DNA* (de l'anglès ADN no codificant) (Wilusz et al., 2009). Bona part d'aquestes regions estan formades per elements repetitius classificables com repeticions en tàndem o repeticions disperses, encara que la resta de la seqüència no respon a un patró definit i classificable.

Gran part de l'ADN intergènic s'havia considerat un artefacte evolutiu sense una funció determinada en el genoma actual (Orgel et al., 1980), per la qual cosa durant dècades aquestes regions s'han denominat *junk DNA* (ADN parasitari o escombraria), denominació que inclou també les seqüències intròniques i pseudogens. No obstant, aquesta denominació no és la més adient donat el paper regulador conegut de moltes d'aquestes seqüències. Actualment s'ha vist que molts elements transposables s'han utilitzat durant l'evolució per crear noves senyals reguladores (Tomilin, 2008). A més, el notable grau de conservació evolutiva d'algunes d'aquestes seqüències sembla indicar que posseeixen altres funcions essencials encara desconegudes o poc conegudes. Per tant, la denominació *junk DNA* no seria correcte i seria millor denominar-lo **ADN no codificant** o **repetitiu** (Cordaux and Batzer, 2009; Deininger and Batzer, 2002).

### 2.1. METILACIÓ DELS ELEMENTS REPETITIUS

Com ja s'ha comentat abans, la metilació de l'ADN es troba majoritàriament en els elements repetitius. Per raons òbvies, hi ha un important biaix en la majoria de les investigacions sobre metilació de l'ADN, que han limitat els estudis a les illes CpG, donada la seva implicació en la regulació de molts gens. Per tant, encara és força desconegut com els mecanismes epigenètics que regulen els elements repetitius poden participar en la regulació de processos fisiològics i patològics com el càncer.

S'ha proposat que les regions del genoma que contenen elements repetitius podrien estar enmascarades per la compartimentalització de la cromatina, donant lloc a una reducció aparent de la mida del genoma (Rollins et al., 2006).

En mamífers i altres vertebrats, les regions silenciades es diferencien, encara que no de manera exclusiva, per la presència de metilació en l'ADN (Liu and Schmid, 1993;

Yoder et al., 1997), de manera que la majoria de les 5-metilcitosines es troben en seqüències repetitives, les quals s'estima que constitueixen aproximadament un 50% del genoma humà (Lander et al., 2001).

Per altra banda, s'ha descrit la hipometilació global del genoma com la primera anormalitat en la metilació de l'ADN detectada en tumors (Feinberg and Tycko, 2004). La hipometilació global es troba en els primers estadis de la carcinogènesis i ha estat associada a la disrupció genòmica i a la inestabilitat genètica. És probable que les seqüències repetitives desmetilades afavoreixin la recombinació mitòtica i promoguin reorganitzacions cromosòmiques (Rodríguez et al., 2006). Això s'ha observat en experiments en els que es va suprimir la metilació de l'ADN mitjançant la disrupció de les DNMTs, provocant aneuploidia (Jaenisch and Bird, 2003). Aquests transposons desmetilats també es podrien integrar en altres regions genòmiques "interrompent", d'aquesta manera, el genoma.

## 2.2. CLASSIFICACIÓ DELS ELEMENTS REPETITIU

### 2.2.1. ADN REPETIT EN TÀNDEM

Es tracta de repeticions que s'ordenen consecutivament, de manera que seqüències idèntiques, o casi idèntiques, es disposen una rera les altres (Ehrlich, 2002; Richard et al., 2008).

#### SATÈL·LITS

El conjunt de repeticions en tàndem de tipus satèl·lit comprèn un total de 250 Mb del genoma humà. Són seqüències entre cinc i varis cents de nucleòtids que es repeteixen en tàndem milers de vegades, generant regions repetides amb mides que oscil·len entre 100 kb fins a diverses megabases (Richard et al., 2008). Tenen un contingut en adenines i timines superior a la mitja del genoma i degut a això són menys denses.

#### MINISATÈL·LITS

Estan formats per una unitat bàsica de seqüència de 6 a 25 nucleòtids que es repeteix en tàndem generant seqüències d'entre 100 i 20.000 pb. S'estima que el genoma humà conté uns 30.000 minisatèl·lits.

Diversos estudis han relacionat els minisatèl·lits amb processos de regulació de l'expressió gènica com el control del nivell de transcripció, l'ajust alternatiu o la imprompta. Així doncs, s'han associat amb punts de fragilitat cromosòmica donat que es situen pròxims a llocs preferents de trencament cromosòmic, translocació genètica i recombinació meiótica. Un 10% dels minisatèl·lits humans són hipermutables, presentant una taxa mitja de mutació d'entre el 0.5% i el 20% en les cèl·lules de la línia germinal, sent així les regions més inestables del genoma humà (Richard et al., 2008).

En el genoma humà aproximadament el 90% dels minisatèl·lits es situen en els telòmers. La seqüència bàsica de sis nucleòtids TTAGGG es repeteix milers de vegades en tàndem, generant regions de 5 a 20 kb. Alguns minisatèl·lits per la seva gran inestabilitat presenten una notable variabilitat entre individus. Es consideren

polimorfismes multial·lèlics, donat que poden presentar-se en un nombre de repeticions molt variable, VNTR (de l'anglès *Variable Number Tandem Repeat*) (Usdin, 2008).

Són marcadors molt utilitzats en genètica forense, ja que permeten establir una empremta genètica característica de cada individu. Poden identificar-se fàcilment per *Southern Blot* i hibridació.

#### MICROSATÈL·LITS

Estan formats per seqüències bàsiques d' 1 a 4 nucleòtids, la repetició en tàndem de la qual origina sovint seqüències de menys de 150 nucleòtids. Alguns exemples importants són el dinucleòtid CA i el trinucleòtid CAG.

Els microsatèl·lits són també polimorfismes multial·lèlics, denominats STR (de l'anglès *Short Tandem Repeats*) i poden identificar-se per PCR. S'estima que el genoma humà conté uns 200.000 microsatèl·lits, que es distribueixen més o menys homogèniament, al contrari que els minisatèl·lits, la qual cosa els fa més informatius com a marcadors (Sharma et al., 2007).

#### **2.2.2. ADN REPETIT DISPERS**

Es tracta de seqüències d'ADN que es repeteixen de manera dispersa per tot el genoma, constituint el 45% del genoma humà (Cordaux and Batzer, 2009). Els elements quantitativament importants són els LINEs i SINEs, que es diferencien entre ells per la mida de la unitat repetida (Lander et al., 2001) (Figura I.12) (Taula I.2 i I.3).

TAULA I.2. FRACCIÓ DEL GENOMA DE VARIS ORGANISMES OCUPAT PER DIFERENTS TIPUS DE REPETICIONS DISPERSES

REPETICIÓ	HOMO SAPIENS	DROSOPHILA MELANOGASTER	C. ELEGANS	ARABIDOPSIS THALIANA
LINE, SINE	33.4%	0.7%	0.4%	0.5%
LTR/HERV	8.1%	1.5%	0%	4.8%
TRANSPOSONS d'ADN	2.8%	0.7%	5.3%	5.1%
TOTAL	44.4%	3.1%	6.5%	10.4%

Aquestes seqüències tenen la capacitat d'autopropagar-se al transcriure's a un ARN missatger intermediari, retrotranscriure's i insertar-se en un altre punt del genoma (Hodgetts, 2004; Hollister and Gaut, 2009) de manera autònoma o no (Figura I.9). Aquest fenomen es produeix poc freqüentment, estimant-se que un de cada 100 o 200

nounats porten una nova inserció d'un element Alu (membre de la família SINE) o L1 (membre de la família LINE) (Batzer and Deininger, 2002). Aquestes noves insercions poden resultar patogèniques per mutagènesis insercional, per desregulació de l'expressió de gens pròxims o per recombinació il·legítima entre dues còpies idèntiques de diferent localització cromosòmica (recombinació intracromosòmica o intercromosòmica), especialment entre elements Alu (Hasler et al., 2007).

			Mida	Nombre de còpies	Fracció del genoma
LINEs	Autonomous	ORF1 ORF2 (pol) AAA	6-8 kb	850,000	21%
	Non-autonomous	A B AAA	100-300 bp		
Retrovirus-like elements	Autonomous	gag pol (env)	6-11 kb	450,000	8%
	Non-autonomous	(gag)	1.5-3 kb		
DNA transposon fossils	Autonomous	transposase	2-3 kb	300,000	3%
	Non-autonomous		80-3,000 bp		

Fig. 1.9. Tipus d'ADN repetit dispers en el genoma humà. Es mostra la capacitat de transcriure's per si mateixes (autonomous) o no (non-autonomous), l'estructura de l'element, la mida, el nombre de còpies en el genoma humà i la fracció d'ocupació en el genoma humà. Figura adaptada de l'article de Lander (Lander et al., 2001).

Tal i com s'ha anat mencionant, els elements transposables s'havien considerat com a seqüències parasitàries, no obstant, i donat els múltiples estudis, s'ha vist que estan involucrats en múltiples processos biològics (Goodier and Kazazian, 2008; Muotri et al., 2007; Ponicsan et al., 2010; Walters et al., 2009). Els retrotransposons tenen gran influència en l'evolució (Liu et al., 2009; Price et al., 2004) i també juguen un paper en la regulació genòmica tant a nivell global com específic (Batzer and Deininger, 2002), generant diversitat genètica i variabilitat a nivell somàtic (Huang et al., 2010; Kidd et al., 2010; Macia et al., 2011).

També cal mencionar que els darrers estudis de seqüenciació massiva han destacat el possible rol dels retrotransposons i la resta d'elements repetitius en la diversitat genòmica i en diferents malalties (Pang et al., 2010; Stankiewicz and Lupski, 2010) així com en la variabilitat estructural entre individus (Iskow et al., 2010; Lupski, 2010; Xing et al., 2009).

En la Taula 1.3, adaptada de l'article de Lander (Lander et al., 2001), es mostren els resultats obtinguts de l'anàlisi *RepeatMasker version 09092000*, utilitzant *RepBaseUpdate 5.08*. En aquesta taula veiem en detall el nombre de còpies, el nombre total de bases en el genoma, la fracció del genoma que ocupen i el nombre de famílies i subfamílies de cada element dispers en el genoma humà.



TAULA I.3. NOMBRE DE CÒPIES I FRACCIÓ DEL GENOMA HUMÀ QUE FORMA PART DE L'ADN REPETIT DISPERS

	Nombre de còpies (x1000)	Nombre total de bases en el genoma (Mb)	Fracció en el genoma (%)	Nombre de famílies (subfamílies)
SINEs	1,558	359.6	13.14	
Alu	1,090	290.1	10.60	1 (~20)
MIR	393	60.1	2.20	1 (1)
MIR3	75	9.3	0.34	1 (1)
LINEs	868	558.8	20.42	
LINE1	516	462.1	16.89	1 (~55)
LINE2	315	88.2	3.22	1 (2)
LINE3	37	8.4	0.31	1 (2)
LTR elements	443	227.0	8.29	
ERV-class I	112	79.2	2.89	72 (132)
ERV(K)-class II	8	8.5	0.31	10 (20)
ERV (L)-class III	83	39.5	1.44	21 (42)
MaLR	240	99.8	3.65	1 (31)
DNA elements	294	77.6	2.84	
hAT group				
MER1-Charlie	182	38.1	1.39	25 (50)
Zaphod	13	4.3	0.16	4 (10)
Tc-1 group				
MER2-Tigger	57	28.0	1.02	12 (28)
Tc2	4	0.9	0.03	1 (5)
Mariner	14	2.6	0.10	4 (5)
PiggyBac-like	2	0.5	0.02	10 (20)
Unclassified	22	3.2	0.12	7 (7)
Unclassified	3	3.8	0.14	3 (4)
Total interspersed repeats		1,226.8	44.83	

Així doncs l'ADN repetit dispers es pot classificar en:

### SINEs

Les sigles **SINEs** vénen d'Elements Nuclears Dispersos Curts (de l'anglès *Short Interspersed Nuclear Elements*). Són seqüències curtes, generalment d'uns pocs cents de bases, que apareixen repetides milers de vegades en el genoma humà. Suposen el 13% del genoma humà (Taula I.2 i I.3). Els elements SINE no tenen capacitat d'autoproagar-se autònomament pel que requereixen d'una retrotranscriptasa present en el medi per la retrotranscripció del seu ARNm (Batzer and Deininger, 2002; Hasler and Strub, 2006; Walters et al., 2009). Dins aquest grup trobem elements MIR i elements Alu, els quals suposen el 10% del genoma (Batzer and Deininger, 2002; Hasler et al., 2007; Lander et al., 2001; Ponicsan et al., 2010; Walters et al., 2009). En l' apartat 3 es parlarà en detall dels elements Alu.

## LINEs

Les sigles **LINEs** venen d'Elements Nuclears Dispersos Llargs (de l'anglès *Long Interspersed Nuclear Elements*). Constitueixen el 20% del genoma humà (Taula I.2 i I.3). La família de major importància quantitativa és la LINE-1 o L1. Es tracta una seqüència de 6kb repetida unes 800.000 vegades de manera dispersa per tot el genoma, encara que la gran majoria de les còpies estan incomplertes al presentar l'extrem 5' truncat per una retrotranscripció incomplerta. S'estima que hi ha unes 5.000 còpies de L1, 90 de les quals són actives i la resta es troben inhibides per la metilació del seu promotor (Lander et al., 2001; Walters et al., 2009).

La seva riquesa en guanines i citosines és del 42%, pròxima a la mitjana del genoma (41%) i es localitzen preferentment en les bandes G dels cromosomes.

Posseeixen un promotor reconegut per l'ARN polimerasa II. Els elements LINE complets són codificants, i en concret LINE-1 codifica dos proteïnes: una proteïna d'unió a l'ARN, codificada pel ORF1 (del anglès *Open Reading Frame 1*) i un enzim amb activitat retrotranscriptasa i endonucleasa codificada pel ORF2. Per tant es consideren retrotransposons autònoms, ja que codifiquen les proteïnes que necessiten per propagar-se (Figura I.9). L'ARN polimerasa II present en el medi transcriu el LINE i aquest ARNm es tradueix en ambdós marcs de lectura produint una retrotranscriptasa que actua sobre l'ARNm generant una còpia d'ADN del LINE, potencialment capaç d'insertar-se en el genoma. Així mateix, aquestes proteïnes poden ser utilitzades per pseudogens processats o elements SINE per la seva propagació (Belancio et al., 2010).

La seqüència dels L1 té la propietat d'inhibir la transcripció, per aquest motiu els nivells d'ARNm i proteïnes codificades pels L1 a les cèl·lules són molt baixos. A més poden modificar la transcripció dels gens. S'ha vist que el 80% dels gens del genoma humà conté algun element L1 en els seus introns i la densitat en L1 correlaciona negativament amb els nivells d'expressió d'aquests gens (Ostertag and Kazazian, 2001).

## TRANSPOSONS LTR / HERV

Es tracta de retrovirus endògens humans i les sigles provenen de l'anglès *Human endogenous retrovirus* (Cordaux and Batzer, 2009; Goodier and Kazazian, 2008). El genoma dels retrovirus està format per ARN, fent que siguin capaços de retrotranscriure's i integrar el seu genoma en el de la cèl·lula infectada (Figura I.9). La mida d'un genoma retroviral complet és al voltant de 6-11 kb, però la majoria dels HERV són còpies incomplertes (Lander et al., 2001).

Així els **HERV** són còpies parcials del genoma del retrovirus integrats en el genoma humà al llarg de l'evolució dels vertebrats. Es tracten de vestigis d'antigues infeccions retrovirals que van afectar a cèl·lules de la línia germinal (Liu et al., 2009; Price et al., 2004). Algunes estimacions estableixen que hi ha unes 98.000 seqüències HERV, metre que d'altres afirmen que són més de 400.000. En qualsevol cas, s'accepta que al voltant d'un 5-8% del genoma humà està format per genomes antigament virals (Taula I.2 i I.3). Al llarg de l'evolució aquestes seqüències sense interès pel genoma hoste han anat acumulant mutacions sense sentit i delecions que les han inactivat.

Encara que la majoria de les HERV tenen milions d'anys d'antiguitat, almenys una família de retrovirus es va integrar durant la divergència evolutiva entre humans i ximpanzés, la família HERV-K (HML2) que suposa un 1% dels HERV (Taula I.3).

### TRANSPOSONS D'ADN

Els **transposons d'ADN** complerts posseeixen la capacitat d'autopropagar-se sense un intermediari d'ARN seguit de retrotranscripció. Un transposó conté en el gen un enzim transposasa, flanquejat per repeticions invertides (Figura I.9).

El seu mecanisme de transposició es basa en un "tallar - enganxar", movent la seva seqüència a una altra banda diferent del genoma. Els diferents tipus de transposases actuen de manera diversa, existint algunes capaces d'unir-se a qualsevol part del genoma mentre que altres s'uneixen a seqüències diana específiques. La transposasa codificada pel propi transposó s'obté realitzant dos talls flanquejants a la cadena d'ADN, generant extrems cohesius, i s'inserta en la seqüència diana, en un altre punt del genoma. Una ADN polimerasa omple els forats generats pels extrems cohesius i una ADN lligasa restableix els enllaços fosfodièsters, recuperant la continuïtat de la seqüència d'ADN (Pace and Feschotte, 2007).

S'estima que el genoma humà conté unes 300.000 còpies d'elements repetits dispersos originats per transposons d'ADN, constituint un 3% del genoma (Taula I.2 i I.3). Hi ha múltiples famílies, entre les quals cal destacar els elements Mariner per la importància patogènica i les famílies MER1 i MER2 per la generació de reordenacions cromosòmiques.

### 3. ELEMENTS ALU

Els **elements Alu** són retrotransposons que poden ser transcrits per l'ARN polimerasa III, però al no codificar per cap proteïna, es creu que requereix dels factors codificats per les LINEs per a la seva ampliació (Batzer and Deininger, 2002; Hasler et al., 2007; Walters et al., 2009).

Dins el grup de seqüències repetitives, els elements Alu, estan restringits a primats. Aquests elements tenen una mida aproximada de 300 pb i constitueixen la família de repeticions més abundant amb més d'un milió de còpies per genoma haploide, representant el 10% del genoma (Lander et al., 2001; Ponicsan et al.) Ponicsan et al., 2010 (Figura I.9 i Taula I.3).

No estan distribuïdes a l'atzar en el genoma humà sinó que solen trobar-se en regions riques en gens (Chen et al., 2002; Grover et al., 2004; Lander et al., 2001). Tendeixen a acumular-se en regions intròniques, UTR3' (*3'Untranslated Regions*) i regions intergèniques i juguen un paper important en la traducció, recombinació meiótica i reorganització del genoma. L'elevat nombre de seqüències Alu proporciona punts de nucleació potencials per recombinacions homòlogues, donant lloc a translocacions, duplicacions o delecions.

Aparentment, la selecció natural no actua sobre aquests elements ja que no són codificants. No obstant hi ha excepcions ja que una inserció pot determinar alteracions en un gen pròxim, promovent l'aparició d'alguna malaltia genètica, o bé pot trobar-se fortament lligada a un gen, pel que les pressions selectives que afectaran a aquest gen actuarien de forma indirecta sobre l'Alu contigua.

S'ha vist que els elements Alu participen en l'arquitectura del genoma delimitant-lo en dominis actius i inactius (Edwards et al., 2010; Rollins et al., 2006) i predisposant al genoma a la metilació en càncer (Danaei et al., 2005; Estecio et al.) (Estecio et al., 2010). També s'ha postulat el possible rol dels elements Alu en la regulació funcional i la implicació en l'evolució dels primats (Hasler and Strub, 2006; Jasinska and Krzyzosiak, 2004).

#### 3.1. ORIGEN I ESTRUCTURA DELS ELEMENTS ALU

Els elements Alu són seqüències de 250-280 nucleòtids presents en 1.500.000 de còpies disperses per tot el genoma. Els elements Alu van sorgir fa 65 milions d'anys, coincidint amb l'origen i expansió dels primats i resulten de la fusió de dues seqüències molt similars provinents del 7SL RNA (Figura I.10) amb el domini S deletat. El 7SL RNA codifica per un ARN que forma part de la partícula de reconeixement del senyal (SRP, de l'anglès *Signal Recognition Particle*), necessària per la síntesi de proteïnes en el reticle endoplasmàtic.

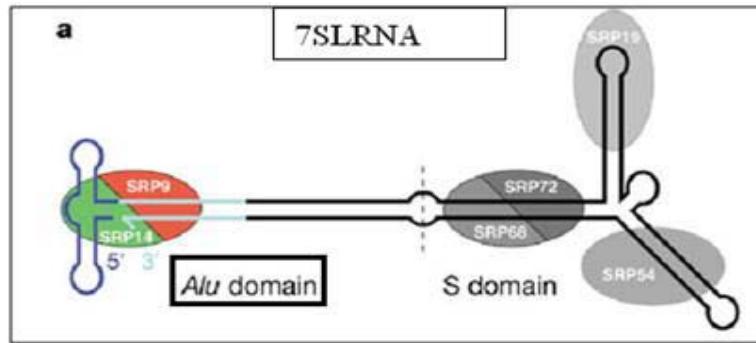


Fig. I.10. Estructura de 7SL ARN del qual segurament deriven els elements Alu.

Les dues seqüències originàries del 7SL ARN s'anomenen FLAM (de l'anglès *Free Left Alu monomer*) i FRAM (*Free Right Alu monomer*) i al fusionar-se formen una estructura dimèrica d'uns 300pb (Bennett et al., 2008; Hasler and Strub, 2006; Kriegs et al., 2007; Mighell et al., 1997) (Figura I.11). Estructuralment els elements Alu són dímers gairebé idèntics, excepte que la segona unitat conté una inserció de 32 nucleòtids que el fa major que la primera.

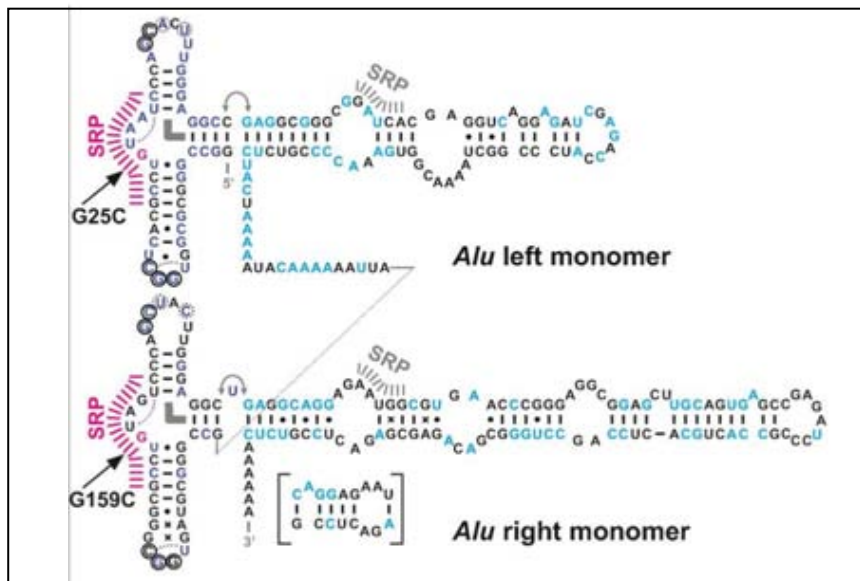


Fig. I.11. Formació de l'element Alu a partir de la fusió de les dos seqüències originàries de 7SL ARN amb el domini S deletionat (Bennett et al., 2008).

En quant a la seva seqüència, tendeix a tenir una considerable riquesa en guanines i citosines (un 56%), i ambdós monòmers presenten una cua poli(A) (seqüència d'adenines) en l'extrem 3' com a vestigi del seu origen d'ARN (Figura I.12). A més posseeixen un promotor de l'ARN polimerasa III per transcriure's en l'extrem 5' (caixes A i B de la Figura I.12).

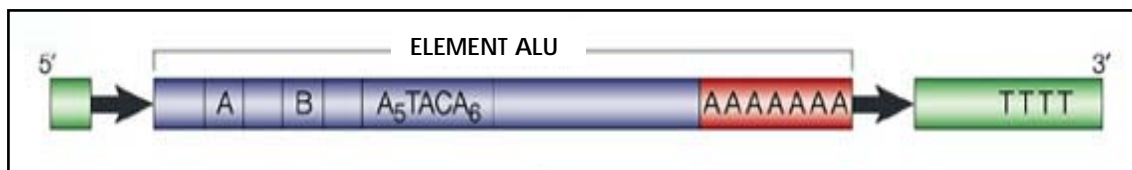


Fig. I.12. Estructura del element Alu. Les caixes A i B representen el promotor de la ARN polimerasa III, la seqüència A<sub>5</sub>TACA<sub>6</sub> es tracta d'una regió força freqüent en molts elements Alu, la caixa vermella és la cua poli(A). Figura adaptada de l'article de Batzer i Deininger (Batzer and Deininger, 2002).

### 3.2. TRANSPOSICIÓ DELS ELEMENTS ALU

Molt poques Alu tenen capacitat de transposar-se i les que ho fan reuneixen les següents característiques: presència d'un promotor ARN Polimerasa III i seqüències flanquejants, estructura secundària, estat de metilació i longitud de la cua poli-A determinades (Batzer and Deininger, 2002; Bennett et al., 2008; Hasler and Strub, 2006; Mighell et al., 1997).

Els elements Alu es consideren retrotransposons no autònoms ja que al no tenir ORF (*Open Reading Frame*) necessiten una retrotranscriptasa present en el medi per la retrotranscripció del seu ARNm (Batzer and Deininger, 2002; Hasler et al., 2007; Walters et al., 2009). Per aquest motiu solen situar-se a prop de LINEs, utilitzant d'aquesta manera la seva maquinària replicativa per dur a terme la mobilització. Aquest procés és similar a una transcripció inversa (Figura I.13).

Les caixes A i B dels elements Alu són necessàries per la transcripció per l'ARN polimerasa III; com que l'element Alu no té terminador intern la transcripció s'acaba quan l'ARN polimerasa III es troba un terminador extern (TTTT). Aquest extrem 3' amb 3 o 4 T (timines) es transformen en U (Uracils) (Figura I.13.B).

S'ha postulat que la cua poli(A) permet la lligació directa al lloc d'integració en el genoma per mitjà de la transcripció inversa. Aquesta primera inserció normalment es du a terme per una endonucleasa facilitada per un element LINE1 en el lloc consens TTAAA (Figura I.13.C). Els mecanismes que donen lloc a la segona inserció a l'altre extrem és una qüestió que encara es desconeix (Figura I.13.D).

Així doncs, només uns pocs elements Alu, anomenats "Alu masters", poden retrotranscriure's. Com ja hem comentat, els elements Alu contenen un promotor intern ARN polimerasa III, però aquest no es suficient per permetre la seva transcripció ja que es requereix de seqüències flanquejants adients per la seva activació. Per tant, la majoria d'elements Alu del genoma, per definició són inactius i no funcionals pel que fa a la retrotransposició. A més, la transposició d'aquests elements, activats de manera fortuïta per la seqüència flanquejant podria ser de curta durada ja que les CpGs que conté són propenses a mutar com a conseqüència de la desaminació dels residus de la 5-metilcitosina. Les mutacions en els dinucleòtids CpG podrien minimitzar o eliminar la capacitat retrotransposició d'un element Alu recentment integrat. Per altra banda, es

creu que les cues poli (A) de les Alu són importants en el procés d'amplificació i poden mutar ràpidament després d'una nova integració. Així doncs el poder de retrotranscripció dels elements Alu és molt limitat fent que l'increment de còpies de l'element Alu no sigui exponencial (Batzer and Deininger, 2002; Belancio et al., 2010; Mills et al., 2007; Roy-Engel et al., 2002a; Witherspoon et al., 2009; Xing et al., 2009).

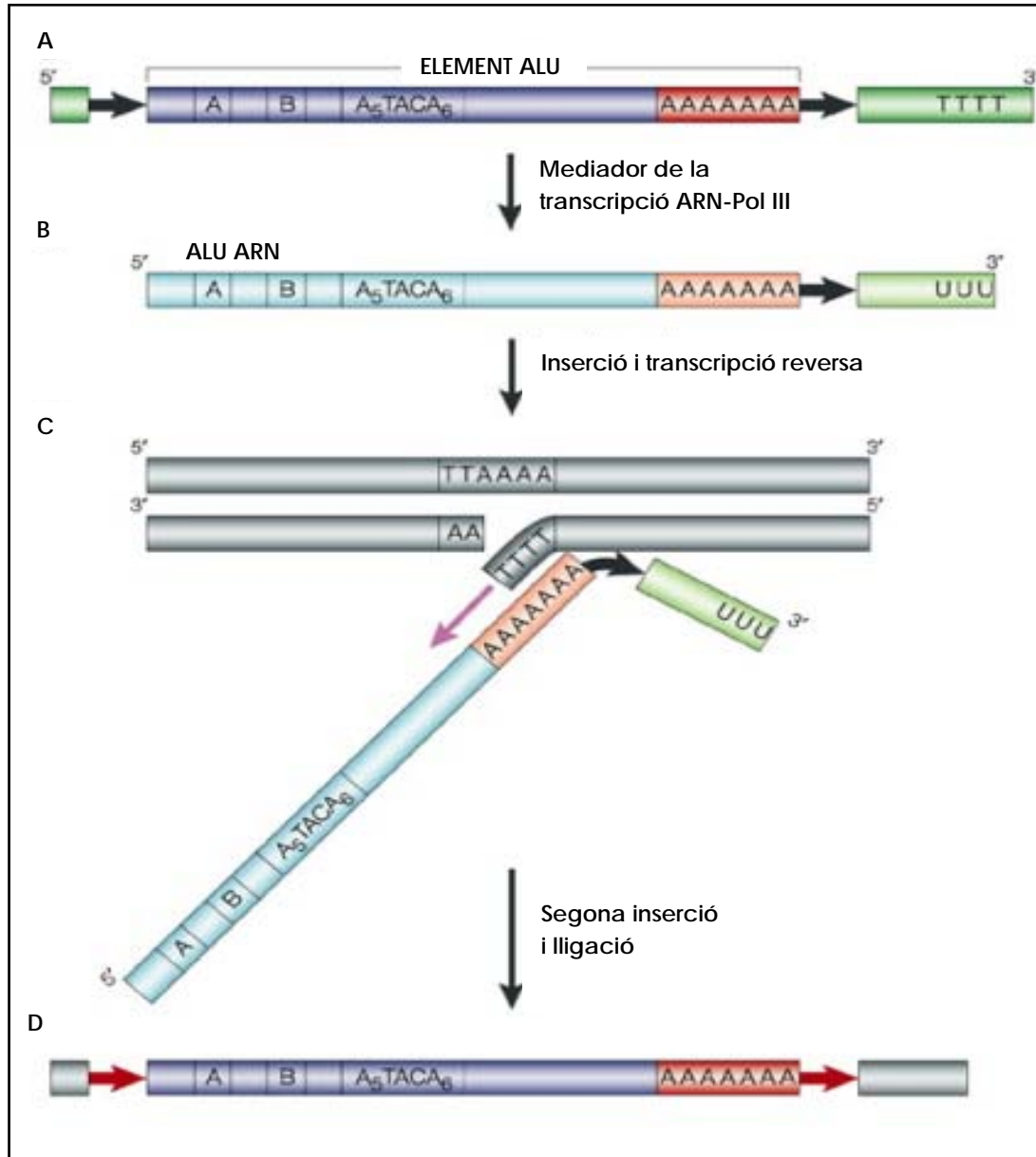


Fig. I.13. Mecanisme de retrotransposició de l'element Alu. A) Estructura del element Alu, B) Transformació a ARNm duta a terme per l'ARN polimerasa III. C) Inserció i transcripció inversa facilitada per l'endonucleasa d'un element LINE1. D) Segona inserció i integració de l'element Alu. Figura adaptada de l'article de Batzer i Deininger (Batzer and Deininger, 2002).

### 3.3. EVOLUCIÓ I CLASSIFICACIÓ DELS ELEMENTS ALU

Originalment es va pensar en el model de transposó, el qual explica que aquests tindrien capacitat de replicar-se a sí mateixos i cada còpia tindria, al mateix temps, capacitat de replicació tal i com ho fan els virus o els transposons, donant-se una amplificació exponencial. Més endavant es va veure que la taxa d'amplificació dels elements Alu no era uniforme, el nombre de còpies no augmentava exponencialment.

L'aparició de la majoria de les repeticions dels elements Alu es va produir fa més de 40 milions d'anys i podria ser deguda a un canvi en l'ambient cel·lular dels primats ancestrals, com per exemple un major potencial transcripcional dels elements Alu o un entorn específic del genoma tals com llocs diana d'elements Alu més assequibles, facilitant aquest èxit. Alternativament, un canvi en el mecanisme de proliferació dels elements Alu, tals com un augment de la quantitat de transcriptasa inversa o un augment de l'activitat dels enzims per la retrotransposició també podrien haver promogut la seva expansió (Ohshima et al., 2003).

En l'evolució dels primats, hi havia aproximadament una nova inserció per cada naixement d'un primat. Actualment aquest rati ha canviat fins l'ordre de una inserció cada 200 naixements. Per tant, el rati d'amplificació dels elements Alu ha disminuït en, almenys, dos ordres de magnitud durant la seva expansió (Kriegs et al., 2007; Roy-Engel et al., 2002b). Encara que les raons de la disminució de la taxa d'amplificació són encara desconeguts, s'hipotetitza sobre una possible causa generada arrel dels canvis en el potencial de mobilització de la retrotransposició d'elements Alu provocada per una transcripció alterada o la transcripció reversa. També podria ser una conseqüència d'una disminució de la disponibilitat de llocs d'inserció per a la integració de les noves còpies Alu, ja que la majoria d'aquests llocs ja estan ocupats per elements Alu antics. D'altra banda, també es podria especular amb el fet que el genoma humà ha evolucionat cap a la restricció de l'amplificació d'aquests elements.

Com ja hem comentat l'acumulació de mutacions fan que el poder de retrotranscripció dels elements Alu sigui limitat. Aquestes mutacions són les que han permès classificar als elements Alu en famílies. Així trobem les Alu J (les més antigues), les Alu S (d'edat intermèdia) i Alu Y (les més recents). Aquests grups al mateix temps es componen de diferents famílies (Batzer and Deininger, 2002; Belancio et al., 2010; Mills et al., 2007; Roy-Engel et al., 2002a; Witherspoon et al., 2009; Xing et al., 2009). La família Alu està formada per diverses subfamílies de diferents edats genètiques que es caracteritzen per una sèrie jeràrquica de mutacions. S'han identificat una sèrie d'elements Alu humans que comparteixen una seqüència consens i comprenen subfamílies que s'han anat expandint al llarg de l'evolució.

La figura I.14 compara les seqüències de consens de diverses subfamílies Alu. Els elements Alu més antics es caracteritzen per un menor nombre de mutacions específiques de subfamília i per un major nombre de mutacions aleatòries (fins al 20% de divergència), confirmant el seu origen ancestral. Per contra, les famílies més joves es caracteritzen per un nombre creixent de mutacions específiques de la subfamília, juntament amb un menor nombre de mutacions aleatòries (fins a un 0,1% de divergència) que s'acumulen després de la seva integració en el genoma.



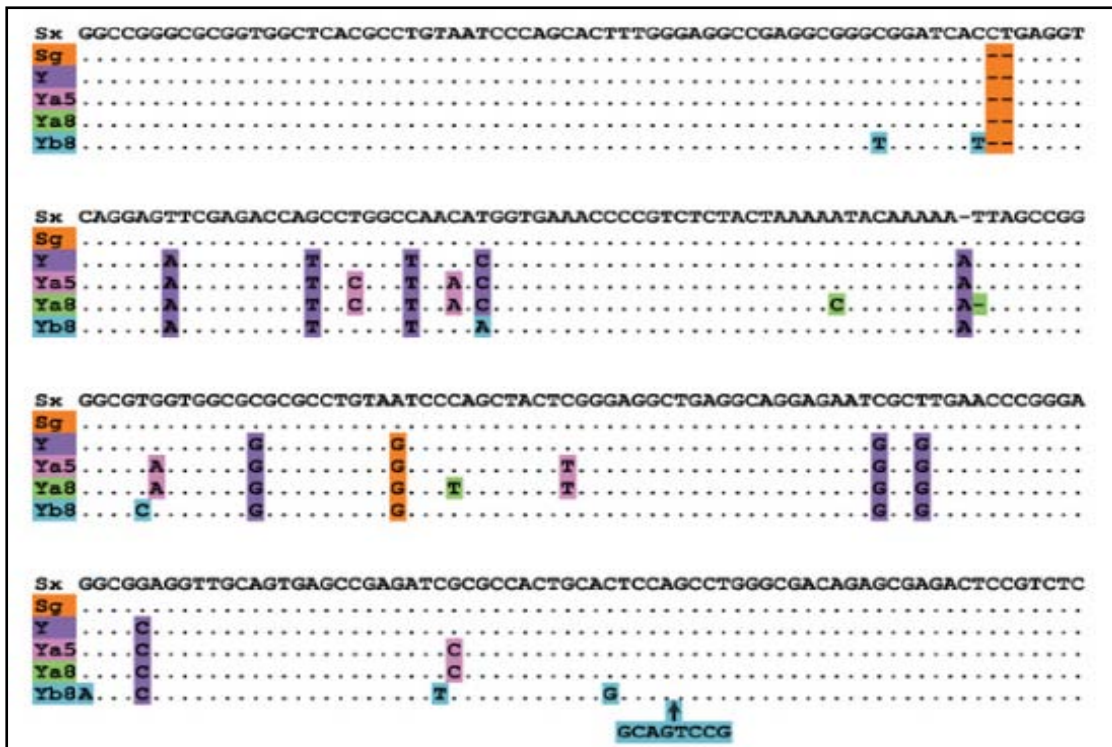


Fig. 1.14. Alineament de les seqüències consens de les subfamílies Alu. La seqüència consens de la subfamília AluSx es troba a dalt i a les files inferiors les Alu cada vegada més joves. El puntejat representen els mateixos nucleòtids que la seqüència consens, les delecions es presenten com a un guió, les mutacions es troben emmarcades en diferents colors depenent on s'ha originat ancestralment. Figura adaptada de l'article de Batzer i Deininger (Batzer and Deininger, 2002).

En la figura 1.15 s'il·lustra el patró d'expansió de les família Alu en els genomes de primats. Dins les Alu més antigues trobem les J i Sx amb més de 850.000 còpies, les Sg amb 40.000 còpies i les Y amb 200.000 còpies. Dins de la família Y trobem un conjunt de subfamílies formades únicament per 5.000 elements que es van integrar en el genoma humà 4-6 milions d'anys després de la divergència entre humans i primats africans. La majoria d'aquests Alu són monomòrfics en els diferents genomes humans, però un 25% s'han inserat en el genoma humà tant recentment que són dimòrfics, degut a la presència o absència d'inserció en alguns genomes. Els elements Alu Y tenen un gran nombre de mutacions específiques de cada subfamília i poques mutacions a l'atzar. Això fa que els elements Alu Y tinguin molt d'interés per estudis evolutius i de variabilitat genòmica en humans. L'estudi d'aquestes joves subfamílies han proporcionat pistes sobre la dinàmica i l'evolució dels elements Alu en el llinatge homínid (Batzer and Deininger, 2002; Price et al., 2004). En canvi, les famílies d'elements Alu més antigues es caracteritzen per un nombre més petit de mutacions, específiques de diagnòstic. A més aquestes famílies acumulen gran nombre de mutacions a l'atzar, la qual cosa confirma el seu origen ancestral.

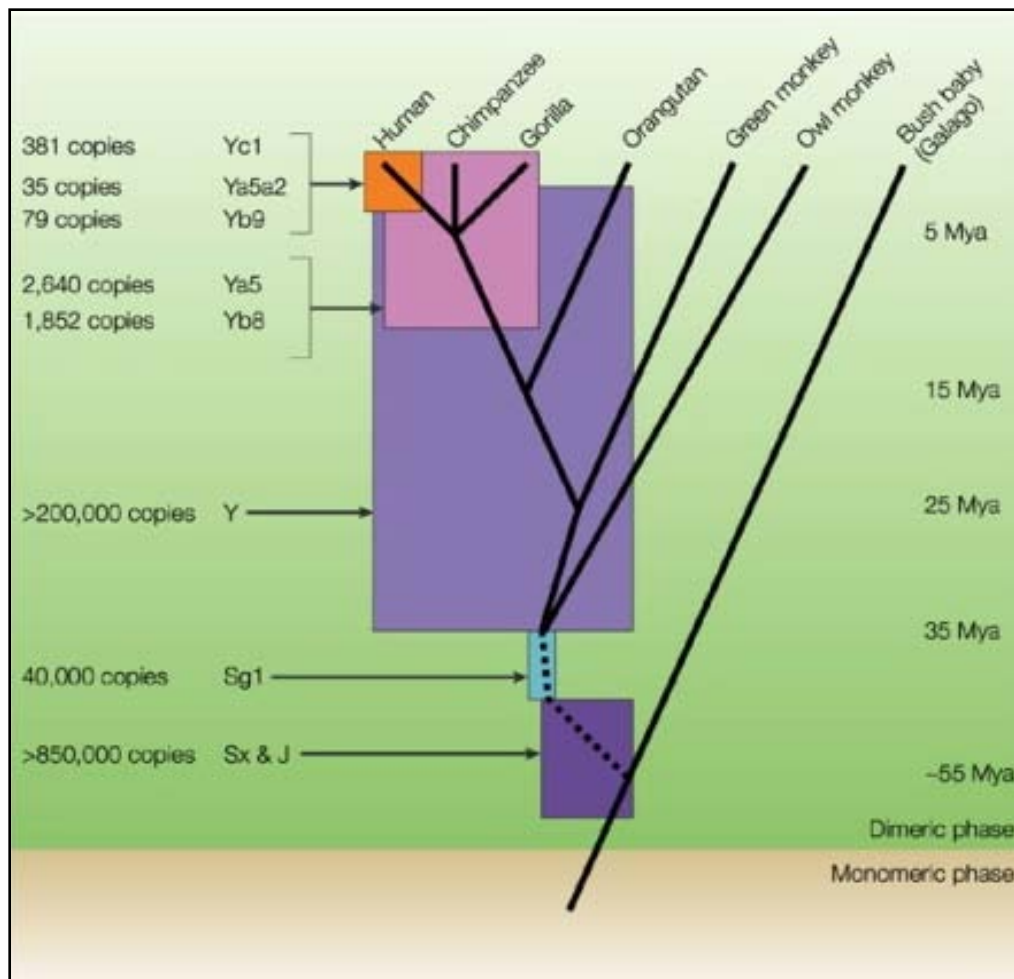


Fig. 1.15. Expansió dels elements Alu en primats. Es destaca l'expansió de les subfamílies Alu (Yc1, Ya5a2, Yb9, Yb8, Sg2, Sx i J) en l'arbre evolutiu dels primats. L'expansió de les diferents subfamílies està marcada en colors segons els pics d'amplificació. Milions d'anys enrere (Mya). Figura adaptada de l'article de Batzer i Deininger (Batzer and Deininger, 2002).

### 3.4. IMPLICACIÓ EVOLUTIVA DELS ELEMENTS ALU

Els elements Alu juguen un paper important la formació i evolució de xarxes reguladores (Cordaux and Batzer, 2009) així com en l'evolució dels *splincings* alternatius en primats (Keren et al., 2010; Lev-Maor et al., 2007).

Els estudis de genòmica comparada es basen en comparar seqüències genòmiques a gran escala mitjançant eines informàtiques. Aquests estudis permeten aprofundir en el coneixement d'aspectes evolutius d'escala temporal i espacial molt diversa (Batzer and Deininger, 2002; Hasler and Strub, 2006; Jasinska and Krzyzosiak, 2004; Price et al., 2004). L'estudi amb genomes de mamífers suggereix que aproximadament el 5% del genoma humà s'ha conservat evolutivament en els últims 200 milions d'anys, la qual cosa inclou la gran majoria dels gens i seqüències reguladores. No obstant, les regions codificants i

les seqüències reguladores actualment conegudes suposen només el 2% del genoma, la qual cosa suggereix que la major part de la seqüència genòmica amb gran importància funcional és desconeguda. Un percentatge important dels gens humans presenta un alt grau de conservació evolutiva. La similitud entre el genoma humà i el del ximpanzé (*Pan troglodytes*) és del 98,77%.

Com ja s'ha comentat, l'aparició de les seqüències Alu es situa fa aproximadament 65 milions d'anys, coincidint amb l'origen i expansió dels primats (Figura I.16). El seu estudi és útil per investigar la diversitat del patrimoni genètic de les poblacions humanes i la evolució dels primats. La majoria de les insercions d'elements Alu són recents i per tant, sovint polimòrfiques, agafant diferents valors de freqüència depenent de la població. Es consideren esdeveniments mutacionals únics ja que és bastant improbable trobar la mateixa inserció Alu en el mateix lloc, per aquest motiu són tant interessants per l'estudi de l'evolució humana. Es tracta d'insercions estables del genoma (Kazazian, 2004).

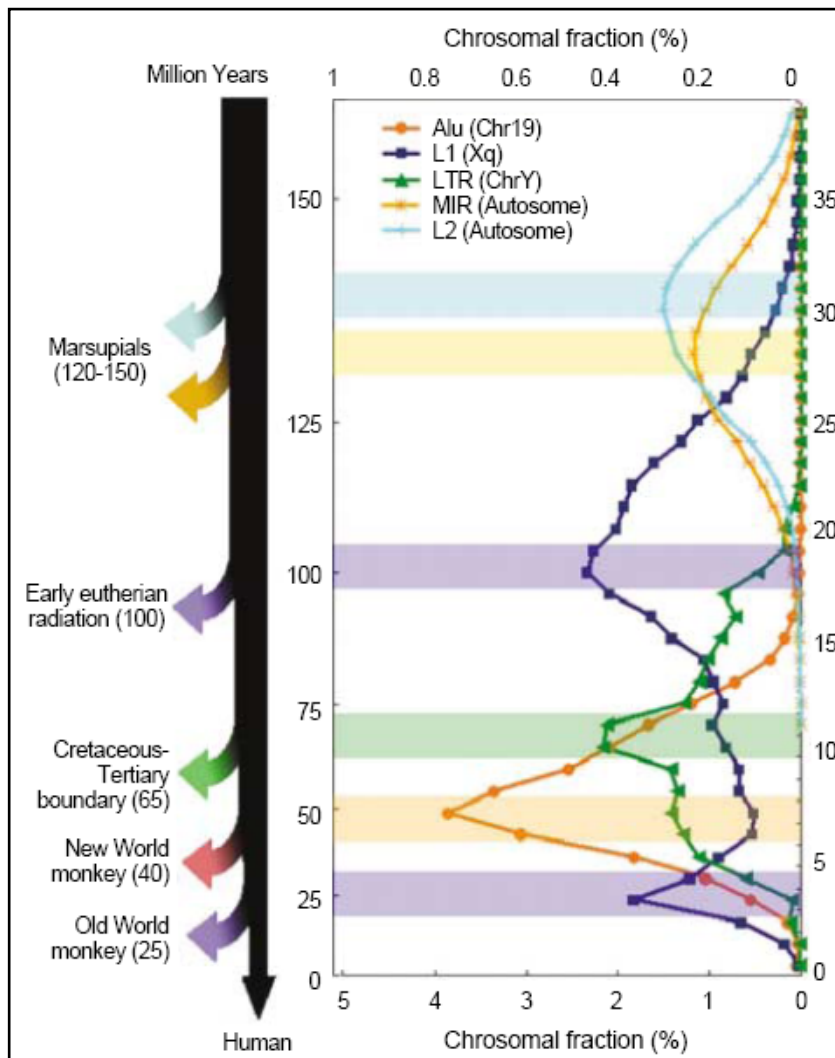


Fig. I.16. Aparició dels elements Alu fa 65 milions d'anys, coincidint amb l'origen i expansió dels primats. *J Korean Med Sci* 2004; 19: 177-85 ISSN 1011-8934

Els elements Alu presenten diversos avantatges pels estudis antropològics entre els que cal destacar el coneixement de l'estat ancestral (l'absència de l'element Alu) d'aquests polimorfismes, la facilitat i rapidesa de la seva detecció i que, al ser aquestes insercions successos únics i estables en la història evolutiva del nostre genoma, es pot assegurar que els al·lels detectats són idèntics per descendència.

La major part de les insercions Alu són relativament recents, per la qual cosa no s'han fixat o perdut, sent polimòrfiques i podent prendre diferents valors de freqüències en diferents poblacions humanes. Moltes d'aquestes insercions han passat en l'últim milió d'anys, i per això la seva presència no és universal. Cal esperar, per tant, que a partir d'un nombre limitat d'insercions Alu es pugui obtenir una bona discriminació entre poblacions de diferents continents, la qual cosa sustentaria l'argument de que aquests marcadors són interessants pels estudis de mestissatge.

En l'article de Carter A.B. (Carter et al., 2004) es va examinar una subfamília d'elements Alu en el genoma humà, coneguda com a AluYb. Van comparar aquests elements entre diferents espècies de primats, incloent ximpanzés, bonobos, gorilles, orangutans, gibons i siamangs. Van trobar que la subfamília AluYb acapara aproximadament el 40% dels elements Alu específics dels humans i és actualment un dels llinatges Alu més actius del genoma. Alguns elements AluYb encara es mobilitzen activament en el nostre genoma, causant mutacions d'inserció que han portat al desenvolupament de nombroses malalties hereditàries.

### 3.5. IMPLICACIÓ GENÒMICA DELS ELEMENTS ALU

Com ja s'ha comentat els elements Alu són els més abundants dels elements repetitius en el genoma humà. S'ha demostrat que la presència dels elements repetitius, i sobretot dels elements Alu, han tingut gran importància en el genoma humà i en particular en la seva evolució (Kreahling and Graveley, 2004; Sela et al., 2007). Per una banda, la integració en regions genòmiques ha causat la inactivació de gens que possiblement haurien estat perjudicials per l'organisme. Per altra banda els elements Alu han portat a duplicacions i delecions de segments de l'ADN mitjançant recombinacions no al·lèliques, fet que ha permès que s'acceleri l'evolució en varis ordres de magnitud (Lev-Maor et al., 2008).

El genoma s'ha adaptat a aquests elements repetitius mitjançant la regulació de l'expressió gènica (Mollet et al., 2010). Dels milers d'elements Alu que es troben en el genoma humà una part es troba en introns mentre que un cert nombre de seqüències Alu totals o parcials també es troben presents en les regions codificants dels ARNm madurs.

Els elements Alu poden actuar com a dianes per la creació de nous llocs d'unió a factors de transcripció, constituint, per tant, un mecanisme alternatiu de regulació en processos biològics proporcionant nous elements reguladors als gens veïns (Polak and Domany, 2006). De fet, diferents elements Alu han esdevingut participants de la transcripció gènica originant nous *enhancers*, promotors i senyals de poliadenilació (Brosius, 1999).

Els elements Alu estan involucrats en un gran nombre d'instabilitats genòmiques i mutacions patològiques (Konkel and Batzer, 2010; Sen et al., 2006; Stenger et al., 2001). Es troben diversos casos on les inclusions constitutives d'un element Alu donen lloc a malalties genètiques com el síndrome de Sly o d'Alport. Les cataractes congènites, la dismòrfia facial i la neuropatia també són síndromes associats a la inserció d'un element Alu dins un gen.

També estan associats a silenciament gènic (Chen et al., 2008) així com a mecanismes de reedició de l'ARN (Barak et al., 2009). El rol dels elements Alu a nivell transcripcional encara està en vies d'estudi i hi ha altres possibles funcions on podrien prendre part, modulant l'expressió de gens a nivell postranscripcional. Com per exemple la localització d'un element Alu en la regió UTR3' (de l'angles *Untranslated Region*) d'un gen podria actuar com a diana per microARNs (Gu et al., 2009; Smalheiser and Torvik, 2006) o també s'ha vist el seu paper en l'*splicing* alternatiu (Cordaux and Batzer, 2009; Keren et al., 2010; Lev-Maor et al., 2007).

### 3.6. CANVIS EPIGENÈTICS EN ELS ELEMENTS ALU

La majoria de 5-metilcitosines es troben en seqüències riques en CpG, dins d'ADN repetit en tàndem o dispers (Gardiner-Garden and Frommer, 1987; Yoder et al., 1997). No obstant, s'ha vist que gran nombre d'elements Alu, així com altres elements repetitius es troben en la cromatina activa i en l'ADN desmetilat (Edwards et al., 2010; Rodríguez et al., 2008b). Aquest fet mostra la complexitat de la regulació genòmica.

Treballs previs estimen que els elements Alu cobreixen el 33% del total de CpGs del genoma (Cordaux and Batzer, 2009) i es troben altament metilats en la majoria de teixits somàtics (Kochanek et al., 1993), encara que una fracció es manté desmetilada en cèl·lules normals i augmenta en cèl·lules tumorals (Danaei et al., 2005).

La desmetilació dels elements Alu s'ha vinculat amb un factor de risc en càncer (Wu et al., 2011). Normalment la hipometilació és una alteració primerenca en la carcinogènesi i s'ha associat a la disrupció genòmica i a la instabilitat genètica (Cho et al., 2007; Eden et al., 2003; Ehrlich, 2002; Rodríguez et al., 2006). Els elements desmetilats faciliten el reorganització a partir de la recombinació mitòtica i la transcripció (Iskow et al., 2010; Schulz et al., 2006). Per altra banda, la desmetilació també pot reactivar les Alu, donant lloc a la seva retrotransposició i conseqüent inserció en un altre punt del genoma. Els elements transposables actius poden interferir en l'expressió gènica per mitjà de l'alteració de l'*splicing* o actuant com a potenciadors (*enhancers*) o promotors (Kreahling and Graveley, 2004; Schulz et al., 2006; Slotkin and Martienssen, 2007). Així doncs, si la inserció té lloc en la regió promotora d'un gen podria afectar a la seva expressió ja sigui alterant l'estat de metilació, interrompent el promotor o introduint noves seqüències reguladores (Feltus et al., 2006). Alternativament, una Alu podria integrar-se en la regió codificant d'un gen i desplaçar el marc de lectura o alterar el *splicing* (Batzer and Deininger, 2002).

Un nombre d'elements Alu provinents de les subfamílies J i Sx es troben solapades amb illes CpG, en les regions promotores. Aquest fet suggereix el rol evolutiu en la regulació

genòmica durant la divergència dels primats (Kazazian, 2004). En humans s'ha estimat que més de 1.000 gens podrien estar regulats per elements Alu associats a illes CpG (Oei et al., 2004).

Per altra banda, certs estudis mostren com els patrons en la pèrdua de metilació en dels elements Alu es troben vinculats amb l'envelliment (Maegawa et al., 2010). Conseqüentment aquest fet contribueix a la desregulació epigenètica i al càncer.

També s'ha observat que la metilació aberrant i la mobilització transcripcional dels elements Alu contribueix a la inestabilitat genòmica en hipòxia (Pal et al., 2010).

Un altre fet interessant és que la remodelació de la cromatina en els elements Alu juga un paper important en la regulació de l'expressió dels ARNm. En l'article de Saito i col·laboradors s'ha vist que aquesta activació de les Alu silenciades associades a ARNm podria ser un nova teràpia pel càncer gàstric (Saito et al., 2009).

Finalment, en altres estudis s'ha observat un enriquiment d'elements SINEs en regions altament hipometilades en cèl·lules mare embrionàries. Aquest fet obre la porta a un nou perfil epigenètic específic de cèl·lules mare (Muramoto et al., 2010).

## 4. CÀNCER

Anomenem **càncer** al conjunt de malalties caracteritzades pel creixement anormal i descontrolat d'un grup de cèl·lules que formen una massa o tumor i que, en fases avançades, tendeix a infiltrar-se al torrent sanguini migrant a altres òrgans o teixits, originant metàstasi o tumors secundaris (Fearon and Jones, 1992).

El procés tumoral es deu a l'acumulació de múltiples alteracions genètiques que confereixen a la cèl·lula tumoral la capacitat de créixer de forma predominant respecte a la resta de les cèl·lules, ja sigui per augment de proliferació com disminució de mort cel·lular (Nowell, 1976).

Els factors determinants d'aquesta malaltia són, per una banda el fons genètic d'un individu i per altra, factors ambientals tals com exposició a agents carcinògens tant siguin químics, físics o biològics que introdueixen canvis genètics i/o afecten la regulació dels gens (Ahmed, 2004). L'envelliment també es considera un factor fonamental en el desenvolupament del càncer (Nardella et al., 2011).

Segons dades de l'Organització Mundial de la Salut (OMS) el càncer és la principal causa de mortalitat a nivell mundial. Se li atribueixen 7,9 milions de morts (aproximadament el 13% de les defuncions mundials) succeïdes en 2007. Les dades anuals de mortalitat per càncer són principalment causades per càncer de pulmó (1,4 milions de morts), d'estómac (866.000 morts), de fetge (653.000 morts), de còlon (677.000 morts) i de mama (548.999 morts). Essent els més freqüents en homes el de pulmó, estómac, fetge, colorectal, esòfag i pròstata; i entre dones, el de mama, pulmó, estómac, colorectal i coll uterí.

Aproximadament el 72% de les morts per càncer succeïdes en 2007 es va registrar en països d'ingressos baixos i mitjos. Es preveu que el nombre de morts anuals mundials per càncer continuarà augmentant i arribarà als 12 milions en el 2030.

### **CÀNCER COLORECTAL**

El càncer colorectal (CCR de forma abreviada) o càncer de còlon i recte es defineix com la proliferació anormala d'un grup de cèl·lules originàries d'aquest sector, que es reproduïxen molt accelerament i formen una massa o tumor. Aquest tumor en les fases avançades tendeix a infiltrar-se per la paret intestinal i òrgans propers i originant metàstasi o tumors secundaris als ganglis limfàtics de la cavitat abdominal i en diversos òrgans o teixits més o menys allunyats de l'intestí gros, com ara el fetge, els ossos, el cervell o els pulmons (Vogelstein et al., 1988).

El CCR es situa en la tercera causa de mort en els països desenvolupats i és el tercer càncer més letal després del de pulmó i pròstata en homes i el de mama en dones (Borras et al., 2001).

L'aparició d'un CCR pot ser conseqüència de factors ambientals i hàbits alimentaris, càncer esporàdic, o degut a una predisposició genètica, el que coneixem com a càncer hereditari i que succeeix en un 5-15% dels casos. Hi ha dos tipus clínics de càncer colorectal heretable en funció del nombre de pòlips en l'intestí gros (Grady, 2005). Trobem doncs el càncer de còlon heretable no poliposi (HNPCC de l'anglès

*Hereditary non polyposis colorectal cancer*) i el síndrome de Lynch (Borras et al., 2001; Grady, 2005).

Els pòlips de còlon i recte generalment són benignes. Pot haver un o múltiples pòlips i són més comuns a mida que la gent envulleix. El 80% dels pòlips són epitelials i s'originen a partir de la capa del teixit mucós que recobreix la superfície interna de l'intestí gros.

Amb el temps, certs tipus de pòlips, anomenats pòlips adenomatosos (ocupen un 90% dels pòlips) poden generar càncer de còlon. De major risc són els pòlips adenomavillosos (ocupen un 5% dels pòlips) dels quals un 40% experimenten un procés de esdevenir malignes. En el còlon també és freqüent trobar un altre tipus de pòlip, que s'anomena hiperplàsic i que sovint no genera càncer (Winawer, 1999).

Els pòlips més grans d'un centímetre tenen major risc de generar càncer que els de menor mida. Altres factors de risc són l'edat, els antecedents familiars de pòlips i càncer de còlon i la presència de pòlips adenomavillosos. D'altra banda cal tenir en compte que determinats processos inflamatoris crònics són sovint un antecedent de càncer (Friedman, 2006).

Hi ha diversos tipus de tumors malignes d'intestí gros segons el tipus de teixit a partir del qual s'han originat. L'adenocarcinoma (més del 95% del total dels casos) es tracta d'un tumor que es forma a partir de la mucosa o capa de teixit mucós que revesteix l'interior de l'òrgan. La resta dels casos es reparteixen fonamentalment entre els fibrosarcomes, originaris de les fibres de teixit connectiu de la paret intestinal, els liposarcomes, formats per teixit gras de la paret intestinal, i els limfomes, que s'originen en les acumulacions limfoides de la paret intestinal.

Segons la direcció en què efectuen el seu desenvolupament es diferencien dos tipus de tumors malignes d'intestí gros: Els estenosants, freqüents al còlon descendent, el sigma i el recte; i els poliploides, presents en el còlon ascendent i el transvers. Aquests últims a diferència dels anteriors, solen trigar més temps a erosionar-se o infiltrar-se pels teixits profunds de la paret intestinal (Ahnen, 1991; Leedham and Wright, 2008).

El càncer segueix una progressió en múltiples estadis, tant a nivell molecular com morfològic (Compton and Greene, 2004). Al llarg del temps s'han postulat diferents sistemes de classificació, entre els quals cal estacar els estadis de Dukes i la classificació TNM (tumor- nòdul- metastasi), que té en compte l'estat del tumor primari, si els nòduls limfàtics estan afectats i la presència de metastasi (Leedham and Wright, 2008) (Figura I.17).

Generalment el CCR sol ser asimptomàtic però senyals com sang en les femtes, fatiga associada amb l'anèmia, dolor abdominal i sagnat pel recte són alguns indicatius per sotmetre's a un examen rectal.

Els tractaments depenen de l'etapa en què es trobi el càncer. El general es sol realitzar cirurgia, quimioteràpia i radioteràpia per destruir les cèl·lules canceroses. També es solen fer anàlisis de sang per detectar marcadors tumorals, tals com l'antigen carcinoembrionari (ACE) i CA 19-9 (Sempere and Jover, 2008; Tinelli et al., 2010).



Amb el temps, els pòlips de còlon i recte adenomatosos (estadi 0) poden convertir-se en malignes i cal extirpar-los. En la majoria de casos els pòlips es poden extirpar al mateix temps que es realitza una colòscòpia. En casos on el càncer ja està present (estadis I, II i III) s'opta per una colectomia (extirpació d'una part del còlon) (Debrix et al., 2008). Per altra banda els fàrmacs més habituals en el tractament quimioteràpic són l'Irinotecan, l'Oxiloplatí i el 5-fluorouracil així com anticossos monoclonals com el Cetuximab, el Panitumumab i el Bevacizumab (Aprile et al., 2009).

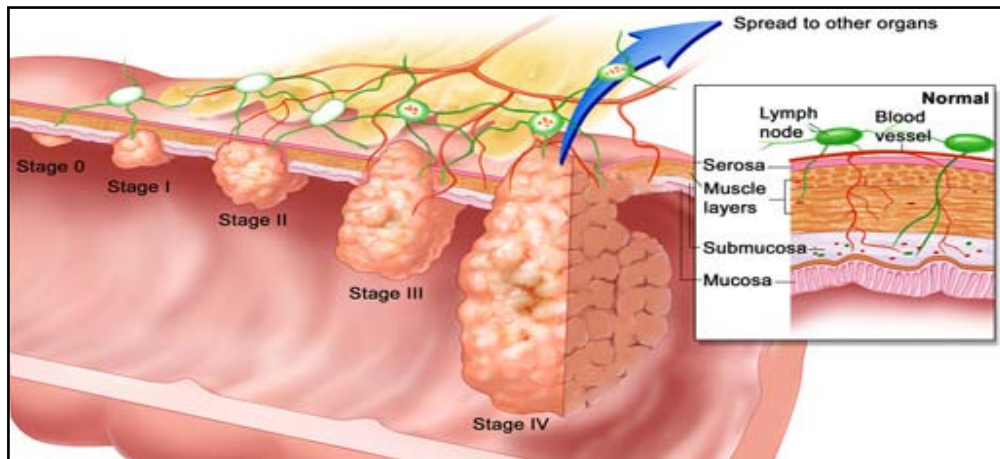


Fig. 1.17. Etapes de progressió del CCR: l'etapa I és el carcinoma in situ: intraepitelial o sense invasió de la làmina pròpia; en l'etapa II el tumor envaeix la capa subserosa, els teixits pericòlics o perirectals no peritonealitzats a través de la mucosa pròpia; en l'etapa III els tumors primaris tenen les mateixes propietats que en l'etapa II però amb afectació dels ganglis limfàtics; l'etapa IV es dona quan hi ha metastasi distant (Leedham and Wright, 2008).

#### 4.1. BASES MOLECULARS DEL CÀNCER

El primer model de progressió tumoral basat en dades moleculars que va integrar el concepte de progressió multietapada amb la identificació d'alteracions genètiques concretes va ser el proposat per Vogelstein en 1988. El model divideix el procés tumoral en diferents etapes (epiteli normal, adenoma primerenc, adenoma intermedi, adenoma tardà i carcinoma). Cadascun d'aquests estadis els relaciona amb la pèrdua d'heterozigotisme (LOH) en una zona cromosòmica diferent. Es tracta de gens supressors de tumors, la repressió dels quals beneficia la tumorigènesi de la cèl·lula (Vogelstein et al., 1988), o d'oncògens, que al expressar-se afavoreixen la transformació neoplàstica. En base aquest model s'ha postulat que els tumors de còlon i recte s'originen com a resultat de mutacions en aquests dos tipus de gens (Fearon and Jones, 1992).

Definim com a **oncògens** als gens responsables del control de la proliferació i la diferenciació cel·lular (Bertram, 2000). Donada la seva naturalesa dominant, la mutació en un al·lel ja seria suficient per desencadenar una proliferació descontrolada. Els oncògens poden actuar incrementant el producte o la seva funció

com a conseqüència d'una mutació, deguda a una ampliació gènica o a una translocació cromosòmica. Segons la seva funció cel·lular els oncogenes es poden classificar com a factors de creixement, receptors de factors de creixement, factors de transducció de senyal intracel·lular, proteïnes nuclears d'unió a l'ADN i factors del cicle cel·lular.

Per altra banda la mutació en gens supressors de tumors s'associa a la pèrdua de funcions cel·lulars. En aquest cas la mutació ha de ser en les dues còpies del gen per la supressió total del mateix ja que normalment es tracta de mutacions recessives. S'han descrit tres tipus de gens supressors de tumors: els guardians del genoma o *gatekeepers*, els cuidadors del genoma o *caretakers* i els cuidadors de l'entorn o *landscapers* (Kinzler and Vogelstein, 1998).

Un model que està ampliament acceptat és el model de Knudson (Knudson, 1985) el qual postula que després d'una mutació puntual és necessària la inactivació de l'altre al·lel. Aquesta inactivació es pot donar mitjançant diversos mecanismes, com per exemple una mutació puntual secundària, pèrdues heterozigòtiques, recombinació mitòtica, modificacions epigenètiques o pèrdua del cromosoma sencer, la qual pot anar acompanyada de la duplicació del cromosoma que porta l'al·lel mutat.

## 4.2. INESTABILITAT GENÈTICA

La **inestabilitat genètica** s'ha proposat com a motor de la tumorigènesi (Tinelli et al., 2010). Les cèl·lules tumorals acumulen un gran nombre de mutacions, aquest fet ha portat a postular l'existència d'un fenotip mutador causant de la inestabilitat genètica (Loeb and Loeb, 1999). La inestabilitat genètica per tant, explicaria l'elevat índex de mutacions en els tumors i permetria l'evolució de la cèl·lula tumoral en un entorn selectiu (Tomlinson and Bodmer, 1999). Així doncs, la inestabilitat genètica seria el pas previ a l'aparició d'un gran nombre de mutacions i per tant es donaria en les fases inicials de la tumorigènesi (Raptis and Bapat, 2006).

Podem definir la inestabilitat genètica com l'estat cel·lular anòmal associat a un augment de la taxa d'alteracions genòmiques, incloent mutacions, reorganitzacions cromosòmiques, delecions i amplificacions.

S'han descrit diverses formes d'inestabilitat genòmica causants d'un tant per cent elevat dels fenotips presents en cèl·lules tumorals (Quigley and Balmain, 2009). Es tracta de la inestabilitat de microsatèl·lits i la inestabilitat cromosòmica.

### 4.2.1. INESTABILITAT DE MICROSATÈL·LITS

Es tracta d'una inestabilitat genètica causada per mutacions puntuals, associada a errors en els gens de reparació de desparellament de l'ADN, generant una xarxa de mutacions de l'ADN molt elevada (Tinelli et al., 2010). És caracteritzada per l'acumulació

de milers d'alteracions somàtiques en la seqüència de l'ADN en forma de petites insercions, delecions i mutacions.

Com ja hem comentat, les seqüències de microsatèl·lits consisteixen en repeticions d'un a sis nucleòtids al llarg de fins a cent parells de bases. Aquestes seqüències són una diana pels errors en la replicació originats per les ADN polimerases.

La inactivació dels gens responsables de la reparació d'errors en l'aparellament de l'ADN provoca que aquests errors no es reparin i per tant, s'acumulin mutacions en la seqüència de microsatèl·lits. Aquest procés d'instabilitat es va detectar a principis dels anys noranta en el càncer colorectal (Ionov et al., 1993). La instabilitat de microsatèl·lits constitueix la majoria de càncers de còlon hereditaris amb absència de pòlips i en fins al 15% dels tumors de còlon i recte esporàdics.

Aquest tipus de instabilitat es dona de manera inicial en els tumors esporàdics i persisteix durant tot el procés tumoral permetent l'avenç del tumor degut a l'adquisició de noves mutacions en oncogenes o gens supressors de tumors (Shibata et al., 1994).

#### 4.2.2. INESTABILITAT CROMOSÒMICA

Es tracta d'un tipus d'instabilitat genòmica associada a un augment dels guanys i pèrdues de cromosomes, així com processos de no disjunció. Cada cèl·lula tumoral presenta un cariotip diferent degut a la instabilitat cromosòmica, la qual es defineix simplement com a l'adquisició d'alteracions que poden ser numèriques o estructurals (Lengauer et al., 1998; Toyota and Suzuki, 2010; Wu et al., 2011). Una alta instabilitat cromosòmica s'associa a un pitjor pronòstic en càncer colorectal (Risques et al., 2003).

Les **alteracions numèriques** poden ser degudes al guany (poliploidies) o pèrdua (aneuploidies) de cromosomes o alteracions en la euploidia (Nowak et al., 2002). Les alteracions numèriques es produeixen quan hi ha un error en la condensació i/o cohesió de les cromàtides germanes, en la formació del fus mitòtic o, en resum, algun problema en la segregació cromosòmica durant la mitosi.

Anomenem **alteracions estructurals** a aquell tipus d'instabilitat genòmica ocasionada per reorganitzacions cromosòmiques, ja siguin delecions, inversions, amplificacions o translocacions (Gasparini et al., 2007). Aquestes reorganitzacions són el resultat d'alteracions en la reparació del dany genòmic i/o pèrdua de la integritat dels telòmers. Normalment la cèl·lula té la capacitat de reparar les lesions de l'ADN, no obstant si la cèl·lula és incapaç de reparar el dany abans de la replicació es genera una mutació o un trencament de doble cadena, donant lloc a les reorganitzacions cromosòmiques. Un clar exemple de instabilitat cromosòmica estructural són les alteracions de regulació cel·lular en fase G1.

## 5. CÀNCER I EPIGENÈTICA

L'epigenètica, juntament amb la genètica, juga un paper fonamental en totes les etapes de la tumorigènesi. De fet, l'alteració dels patrons de metilació de l'ADN i de modificacions de les histones (Baylin and Schuebel, 2007) són un tret comú en tots els càncers humans, si bé, cada tipus de càncer presenta patrons característics (Esteller, 2007; Esteller, 2008; Jones and Baylin, 2007; Kouzarides, 2007).

Pel que fa a les alteracions en els patrons de metilació de l'ADN és sol parlar d'hipermetilació local i d'hipometilació global. Tot i que la primera troballa va ser la hipometilació global de l'ADN tumoral (Ehrlich, 2002; Feinberg and Tycko, 2004), la majoria d'estudis es van focalitzar en les metilacions *de novo* que succeïen en les illes CpG i la seva correlació amb el silenciament transcripcional d'un gran nombre de gens implicats en càncer (Esteller, 2008; Jones and Baylin, 2007).

L'activació d'oncogenes i la inactivació de gens supressors de tumors pot donar-se per diferents mecanismes, ja sigui per mutació, per pèrdua d'heterozigosi o per inactivació epigenètica dels gens supressors de tumors (Esteller, 2008; Feinberg and Tycko, 2004). Aquest darrer mecanisme està guanyant protagonisme últimament. El nombre de gens relacionats amb càncer afectats pel silenciament epigenètic és major que els alterats pels altres mecanismes (Esteller, 2007). A més, per un gran nombre de gens, la inactivació epigenètica resulta ser l'únic mecanisme possible de supressió (Berdasco and Esteller, 2010; Esteller, 2008; Feinberg; Jones and Baylin, 2007; Kouzarides, 2007).

Estudis recents amb noves metodologies com l'estudi genòmic a gran escala han descobert que la majoria de canvis de metilació en CpG associats al desenvolupament o al càncer no succeeixen en illes CpG sinó en les *CpG island shores* (Irizarry et al., 2009). Les diferències de metilació en les *CpG island shores* són visible no només entre cèl·lules normals i tumorals sinó també entre fibroblasts, cèl·lules induïdes pluripotents i cèl·lules mare embrionàries. Aquest fet fa palesa l'existència de mecanismes addicionals de reprogramació epigenètica presents en aquestes zones del genoma (Doi et al., 2009).

Els termes d'hipometilació i hipermetilació de l'ADN són termes relatius i indiquen major o menor grau de metilació de l'ADN referit a un estàndard (teixit normal, no tumoral, d'un individu) (Sharma et al., 2010). La pèrdua de metilació genòmica (hipometilació) com succés primari sovint s'associa amb el procés neoplàsic i és proporcional a la severitat de la malaltia (Jones and Baylin, 2002). Els genomes de les cèl·lules preneoplàsiques, canceroses i envellides comparteixen canvis importants en els nivells de metilació, com a successos primerencs en el desenvolupament d'alguns tumors. Els dos nivells de metilació poden presentar-se de forma individual o simultània.

En general, la hipermetilació que es dona en promotors d'alguns gens està associada amb el seu silenciament transcripcional i la hipometilació amb la sobreexpressió de certes proteïnes involucrades en els processos d'invasió i metàstasi. La hipometilació s'associa a inestabilitat cromosòmica i a pèrdua d'imprompta (Esteller, 2008; Portela and Esteller, 2010).

## 5.1. HIPERMETILACIÓ

La **hipermetilació** d'illes CpG és l'alteració més ben estudiada a nivell epigenètic en cèl·lules tumorals (Baylin and Herman, 2000; Esteller, 2002). La hipermetilació es troba freqüentment en illes CpG i s'observa en una gran varietat de càncers, en canvi és molt poc freqüent la hipermetilació de seqüències repetitives en càncer (Yan et al., 2000).

El paper de la hipermetilació en la carcinogènesis va lligat al silenciament transcripcional de gens implicats en la prevenció del càncer. Múltiples estudis han desvelat l'extensa distribució d'aquest gens per totes les regions cromosòmiques, afectant totes les vies cel·lulars conegudes (Esteller, 2008; Weber and Schubeler, 2007).

Els gens silenciats degut a la hipermetilació solen classificar-se com a gens supressors de tumors clàssics, candidats o de funció desconeguda. L'existència d'aquest gran nombre de gens hipermetilats i silenciats en càncer ha portat a postular l'existència d'un possible "fenotip metilador" (Toyota and Suzuki, 2010). Aquesta teoria suggereix que aquest "fenotip metilador" obeeix a la pertorbació d'algun sistema cel·lular de control d'aquests mecanismes epigenètics (Figura I.18), encara que no s'ha identificat encara la base molecular d'aquest fenotip.

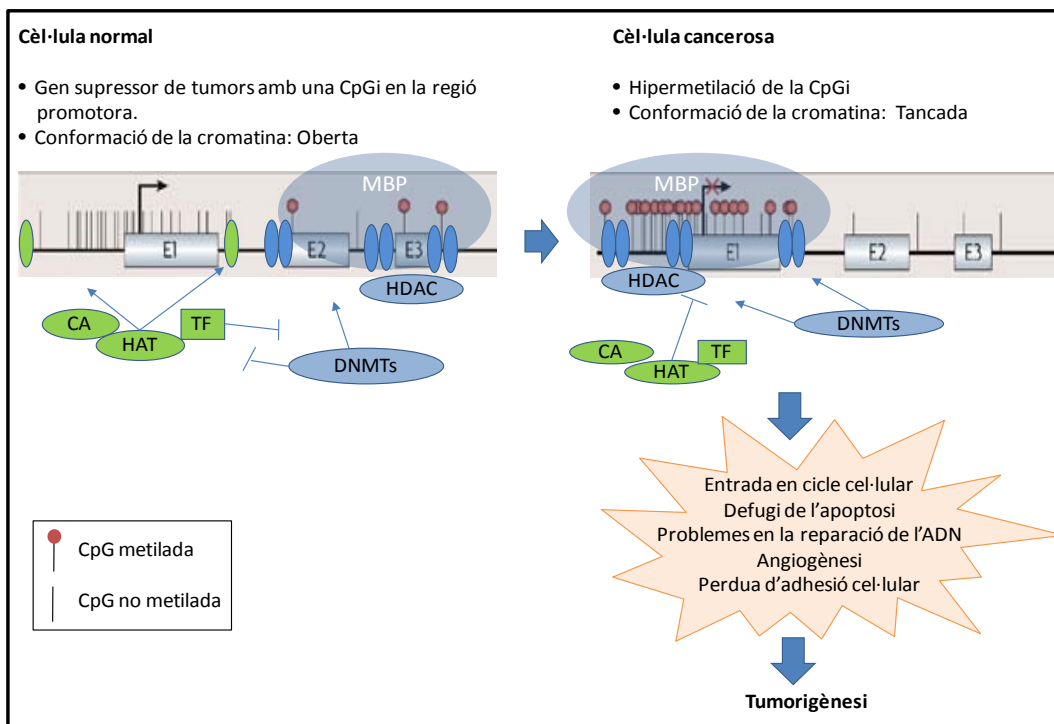


Fig. I.18. Patrons de hipermetilació en cèl·lules canceroses. Cèl·lules normals: Típica configuració d'activació amb el promotor desmetilat. Cèl·lula cancerosa: El gen mostra una metilació aberrant i per tant un silenciament transcripcional. Les CpGs dins de l'illa CpG estan metilades fent que la regió sigui accessible a les MBPs (de l'anglès Methylated Binding Proteins), els nucleosomes contenen HDAC i la cromatina està més compactada. Les illes CpG metilades interaccionen amb els complexos MBP-HDAC i els complexos de transcripció (CA, HAT, TF) no poden interactuar. La regió downstream de l'illa mostra la hipometilació típica de les cèl·lules canceroses. Figura adaptada de l'article de Baylin i Herman (Baylin and Herman, 2000) i de l'article de Esteller (Esteller, 2007).

En un estudi del RLGS (de l'anglès *Restriction Landmark Genomic Scanning*) per analitzar la metilació de casi 1.200 illes CpG en 98 tipus de càncers primaris (mama, còlon, cervell, cap i coll, testicles, leucèmia mieloide aguda i tumors neuroectodèrmics) es va estimar que un promig de 600 sobre 45.000 de les illes CpG del genoma estan hipermetilades en càncer respecte el teixit normal (Costello et al., 2000). Entre els tumors amb major freqüència d'hipermetilació en illes CpG trobem el còlon i el pulmó. S'ha vist que la hipermetilació no únicament es produeix en els gens amb una funció coneguda i implicada en la carcinogènesi sinó que també succeeix en gens no involucrats.

Així doncs en càncer, el silenciament epigenètic s'ha associat sempre com un esdeveniment local per silenciar gens concrets, no obstant també s'ha vist que grans regions cromosòmiques poden ser silenciades, aquest procés es defineix com a *Long Range Epigenetic Silencing* (LRES). Aquestes regions LRES poden cobrir megabases d'ADN formant regions d'heterocromatina amb el conjunt d'illes CpG de la regió hipermetilades. El mecanisme del LRES encara no es coneix, en alguns casos sembla que al produir-se el silenciament d'un gen concret la seqüència adjacent es veu afectada, hipermetillant-se per defecte. Per tant, semblaria que certes CpGi actuarien com a llocs diana i que al hipermetillar-se promourien la propagació de la metilació a les illes CpG veïnes. Així mateix, podria ser que l'elevada densitat d'hipermetilació de l'ADN provoqués la metilació de les marques d'histona donant com a resultat l'expansió heterocromatinització global entre aquestes regions riques en CpGs (Frigola et al., 2006).

## 5.2. HIPOMETILACIÓ

S'ha donat molta menys importància al paper de la hipometilació de l'ADN en la carcinogènesi (Baylin and Herman, 2000; Baylin and Schuebel, 2007; Issa, 2000) malgrat descobrir-se abans (Ehrlich, 2002; Vogelstein et al., 1988) i saber-se que és un fenomen molt estès en la majoria de tumors (Esteller, 2007).

La pèrdua de metilació en l'ADN és característica de la majoria de càncers i afecta a la inestabilitat genètica i l'activació gènica. Així com la hipermetilació afecta sovint a illes CpG de regions promotores de gens supressors de tumors, la **hipometilació** es dóna principalment en elements repetitius (Ducasse and Brown, 2006; Esteller, 2005).

La hipometilació pot afectar a l'estructura compacta de la cromatina provocant reorganitzacions mitòtiques i les conseqüents delecions i translocacions (Eden et al., 2003) (Figura I.19).

Aquest fet es va veure en l'estudi amb ratolins deficients de DNMTs, els quals patien un augment d'aneuploidies i translocacions (Karpf and Matsui, 2005). Altres estudis amb inhibidors de la metilació de l'ADN tant *in vivo* (Thomas and Williams, 1992) com *in vitro* (Denda et al., 1985) evidencien la importància de la hipometilació en càncer (Eden et al., 2003; Gaudet et al., 2003).

Sobre els anys vuitanta es va veure que diferents tipus de càncer tenien un baix contingut amb metilcitosines respecte als teixits sans. En un altre estudi, on es va

comparar càncer de còlon i mucosa normal adjacent, van observar una reducció d'un 10% de metilació en els càncers i pòlips benignes respecte a la mucosa normal (Feinberg, 1988). També s'ha observat el mateix fenomen al comparar tumors de pròstata metastàtics amb pròstata normal, leucòcits de pacients amb leucèmia limfocítica crònica amb leucòcits normals, hepatòcits tumorals amb normals i càncer de cèrvix amb teixit normal (Ehrlich, 2002). En resum, la hipometilació de diferents tipus de càncer respecte a teixits normals és força freqüent.

Les regions del genoma que es poden desmetil·lar són majoritàriament formades per seqüències i elements repetitius i una part minoritària formada per illes CpG associades a gens silenciats del cromosoma X inactiu, gens imprintats i gens metilats específics de teixit. Així doncs la hipometilació es pot classificar en hipometilació regional i hipometilació global.

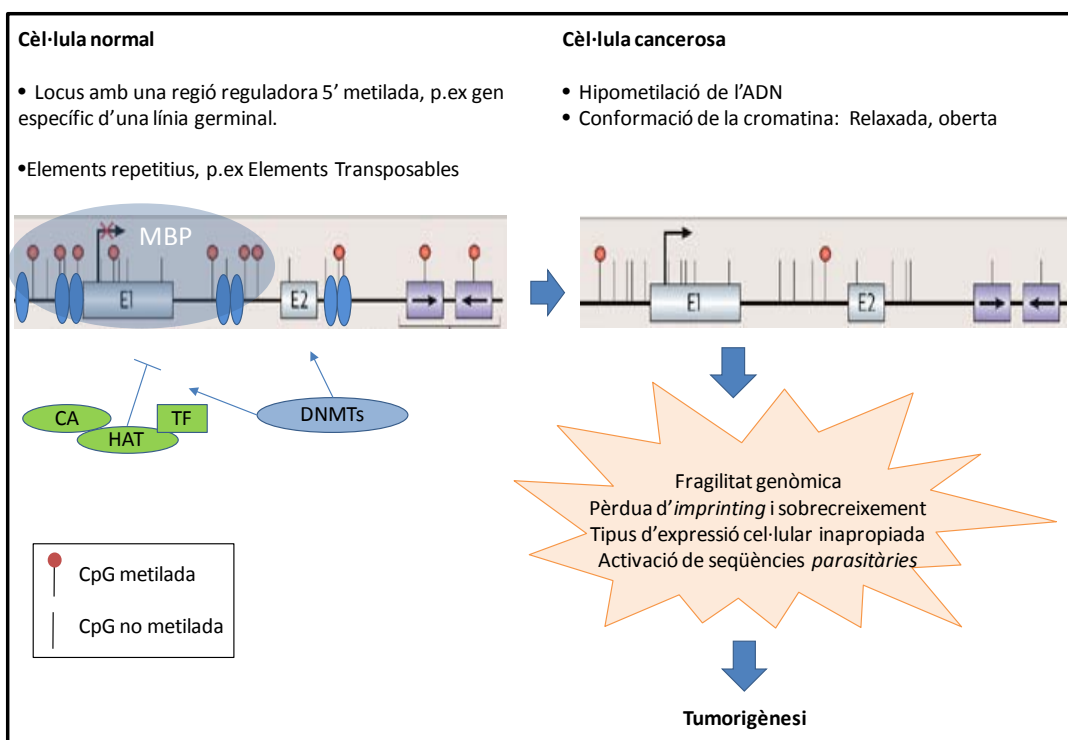


Fig. 1.19. Patrons de hipometilació en cèl·lules normals i canceroses. La hipometilació d'illes CpG en gens supressors de tumors és una alteració comú en cèl·lules canceroses, i provoca la inactivació transcripcional d'aquests gens i la pèrdua de la seva funció cel·lular normal. El genoma de les cèl·lules canceroses pateixen una hipometilació global de les seqüències repetitives, i els gens específics de teixit i els "imprintats" presenten també una pèrdua en la metilació de l'ADN. En alguns casos aquesta hipometilació contribueix al fenotip de cèl·lules canceroses, causant canvis com la pèrdua "imprinting" i pot contribuir a la inestabilitat genòmica de certs tumors. Figura adaptada de Baylin i Herman (Baylin and Herman, 2000) i Esteller (Esteller, 2007).

### 5.2.1. HIPOMETILACIÓ LOCAL

La **hipometilació local** afecta a regions discretes, bàsicament a illes CpG metilades. Aquesta s'associa a la reactivació transcripcional de gens silenciats durant el desenvolupament i la diferenciació (Filkowski et al., 2010; Pufulete et al., 2005; Wilhelm et al., 2010).

Les regions metilades i, per tant transcripcionalment silenciades poden ser susceptibles de ser hipometilades. Aquest gens reactivats normalment són oncogèns. Entre gens específics de teixit hipometilats en tumors trobem hRAS (Vogelstein et al., 1988), ciclina D2, Maspin 1, MAGEA1.

La imprompta genòmica també es pot perdre degut a la hipometilació. En aquest cas els gens perden la seva regulació monoal·lèlica i es sobreexpressen en tumors. Com a exemple de la pèrdua d'imprompta o LOI (de l'anglès *Loss Of Imprinting*) trobem el cas del locus IGF2/H2 en varis tumors (Cui et al., 2002; Feinberg, 2007).

### 5.2.2. HIPOMETILACIÓ GLOBAL

La **hipometilació global** afecta als elements repetitius. Aquesta desmetilació s'associa principalment a la inestabilitat genètica (Toyota and Suzuki, 2010; Wu et al., 2011).

El genoma humà consta d'un 45% d'elements repetitius (Lander et al., 2001) els quals normalment es troben reprimits i estan dotats d'un alt contingut de 5-mC. No obstant, els elements repetitius, com els LINEs o els Alus, sovint es troben hipometilats en càncer (Rollins et al., 2006). De fet, els elements repetitius són els que més pateixen la hipometilació en cèl·lules canceroses. Els elements transposables desmetilats estan actius, poden translocar-se a altres regions del genoma, contribuint d'aquesta manera a la inestabilitat genòmica (Takai and Jones, 2002).

S'ha postulat que les pèrdues de metilació globals descrites en càncer podrien tenir lloc de manera massiva en aquestes seqüències repetitives. Aquest fet s'ha demostrat en diversos elements repetitius com l'ADN satèl·lit pericentromèric (Stephens et al., 1998; Weisenberger et al., 2005) i en els elements Alu i LINEs (Choi et al., 2007; Weisenberger et al., 2005; Yang et al., 2004).

Així doncs la hipometilació global en càncer sol associar-se a la presència d'elements repetitius, normalment metilats. No obstant, Hansen i col·laboradors van observar la majoria de la metilació global era causada per regions hipometilades independentment de la presència o no d'elements repetitius en les regions. En el seu estudi van cobrir el 57% d'elements L1, el 94% de L2, el 95% de seqüències Mir i el 18% d'elements Alu. Aquests elements analitzats no van mostrar una associació específica a la hipometilació, malgrat tot també comenten que els elements repetitius no coberts en el seu estudi podrien estar més hipometilats contribuint així a la hipometilació global (Hansen et al., 2011).



## 6. DIFERENCIACIÓ I EPIGENÈTICA

El balanç entre pluripotència i diferenciació de les cèl·lules mare embrionàries també es troba regulat per processos epigenètics, els quals controlen l'expressió de factors de transcripció propis de l'estat pluripotent, així com específics del llinatge cel·lular. Cada cèl·lula de l'organisme posseeix el seu propi patró epigenètic, que és un reflex de les influències ambientals a les que ha estat sotmesa, i que determina, en última instància el seu fenotip cel·lular (Ptashne, 2007).

Els processos epigenètics més importants que regulen la pluripotencialitat i els processos de diferenciació de les cèl·lules mare embrionàries són diversos:

### 6.1. METILACIÓ DE L'ADN

La metilació tal i com ja s'ha comentat, també juga un paper molt important en el desenvolupament embrionari, així com en la inactivació del cromosoma X en les femelles, la imprompta genètica i el silenciament de transposons.

Durant el desenvolupament embrionari es produeix la modificació dels patrons de metilació del genoma, de forma que l'ADN es troba desmetilat en el moment de la implantació, exceptuant els gens sotmesos a la imprompta genòmica. Després de la implantació, el patró de metilació propi de l'embrió s'aconsegueix mitjançant l'acció *de novo* de les ADN metiltransferases DNMT3a i DNMT3b (Okano et al., 1999).

El perfil de metilació de l'ADN en cèl·lules mare embrionàries es diferencia dels existents en teixits adults (Bibikova et al., 2006) i malgrat que els nivells globals de metilació no es modifiquen durant la diferenciació cel·lular (Meissner et al., 2008), es produeixen canvis en la metilació de promotors de gens específics. Entre els que incrementen els seus nivells de metilació es troben gens implicats en el manteniment de la pluripotència i gens específics de la línia germinal (Shen et al., 2006), de forma que es produeix un silenciament. Simultàniament, es desmetilen gens implicats en la diferenciació cel·lular, permetent la seva expressió. D'aquesta forma, la metilació de l'ADN regula l'expressió d'alguns marcadors de pluripotència, com Oct4 i Nanog, que durant l'estat no diferenciat presenten els seus promotors i potenciadors o *enhancers* desmetilats i es comencen a metilar seqüencialment quan s'inicia el procés de diferenciació cel·lular a fi de reduir paulatinament l'expressió d'aquests gens (Hattori et al., 2004). El patró de metilació que mostren les cèl·lules mare embrionàries durant la diferenciació depèn del llinatge cel·lular cap al que es diferencien, que en última instància determinarà els gens, l'expressió dels quals es necessiti.

Malgrat això, s'ha demostrat que els processos de metilació de l'ADN són prescindibles pel manteniment de l'estat pluripotent de les cèl·lules mare embrionàries, encara que resulten essencials per portar a terme la diferenciació cel·lular, ja que la seva absència la impossibilitaria, tal i com demostren certs models murins que no expressen ADN metiltransferases (Gilbert et al., 2007).

## 6.2. MODIFICACIÓ D' HISTONES

Les cèl·lules mare embrionàries es caracteritzen per tenir una estructura dinàmica i permissiva de la cromatina, donat el gran nombre d'histones acetilades que afavoreixen l'accés de nucleases i de la maquinària transcripcional a l'ADN (Keenen and de la Serna, 2009).

Així, durant l'estat no diferenciat, els promotors de gens pluripotents (com Oct4 i Nanog) posseeixen elevats nivells d'acetilació de les histones H3 i H4, així com la marca activadora de la transcripció H3K4me3 (Bernstein et al., 2006), permetent que ambdós gens s'expressin. No obstant durant la diferenciació la cromatina adquireix una estructura més restrictiva, produint-se un augment global de la marca d'inactivació H3K9me3 i una disminució dels nivells d'acetilació de les histones H3 i H4 (Atkinson and Armstrong, 2008; Meshorer and Misteli, 2006).

Així doncs, les cèl·lules mare embrionàries utilitzarien els mateixos mecanismes epigenètics que les cèl·lules somàtiques per controlar la transcripció gènica. No obstant, també presenten algunes característiques específiques per silenciar gens que controlen la diferenciació cel·lular (Atkinson and Armstrong, 2008; Barrero et al., 2010). És el cas dels **dominis bivalents** (Azura et al., 2006; Bernstein et al., 2006), formats per zones riques en la marca repressiva de la transcripció H3K27me3 i, al mateix temps, la marca activa H3K4me3. Aquests dominis s'han trobat propers a inicis de transcripció gènica implicats en diferenciació cel·lular i encara que també es poden trobar en teixits diferenciats i en càncer, la seva dualitat funcional és molt més clara en cèl·lules mare (Voigt et al., 2013).

S'ha postulat que l'existència dels dominis bivalents s'encarregarien de mantenir reprimida la transcripció gènica d'aquests però serien capaços de produir una conformació transcripcionalment permissiva de la cromatina al rebre els estímuls que les induïssin a diferenciar-se, reprimint H3K4me3 en el cas de silenciar l'expressió o reprimint H3K27me3 en el cas de promoure l'expressió gènica (Azura et al., 2006; Bernstein et al., 2006).

El silenciament gènic és un procés complex que requereix, tal i com ja s'ha comentat, de l'acció simultània de múltiples mecanismes perquè es dugui a terme de manera efectiva. En aquest sentit, s'ha demostrat que les modificacions de les histones i la metilació de l'ADN actuarien de manera sinèrgica per regular la transcripció.

## 7. RESULTATS PREVIS DEL NOSTRE LABORATORI: QUANTIFICACIÓ I IDENTIFICACIÓ D'ELEMENTS ALU

En estudis previs del nostre grup es van analitzar una sèrie de 17 tumors de còlon i recte i les seves parelles de teixit normal per mitjà de la tècnica QUMA (*Quantification of UnMethylated Alus*) (Rodríguez et al., 2008b) (tècnica explicada en detall en material i mètodes, apartat 2.3.2) a fi de quantificar els elements Alu desmetilats (Figura I.20).

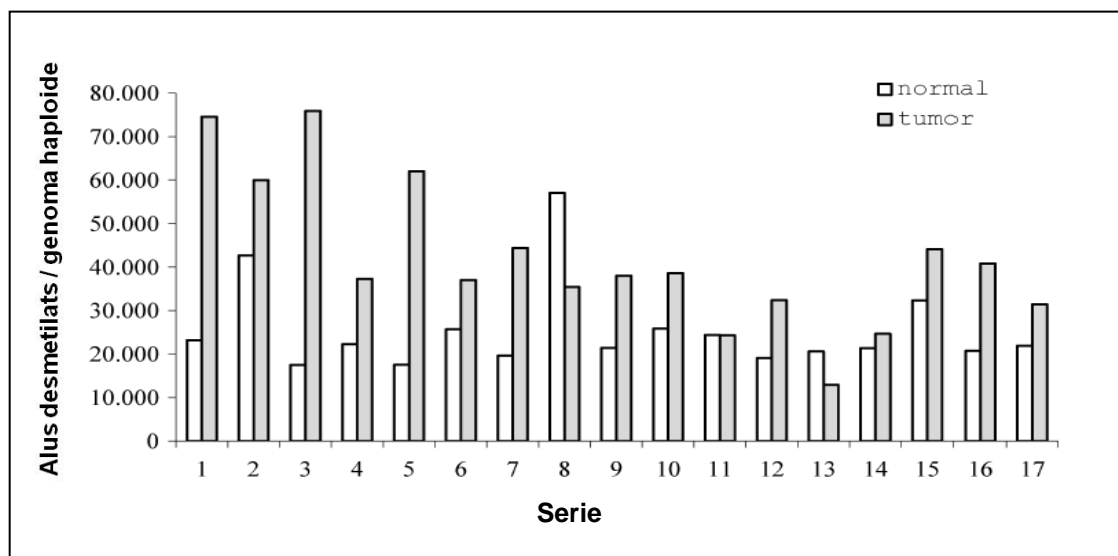


Fig. I.20. Quantificació dels elements Alu desmetilats en 17 parelles de mucosa normal i carcinoma colorectal per la tècnica QUMA. Els valors representen el nombre de Alus desmetilades per genoma haploide. La majoria de tumors presenten major nombre d'hipometilacions al comparar-los amb la seva parella normal. Figura adaptada de l'article de Rodríguez J i col·laboradors. (Rodríguez et al., 2008b).

Els resultats obtinguts van demostrar que les cèl·lules epitelials normals de còlon contenen aproximadament 25.000 Alus desmetilades ( $25.486 \pm 10.157$ ), en canvi en cèl·lules tumorals aquest valor gairebé és duplicava ( $41.995 \pm 17.187$ ).

Es van identificar de manera computacional un total de 198.309 elements Alu que contenen la seqüència AACCCGGG (diana Smal utilitzada en la tècnica QUMA), i es va estimar que un 15.1% dels elements Alu amb aquesta seqüència estaven desmetilats en mucosa normal mentre que un 24.9% ho estava en carcinoma de còlon.

Considerant que el genoma humà conté aproximadament un milió d'elements Alu, aquesta estimació indica que les Alu desmetilades constitueixen entre el 2.3 i el 3.8% del total d'elements Alu en teixit normal i tumoral respectivament.

Per altra banda una sèrie més gran de mostres, que incloïa aquestes parelles N-T, es van analitzar per la tècnica AUMA (*Amplification of UnMethylated Alus*) (tècnica explicada en detall en Material i Mètodes, apartat 2.3.2). Aquesta metodologia va permetre identificar una col·lecció d'elements Alu desmetilats en teixit normal i/o tumoral així com un gran nombre de regions fora d'Alu hipermetilades i hipometilades en tumors de còlon i recte.

**HIPÒTESI I OBJECTIUS**

Una aclaparant majoria dels més d'un milió d'elements Alu es troben en dominis inactius i estan metilats. L'excepcionalitat de l'estat de cromatina activa d'alguns elements Alu, tant en situacions fisiològiques com patològiques, i específicament en càncer, els converteix en factors amb una potencial funcionalitat, i per tant en candidats pel seu estudi.

Si bé la possible activitat dels elements Alu està determinada per diferents marques epigenètiques, aquells que estiguin desmetilats serien bons candidats per tenir un paper més rellevant. La confecció d'un mapa genòmic dels elements Alu desmetilats i previsiblement actius a nivell de cromatina pot representar una eina fonamental per entendre el seu paper sobre l'expressió gènica i el manteniment estructural de la cromatina. Tanmateix, canvis en l'estat epigenètic de determinats elements Alu poden constituir marcadors amb un impacte sobre el risc i l'evolució de la malaltia.

Malgrat que el nombre de metodologies de *screening* de metilació (Butcher and Beck, 2008; Jorda et al., 2009; Laird, 2010) s'ha incrementat en els darrers anys, i que s'han utilitzat diferents aproximacions per estimar els nivells de metilació en els elements repetitius, encara hi ha una manca d'estratègies que permetin la identificació específica d'alteracions de la metilació de l'ADN en elements repetitius.

En el nostre laboratori s'han desenvolupat dues tècniques per identificar i quantificar les seqüències Alu desmetilades (AUMA - *Amplification of UnMethylated Alus* - i QUMA - *Quantification of UnMethylated Alus* -) (Rodríguez et al., 2008b), que ens han confirmat la hipometilació dels elements Alu en el càncer colorectal i suggereixen que aquests elements es troben localitzats en la cromatina activa participant en els processos cel·lulars, incloent funcions implicades en el desenvolupament i la identitat cel·lular. De fet, el perfil epigenètic d'alguns d'aquests elements pot ser una font important de l'herència i la variabilitat somàtica amb manifestacions fenotípiques com el risc de càncer, la regulació anormal de les vies de senyalització de les cèl·lules canceroses i el paper en els processos de diferenciació.

Així doncs, creiem que els elements Alu desmetilats presenten unes característiques dins l'ALUOMA que els concedeixen les propietats adients per actuar com a elements reguladors amb la capacitat de modificar l'expressió d'un gen concret a diferents nivells.

L'objectiu principal d'aquesta tesi és identificar elements Alu amb un estat de cromatina activa (inicialment detectats per la seva desmetilació) i determinar el seu paper en la regulació de l'expressió gènica.

Per assolir aquest objectiu es proposen els següents objectius específics:

1. Caracteritzar els elements Alu desmetilats tant a nivell estructural (seqüència), funcional (context genòmic) i evolutiu.
2. Definir els patrons epigenètics dels elements Alu desmetilats associats a diferents tipus cel·lulars.
3. Establir els mecanismes de regulació gènica en els que participen els elements Alu, segons el seu estat de metilació, dins la tumorigènesi i la diferenciació.

**MATERIAL I MÈTODES**

## 1. MATERIAL

### 1.1. LÍNIES CEL·LULARS

S'ha treballat amb diferents línies de carcinoma i adenocarcinoma de còlon humà obtingudes de l'ATCC (*American Type Culture Collection*):

HCT116, àmpliament utilitzada en estudis de la biologia del càncer, presenta inestabilitat de microsatèl·lits i una mutació en el codó 13 del protooncogen *Ras*, a més sobreexpressa *TGFβ1* i *TGFβ2* (transforming growth factor β1 i β2). Es tracta d'una cèl·lula de carcinoma de còlon humà amb cariotip diploide.

HCT116-DKO, és una variant de la línia HCT116 deficient de les *DMNT1* i *DMNT3b* (ADN metiltransferasa 1 i 3b) (Paz, Wei et al. 2003), línia cedida pel Dr. Esteller (PEBC-IDIBELL).

Sw480, presenta mutacions en els gens *p53*, *myc*, *ras*. Es tracta d'una línia hipotriplòide amb inestabilitat cromosòmica.

LoVo, presenta mutacions en els gens *p53*, *myc*, *ras*. Es tracta d'una línia d'adenocarcinoma de còlon humà amb cariotip hiperdiploide i inestabilitat de microsatèl·lits.

CaCo2, presenta mutacions en *Tp53* i *APC* i inestabilitat cromosòmica. Es tracta d'una línia d'adenocarcinoma de còlon humà amb cariotip hiperdiploide.

HT29, presenta mutacions en *Tp53* i *APC*, ampliació de *c-myc* i un fenotip d'inestabilitat cromosòmica i aneuploidia i és d'adenocarcinoma de còlon humà.

Totes elles presenten una morfologia epitelial i creixement adherit al flascó de cultiu. El medi amb el que creixen és DMEM:F12 (Dulbecco's Modified Eagle medium : F12) complementat amb el 10% de SBF (sèrum boví fetal) i amb 20 μM de L-glutamina i 10 μM de piruvat (tots els reactius de GIBCO® Invitrogen, Carlsbad, CA, USA). Les condicions de creixement són 37°C, 5% de CO<sub>2</sub>, 95% d'aire i 100% d'humitat relativa.

Per altra banda també s'han utilitzat altres línies cel·lulars cedides per la Dra. M. José Barrero (Centre de Medicina Regenerativa de Barcelona, CMRB): línies cel·lulars de fibroblasts (HFF i F) i pellets cel·lulars de cèl·lules mare (hESC2, hESC4, hESC6), *embrioid bodies* (EBs), cèl·lules mare pluripotents induïdes originàries de queratinòcits (KIPS4F1, KIPS4F8) i de queratinòcits de bulb pilós [H]KIPS4F1.



## 1.2. MOSTRES HUMANES

Els teixits estudiats procedeixen de pacients diagnosticats de càncer colorectal en l'Institut Català d'Oncologia (Hospital de Bellvitge, Barcelona) i en l'Hospital Universitari Germans Trias i Pujol (Badalona). Les mostres van ser prèviament caracteritzades per inestabilitat de microsatèl·lits (MSI), mutacions en els gens p53 i K-ras.

De tots ells es va obtenir mucosa normal i teixit tumoral en forma de teixit fresc conservat a -80°C o bé en blocs de parafina i d'OCT. Les mostres es van recollir seguint el protocol aprovat pel comitè ètic de l'Hospital corresponent.

Les mostres de sang van ser obtingudes en l'Institut Català d'Oncologia (Hospital de Bellvitge, Barcelona) de 10 donants sans.

## 1.3. MOSTRES DE PRIMATS NO HUMANS

S'han utilitzat mostres de sang de diferents primats: quatre casos de *Pan Troglodytes* (ximpanzé), cinc de *Gorilla Gorilla* (goril·la) i cinc casos de *Pongo pygmaeus* (orangutan).

L'ADN dels primats es van obtenir del *Banc Taxonlife* (fundació Echevarne, Barcelona). El protocol d'obtenció de l'ADN que van emprar va ser el del kit EZ1 DNA Blood 200ul (QIAGEN®, Venlo, Holanda).

## 2. MÈTODES

### 2.1. TRACTAMENTS/ DROGUES

Per estudiar els efectes de la metilació en la línia d'adenocarcinoma de còlon HCT116 es van utilitzar diferents inhibidors de la metilació de l'ADN: 5-aza-2'deoxicitidina (5-AzadC) (*Sigma Aldrich, St Lois, MO, USA*) com agent desmetilant, trichostatina A (TSA) (*Sigma*) com inhibidor de la histona deacetilasa i un combinat d'ambdós agents.

Es van sembrar quatre plaques per tractament de les cèl·lules HCT116 en plaques de 10 cm de diàmetre, a diferents concentracions a fi que a l'hora de recollir-les, i després dels diferents tractaments, les concentracions s'igualesin. Així doncs per les HCT116 control es van sembrar  $3.5 \cdot 10^5$  cèl·lules, per les HCT116 que seran tractades amb 5-AzadC  $5 \cdot 10^5$  cèl·lules, per les HCT116 que seran tractades amb TSA  $4 \cdot 10^5$  cèl·lules i per les HCT116 tractades amb ambdós agents  $6 \cdot 10^5$  cèl·lules.

Passades 24 hores s'afegeix 5-AzadC a una concentració final de  $0.5 \mu\text{M}$  a les plaques corresponents. 24 hores més tard es realitza un canvi de medi a totes les plaques i es torna a afegir 5-AzadC a una concentració final de  $0.5 \mu\text{M}$  a les plaques corresponents. Al dia següent, de nou, es canvia el medi a totes les plaques i s'afegeix TSA a una concentració final de  $0.3 \mu\text{M}$  a les plaques corresponents.

Al cap de 16 hores es recullen les cèl·lules, es renten amb PBS (tampó fosfat salí) fred i els precipitats es guarden a  $-80^\circ\text{C}$  per posteriors anàlisis.

L'eficiència dels tractaments es va avaluar a nivell d'expressió gènica i metilació de l'ADN.

Els gens control utilitzats en aquest estudi van ser *EN1* (*Engrailed 1*) i *CPLX2* (*Complexina 2*) (seqüència i condicions dels encebadors en l'*annex I taula A1.1*), tots dos hipermetilats i, per tant silenciat. En estudis previs ja s'havia vist la reexpressió d'aquests gens a partir del tractament amb els agents desmetilats 5-AzadC i TSA (Mayor et al., 2009; Rodriguez et al., 2008b).

### 2.2. EXPRESSIÓ A NIVELL D'ARN

#### 2.2.1. EXTRACCIÓ DE L'ARN TOTAL

L'extracció d'ARN total es va fer seguint el protocol convencional d'extracció amb *Trizol* (*Invitrogen, Carlsbad, CA, USA*). La qualitat de l'ARN és resol en un gel d'agarosa a l'1%. Després es procedeix a quantificar l'ARN amb l'espectròmetre *Nanodrop 1000* (*Thermo Scientific, Rockford, IL, USA*). Es guarda l'ARN a una concentració de  $250 \text{ ng}/\mu\text{l}$ .

### 2.2.2. OBTENCIÓ DE L'ADN COMPLEMENTARI

L'ARN obtingut es transforma a ADNc (ADN complementari o de cadena senzilla) per mitjà de la retrotranscriptasa inversa.

Primerament s'afegeix 1 µl d'hexamers aleatoris d(N)<sub>6</sub> 150 ng/µl (*GE healthcare, Piscataway, NJ, USA*) i 2 µl de dNTPs 5mM (*GE healthcare*) a 500 ng d'ARNm. S'escalfa 5 minuts a 65°C per eliminar possibles estructures secundàries de l'ARN. Posteriorment s'afegeixen la resta de reactius: 1µl de l'enzima de retrotranscripció M-MLV (*Invitrogen*) a 200 U/µl, 0.35 µl d'inhibidor de ribonucleasa *RNAasin*<sup>®</sup> 40 U/µl (*Promega, Madison, WI, USA*), 2 µl DTT (Dithiothreitol) 0.1M i 5 µl de tampó 5X.

La reacció es realitza en un termociclador, 10 minuts a 25°C, 50 minuts a 37°C per permetre l'òptima actuació de la retrotranscriptasa inversa i 15 minuts a 70°C per inactivar la reacció.

### 2.2.3. PCR QUANTITATIVA A TEMPS REAL

Mitjançant la tecnologia de la PCR (Reacció en Cadena de la Polimerasa) amb l'agent intercalant SYBR<sup>®</sup> Green i la utilització del termociclador *Light Cycler*<sup>®</sup> 480 (*Roche Diagnostics Corporation, IN, Indiana, USA*) que integra una font làser d'emissió de llum i un sistema de detecció mitjançant una càmera CCD. Una seqüència d'àcid nucleic pot ser detectada i quantificada ja que l'increment d'ADN en cada cicle d'amplificació és proporcional a la fluorescència emesa.

#### TÈCNICA

La reacció es fa amb una mescla d'1 µl d'ADNc a 50 ng amb 5 µl *Fast Start DNA master SybrGreen I* (mescla que ja conté els dNTPs, Magnesi, tampó i Taq polimerasa) (*Roche*), 3 µl d'aigua bidestilada, 0.5 µl d'encebador invers 20 µM i 0.5 µl d'encebador directe 20 µM.

Cada punt es fa per triplicat.

El programa de la reacció és similar en tots els casos, variant únicament la temperatura d'amplificació, limitada per les característiques dels encebadors utilitzats (Figura M.1).

Program Name	denaturation						
Cycles	1	Analysis Mode	None				
Target (°C)	Acquisition Mode	Hold (hh:mm:ss)	Ramp Rate (°C/s)	Acquisitions (per °C)	Sec Target (°C)	Step size (°C)	Step Delay (cycles)
95	None	00:10:00	4.80		0	0	0
Program Name	amplification						
Cycles	50	Analysis Mode	Quantification				
Target (°C)	Acquisition Mode	Hold (hh:mm:ss)	Ramp Rate (°C/s)	Acquisitions (per °C)	Sec Target (°C)	Step size (°C)	Step Delay (cycles)
95	None	00:00:15	4.80		0	0	0
62	None	00:00:25	2.50		0	0	0
72	Single	00:00:30	4.80		0	0	0
Program Name	melting						
Cycles	1	Analysis Mode	Melting Curves				
Target (°C)	Acquisition Mode	Hold (hh:mm:ss)	Ramp Rate (°C/s)	Acquisitions (per °C)	Sec Target (°C)	Step size (°C)	Step Delay (cycles)
95	None	00:00:15	4.80		0	0	0
70	None	00:01:00	2.50		0	0	0
96	Continuous		0.06	10	0	0	0
Program Name	cooling						
Cycles	1	Analysis Mode	None				
Target (°C)	Acquisition Mode	Hold (hh:mm:ss)	Ramp Rate (°C/s)	Acquisitions (per °C)	Sec Target (°C)	Step size (°C)	Step Delay (cycles)
40	None	00:00:30	2.50		0	0	0

Fig. M.1. Programa del Light Cycler®, permet programar les temperatures i temps de desnaturalització, amplificació i fusió de la mostra. En vermell està marcada la temperatura d'amplificació, variable, que canviem dependent de les característiques de l'encebador utilitzat.

## DISSENY DELS ENCEBADORS

Els encebador tindran de 18 a 22 nucleòtids. El percentatge de CG ha de ser similar entre l'encebador directe i indirecte per tenir una temperatura de fusió ( $T_m$ , de l'anglès *melting temperature*) sigui similar. Cal tenir en compte la següent fórmula:

$$T_m = ( 59.9 + 0.45 * \%CG - \frac{625}{pb} )$$

On  $pb$  són els número de nucleòtids totals del encebador i  $\%CG$  és el percentatge de dinucleòtids CG que hi ha en l'encebador. Preferiblement, les temperatures de fusió haurien de ser pròximes als 58°C.

Un cop s'han dissenyat les parelles de encebadors es determina la presència de *hairpins* i *dimers* utilitzant programes informàtics com el *Windows 32 PrimerSelect 3.10 ©1993-1998 DNASTAR Inc.* A més també es fa un Blat per veure si la parella d'encebadors són específics pel nostre fragment. El producte amplificat amb la PCR hauria de ser, com a molt, pròxim a 300 pb per que el resultat de la PCR a temps real sigui bo.

## NORMALITZACIÓ

Primer caldrà definir alguns conceptes:

$C_p$ : Valors que obtenim del LightCycler. Són el punt de tall o de l'anglès, *Crossingpoints* ( $C_p$ )(Figura M.2a). El  $C_p$  correspon a la fracció del número de cicle en el que la fluorescència generada pel SybrGreen® sobrepassa el llindar en que l'amplificació és

exponencial. Aquest valor és considera inversament proporcional a la concentració de la mostra.

T<sub>m</sub>: Temperatura de fusió específica del fragment amplificat que s'està buscant (Figura M.2b).

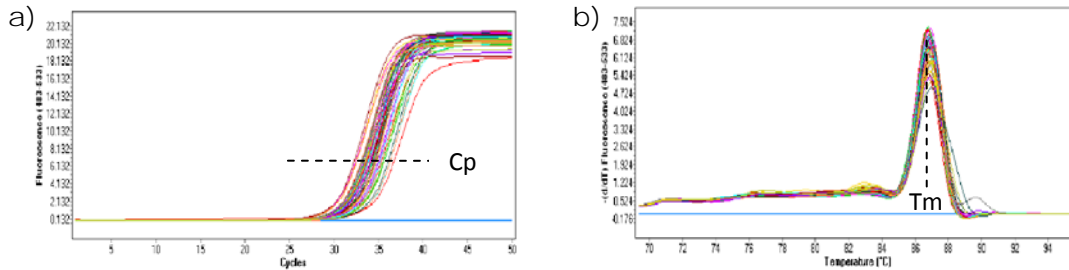


Fig. M.2. Valors obtinguts amb el software del Light Cycler®.

a) Corba d'amplificació del Light Cycler®, el Cp (del anglès Crossing point) és el punt de tall en el que l'amplificació és exponencial.

b) Pic de fusió del Light Cycler®, marca el T<sub>m</sub> (del anglès melting temperature).

Eficiència: l'eficiència (E) és el valor de la pendent de la recta patró que s'obté després de fer una dilució seriada de 6 punts i cada punt per triplicat. Respon a la fórmula:

$$E = 10^{\left(-\frac{1}{pendent}\right)}$$

Així doncs, a l'hora de normalitzar les nostres dades d'expressió utilitzem aquesta fórmula:

$$RATIO = \frac{E^{-\Delta CP_{mostra}}}{E^{-\Delta CP_{standard}}}$$

L'expressió gènica dels gens d'interès es va fer mitjançant la PCR a temps real. Els gens d'interès van ser aquells que tenien un element Alu proper i les seqüències i condicions dels encebadors es troben en l'annex I taula A1.3, A1.4 i A1.5. Els estàndards que s'utilitzen són  $\beta 2$  microglobulina, PUM1, 18S, Ciclofilina A PPA, PMSC4 i MRLP19 (seqüència dels encebadors en l'annex I taula A1.2) i es fa un promig de tots 5 per fer la normalització.

La desviació, es calcula fent la desviació estàndard dels CPs de l'Amostra (per triplicat) o dels estàndards (per triplicat).

## 2.2.4. NORTHERN BLOT

### DISSENY DELS ENCEBADORS

Primer obtenim l'ADN complementari a partir d'una RT-PCR convencional (protocol en l'apartat 2.2.2) i amplifiquem la seqüència que s'utilitzarà com a sonda per PCR convencional (Taula M.5) (el disseny i condicions dels encebadors es troba en l'annex I taula A1.11).

Amb el producte obtingut fem una lligació amb el kit pGEM-T EASY VECTOR I (Promega Biotech Iberica). Utilitzem 1.5µl de l'ADN purificat més 2.5 µl del tampó 2x, 0.5 µl del vector pGEMT i 0.5 µl de T4lligasa. Ho deixem 16h a 16°C.

El producte de la lligació el transformem utilitzant les cèl·lules competents DH5 Barregem suaument 5 µl de lligació amb 50 µl de bacteries E.Coli competents (DH5α) i incubem 20 minuts a 4°C (per permetre la unió de l'ADN a la paret bacteriana), 45 segons a 42°C (per permetre l'entrada de l'ADN dins DH5α) i 2 minuts a 4°C.

Fem créixer els bacteris 45 minuts a 37°C en 950 µl de LB i passat aquest temps sembrem 200 µl del cultiu en una placa d'LB agar amb ampicil·lina (50 µg/ul) com antibiòtic de selecció i 30 µl de X-Gal (Promega) per indicar els vectors que han integrat l'insert. Deixem les plaques a 37°C 16h. Passat aquest temps piquem 10 de les blanques resultants i les resuspenem en 15 µl d'H<sub>2</sub>O.

Comprovem si és el producte esperat per PCR, utilitzant encebadors específics de la seqüència d'interès i encebadors M13 (*Invitrogen*) que inclouen la seqüència del vector pGEMT-Easy.

La reacció de la PCR (Taula M.5) és fa amb una mescla de 3 µl d'ADN amb 2.5 µl de dNTPs, 1mM, 0.2 µl de Magnesi 100 mM, 2.5 µl de tampó10x, 0.2 µl Taq polimerasa, 14.8 µl d'aigua bidestilada, 1 µl d'encebador invers extern 20 µM (5' GAGCGGATAACA ATTCACACAGG 3') i 1 µl d'encebador directe extern 20 µM (5' GCCAGGGTTTT CCCAGT CACGA 3').

TAULA M.5. PROGRAMA DE PCR PER NORTHERN BLOT

PCR				
Programa		°C	temps	Cicles
<b>Desnaturalització</b>		95	5 min	1
<b>Amplificació</b>	Desnaturalització	95	1 min	28
	Hibridació	68	45 seg	
	Elongació	72	1 min 15 seg	
<b>Elongació</b>		72	5 min	1
		16	∞	1

El producte transformat correctament es resuspen amb LB amb antibiòtic (ampicil·lina) i s'incuba durant 16 hores a 37°C.

Finalment es fa una *miniprep* per purificar el producte crescut utilitzant el *kit Gene elute Plasmid (Sigma)*. La identitat del producte purificat es comprova per seqüenciació. Si aquest es correcte es fa una PCR convencional i es purifica el producte amb el kit de Genomed (*Zymoresearch*), es quantifica en el Nanodrop i es corre en un gel d'agarosa.

### MARCATGE RADIOACTIU DE LA SONDA

Agafem 45 ng del producte purificat, l'escalfem 3 minuts a 95°C i el lliguem fent servir les *beads* del kit *Ready-to-Go DNA Labelling Beads (dCTP)* (*GE Healthcare*). Després afegim 5 µl del fòsfor radioactiu dCTP [ $\alpha$ -<sup>32</sup>P]- 3000 Ci/mmol i escalfem 30 minuts a 37°C. Purifiquem amb el kit *Illustra ProbeQuant G-50 Micro Columns* (*GE Healthcare*).

Les sondes marcades es poden guardar a -20°C fins al seu ús. Cal tenir en compte que es tracta de marcatge radioactiu i per tant la radioactivitat va decaient al llarg del temps.

### ELECTROFORESI

Es preparen les mostres a fi de tenir 20 µg d'ARN per pou. La mostra es prepara amb 50% formamida desionitzada, 6% formaldehid i 1% MOPS1X. La mostra s'escalfa 10 minuts a 65°C i li afegim LB6X (de l'anglès *Loading Buffer*) a concentració final de 1x. Com a marcadors de carrega utilitzem, el *Ladder ssRNA* (*New England Biolabs*) per l'ARN i el  $\lambda$  Bstell per l'ADN.

La mostra es corre en un gel desnaturalitzant 1.1% d'agarosa, 1x MOPS i 6% formaldehid amb un tampó de MOPS1X durant 18 hores aproximadament a 20V.

### TRANSFERÈNCIA

El gel amb les mostres es renta tres cops amb solució 20X SSC pH 7 (175.3 g/l de NaCl i 88.2 g/l de Na<sub>3</sub>citrat-2H<sub>2</sub>O) en agitació i a temperatura ambient durant 15 minuts.

La transferència es fa en una safata amb tampó 20x SSC i un paper Whatman humit que està en contacte constant amb el tampó de la safata. A sobre va el gel sobre el qual es posa la membrana prehidratada de Nylon carregada positivament (*Roche*), varis papers Whatman, paper absorbent i un pes per permetre l'absorció del tampó SSC20x i la transferència de l'ARN del gel a la membrana per capil·laritat (Figura M.9).

El muntatge es deixa unes 18 hores aproximadament, moment en el que es recupera la membrana i es fixa l'ARN que hi ha en ella exposant la membrana a la llum ultraviolada a través de *Stratalinker® 2400 UV Crosslinker, 230V* (*Cultek S.L.U. Madrid, Espanya*), segons les condicions recomanades.

Posteriorment rentem la membrana amb H<sub>2</sub>O DEPC i la tenyim amb blau de metilè per visualitzar l'ARN ribosòmic (18S i 28S).

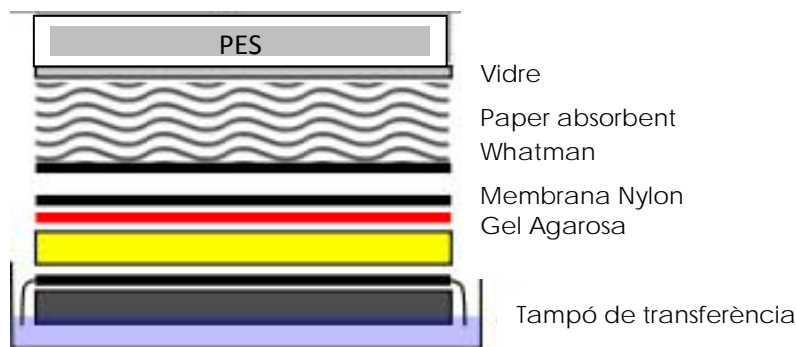


Fig. M9. Esquema del muntatge d'un Northern Blot en el moment de la transferència.

### HIBRIDACIÓ DE LA SONDA

Prehibridem la membrana en un volum apropiat del tampó comercial *DIG Easy Hyb* (Roche) a 42°C durant, almenys, 30 minuts en agitació.

La sonda, prèviament marcada, l'escalfem 5 minuts a 95°C i l'afegim al tampó *DIG Easy Hyb*. S'incuba aproximadament 18 hores a 42°C i en agitació.

Després d'hibridar fem diferents rentats. Primer es fa un rentat de baixa astringència (2XSSC, 0.1%SDS) durant 5 minuts a temperatura ambient un parell de vegades i després es renta amb un tampó d'alta astringència (0.1%SSC, 0.1%SDS) 15 minuts a 65°C. Amb el comptador Geiger ens assegurem de que la senyal és localitzada i la membrana està suficientment neta.

### DETECCIÓ

Les hibridacions amb sondes control o estàndards (Ciclofilina) es revelen en una pel·lícula i les sondes específiques del gen s'analitzen amb el Thypoon 9400 (Biorad).

## **2.3. RACE**

La tècnica RACE prové de les sigles en anglès 5' and 3' *Rapid Amplification of cDNA Ends* (Zhumabayeva et al., 2004). Utilitzem el kit *The SMARTer™ RACE cDNA Amplification Kit* de Clontech, permetent la ràpida ampliació dels extrems 3' i 5' del nostre gen d'interès.

Després de comprovar i posar a punt la tècnica RACE amb una mostra control i uns encebadors específics facilitats en el kit, realitzem l'amplificació dels extrems 3' i 5' en les nostres mostres.

### DISSENY DELS ENCEBADORS GSP

En la tècnica RACE s'utilitzen 2 tipus encebadors. Un universal i facilitat pel kit de Clontech (UPM de l'anglès *Universal Primer Mix*) i un altre específic del fragment que volem amplificar (zona coneguda) (GSP1-R en el cas de l'extrem 5' i GSP-F en el cas de l'extrem 3').

Pel que fa als encebadors GSPs es recomana que es trobin entre 23 i 28 nt, amb un contingut de GC del 50-70%, una Tm superior a 70°C i específics a la seqüència d'interès. Les seqüències i condicions dels encebadors GSP dissenyats per aquest estudi es troben en l'*annex I taula A1.12*.

En la figura M.10 es mostra un esquema d'on es situarien els encebadors GSP i que les regions que s'amplifiquen per 5'-RACE i per 3'-RACE.



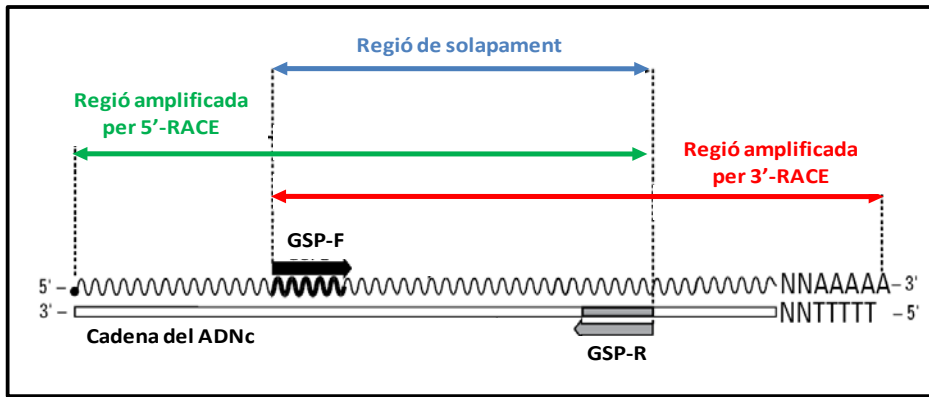


Fig. M.10. Disseny dels cebadors GSP, específics de la regió d'interès. En verd es mostra la regió que seria amplificada pel 5'-RACE, en vermell la que s'amplificaria pel 3'-RACE i en blau la regió de solapament.

### OBTENCIÓ DEL RACE ADNc 5' I 3'

En cada cas partim d'1µg d'ARN total per obtenir l'ADNc 5' i l'ADNc 3'.

Primerament en un tub preparam 1µl RNA (1µg), 1.75 µl H<sub>2</sub>O i encebador 5' CDS A (serà la mostra 5'-RACE ADNc) i en un altre tub 1µl RNA (1µg), 2.75 µl H<sub>2</sub>O i encebador 3' CDS A (serà la mostra 3'-RACE ADNc). Barregem i incubem 3 minuts a 72°C i 2 minuts a 42°C. Després afegim 1µl de SMARTER II oligo en el tub de 5'RACE (el SMARTER II oligo conté una extensió terminal de residus G que anella amb la cua de l'ADNc rica en dC servint com a base per iniciar la retrotranscripció en l'extrem 5').

Posteriorment afegim 2 µl 5X *First Strand Buffer*, 1µl DTT, 1µl dNTPs (5mM), 0.25 µl RNase Inhibitor (40 U/µl) i 1µl *SMARTScribe Reverse Transcriptase* (100 U) (aquesta transcripció reversa té activitat transferasa terminal i afegeix 3-5 residus, predominantment dC, a l'extrem 3' de la primera cadena d'ADNc). Barregem i incubem 90 minuts a 42°C i 10 minuts a 70°C. Diluïm el producte amb 100 µl de tampó Tricine-EDTA.

### AMPLIFICACIÓ DELS RACE ADNc

Un cop tenim els RACE ADNc, es procedeix a amplificar el producte utilitzant els cebadors específics del fragment que volem amplificar (zona coneguda) (GSP1-R en el cas de l'extrem 5' i GSP-F en el cas de l'extrem 3') i l'encebador universal (*UPM*). En la taula M.6 mostrem els diferents components de la PCR.

La master mix consta de 34.5 µl H<sub>2</sub>O, 5 µl 10X tampó Advantage 2 PCR, 2 µl dNTP (5 mM) i 1 µl 50X Advantage 2 Polymerase Mix.

TAULA M.6. COMPONENTS DE LA PCR D'AMPLIFICACIÓ DEL RACE ADNc 3' I 5'

components	5' - RACE				3' - RACE			
	MOSTRA	GSP R+F (C+)	UPM (C-)	GSP1-R (C-)	MOSTRA	GSP R+F (C+)	UPM (C-)	GSP1-F (C-)
5' -RACE ADNc	2.5µl	2.5µl	2.5µl	2.5µl	-	-	-	-
3' -RACE ADNc	-	-	-	-	2.5µl	2.5µl	2.5µl	2.5µl
encebador 5' GSP1- R (10µM)	1µl	1µl	-	1µl	-	1µl	-	-
encebador 3' GSP1- F (10µM)	-	1µl	-	-	1µl	1µl	-	1µl
encebador UPM mix (10x)	5µl	-	5µl	-	5µl	-	5µl	-
H <sub>2</sub> O	-	4µl	1µl	5µl	-	4µl	1µl	5µl
Master Mix	41.5µl	41.5µl	41.5µl	41.5µl	41.5µl	41.5µl	41.5µl	41.5µl
Volum final	50µl	50µl	50µl	50µl	50µl	50µl	50µl	50µl
Mida esperada del fragment	mida entre R i UPM	mida entre R i F	no producte	no producte	mida entre F i UPM	mida entre R i F	no producte	no producte

Tant en 5'-RACE com 3' fem servir varis controls per assegurar el correcte funcionament de la tècnica. Com a control positiu fem servir una PCR on amplifiquem utilitzant els encebadors GSP-F i GSP-R i com a control negatiu fem servir dos, un sense posar l'encebador específic (GSP) i l'altre sense posar l'encebador universal (UPM). El programa de PCR (Taula M.7) dependrà dels diferents encebadors GSP dissenyats.

TAULA M.7. PROGRAMA DE PCR PER RACE

PCR				
Programa		°C	temps	Cicles
<b>Desnaturalització</b>		94	5 seg	5
		72	3 min*	
<b>Amplificació</b>	Desnaturalització	94	30 seg	5
	Hibridació	70	30 seg	
	Elongació	72	3 min*	
<b>Amplificació</b>	Desnaturalització	94	30 seg	25
	Hibridació	68	30 seg	
	Elongació	72	3 min*	

\*Si el fragment esperat és major a 3kb afegirem 1minut per cada kb addicional.

El producte de PCR es resol en un gel d'acrilamida al 2% per veure els diferents fragments i la mida d'aquests.

#### CLONACIÓ I SEQÜENCIACIÓ DEL PRODUCTE RACE

Les diferents bandes del gel es retallen, es purifiquen amb el kit Nucleospin Gel and PCR Clean-up (Machery-Nagel, GmbH&Co. KG, Germany) i es quantifiquen en el NanoDrop.

Amb el producte obtingut fem una lligació amb el kit pGEM-T EASY VECTOR I (Promega Biotech Iberica). Utilitzem 1.5 µl de l'ADN purificat més 2.5 µl del tampó 2x, 0.5 µl del vector pGEMT i 0.5 µl de T4lligasa. Ho deixem 16h a 16°C.

El producte de la lligació el transformem utilitzant les cèl·lules competents DH5. Barregem suaument 5 µl de lligació amb 50 µl de bacteris *E. Coli* competents (DH5α) i incubem 20 minuts a 4°C, 45 segons a 42°C i 2 minuts a 4°C.

Fem créixer els bacteris 45 minuts a 37°C en 950 µl de LB i passat aquest temps sembrem 200 µl del cultiu en una placa d'LB agar amb ampicil·lina (50 µg/µl) com antibiòtic de selecció i 30 µl de X-Gal (Promega) per indicar els vectors que han integrat l'insert. Deixem les plaques a 37°C 16h. Passat aquest temps piquem 10 de les blanques resultants i les resuspenem en 15 µl d'H<sub>2</sub>O.

Comprovem si és el producte esperat per PCR, utilitzant encebadors específics de la seqüència d'interès i encebadors M13 (*Invitrogen*) que inclouen la seqüència del vector pGEMT-Easy.

La reacció de la PCR (Taula M.8) es fa amb una mescla de 3 µl d'ADN amb 2.5 µl de dNTPs 1mM, 0.2 µl de Magnesi 100 mM, 2.5 µl de tampó10x, 0.2 µl Taq polimerasa, 14.8 µl d'aigua bidestilada, 1 µl d'encebador invers extern 20 µM (5' GAGCGGATAACA ATTCACACAGG 3') i 1 µl d'encebador directe extern 20 µM (5' GCCAGGGTTT CCCAGT CACGA 3').

TAULA M.8. PROGRAMA DE PCR CARACTERITZAR EL PRODUCTE DEL RACE

PCR				
Programa		°C	temps	Cicles
<b>Desnaturalització</b>		95	5 min	1
<b>Amplificació</b>	Desnaturalització	95	1 min	28
	Hibridació	68	45 seg	
	Elongació	72	1 min 15 seg	
<b>Elongació</b>		72	5 min	1
		16	∞	1

Purifiquem el producte obtingut amb el kit de Genomed (*Zymoresearch*), es quantifica en el Nanodrop i es corre en un gel d'agarosa. La identitat del producte purificat es comprova per seqüenciació.

## 2.4. WESTERN BLOT

### EXTRACCIÓ DE PROTEINES

Partim d'un precipitat d'1\*10<sup>6</sup> cèl·lules, el qual resuspenem amb 100 µl de tampó RIPA (150 mM NaCl, 1% X-Tritó100, 0.5% deoxycholate, 0.1%SDS, 50 mM Tris pH8, 3mM DTT, 30

mM PMSF, 1 mM ortovanadat, 5 mM NaF ) i es deixa 30 minuts en gel. Després *vortexem* 10 segons i fem disrupció mecànica amb una xeringa d'insulina. Centrifuguem 30 minuts a 16000 x g a temperatura ambient. Recuperem sobrenedant i el quantifiquem en el fent servir el sistema de luminescència BCA (Thermo Scientific).

Per enriquir l'extracte amb proteïnes nuclears (*Andrews and Faller, 1991*) partim d'un precipitat de  $5 \times 10^6$  cèl·lules, el qual resuspenem amb 400 µl de tampó A (10 mM HEPES-KOH pH 7.9, 1.5 mM MgCl<sub>2</sub>, 0.5 mM DTT i 0.5 mM PMSF) i es deixa 10 minuts en gel. Després *vortexem* 10 segons i centrifuguem 10 segons a 16000 x g a temperatura ambient. Descartem el sobrenedant. Al precipitat resultant afegim 50 µl tampó C ric en sals, (20mM HEPES-KOH pH 7.9, 25% glicerol, 420 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.5 mM DTT, 5 mM PMSF, 0.5 mM aprotinina, 0.5 mM pepstatina) i es deixa 20 minuts en gel. Després centrifuguem a 16000 x g 2 minuts a 4°C. Recuperem sobrenedant i el quantifiquem en el Nanodrop.

### ELECTROFORESI

Els polipèptids es separen per pes molecular mitjançant electroforesi en poliacrilamida i SDS (SDS-PAGE).

Depenent del pes de la nostra proteïna d'interès preparam un gel separador a un tant per cent determinat de poliacrilamida (0.375 M Tris-HCl pH 8.8, X% Acrilamida, 0.1%, SDS, 0.02% TEMED, 0.004% persulfat d'amoni -APS-) i la part apiladora al 3% (0.125 M Tris-HCl pH 6.8, 3% Acrilamida, 0.1% SDS, 0.02% TEMED, 0.004%APS).

La mostra es prepara utilitzant 30 µg de proteïna amb tampó de càrrega Laemmli (Laemmli, 1970) (60 mM Tris-HCl pH 6.8, 2% SDS, 10% glicerol, 5% β-mercaptoetanol, 0.01% blau de bromofenol) a fi de desnaturalitzar les proteïnes durant 5 minuts a 95°C.

El gel es corre amb tampó d'electroforesi 1X partint d'una solució 5X (0.05 M Tris-HCl pH 8.8, 0.834 M glicina pH 8.3, 0.1% SDS). Utilitzem com a marcador de pes molecular el *Precision Plus Protein All Blue Standards* (Biorad, Hercules, CA 94547, U.S.A) que té un rang de 10 fins a 250 kDa.

### ELECTROTRANSFERÈNCIA

Utilitzem una membrana de PVDF *Immobilon-P Membrane, PVDF, 0.45 µm* (Millipore Iberica S.A.) i realitzem la transferència en el tampó Towbin (Towbin et al., 2001) que conté 25 mM Tris pH 8.3, 192 mM glicina i 20% metanol, durant 1 hora a 100V.

### BLOQUEIG I INCUBACIÓ ANTICÓS

Es bloqueja la membrana en solució bloquejant (TBS1X, 0.1% Tween 20, 5% Moliko milk) a 4°C tota la nit i en agitació. Rentem dos cops la membrana 15 minuts amb TBS1X amb 0.1% Tween 20.

Hibridem la membrana amb l'anticòs primari (Taula M.5) diluït amb solució bloquejant durant 1 hora i 30 minuts. Després repetim els rentats amb TBS1X amb 0.1% Tween 20.

TAULA M.9. ANTICOSSOS EMPRATS

	ANTICÓS	Casa comercial	referència	dilució	procedència
primari	Anti-DIEXF	Abcam	Ab111058	1/250	Conill
	Anti-GLDC	Abcam	Ab97625	1/250	Conill
	Anti-Tubulina	Sigma	T5168	1/4000	ratolí
secundari	Anti-mouse HRP	Dako	P0447	1/10000	cabra
	Anti-rabbit HRP	Dako	P0448	1/3000	cabra

Hibridem la membrana amb l'anticòs secundari corresponent (Taula M.9) diluït amb solució bloquejant durant 1 hora i 30 minuts. Després repetim els rentats amb TBS1X amb 0.1% Tween 20.

### DETECCIÓ

Per la detecció utilitzem *Super Signal West Pico Chemiluminescent Substrate* (Thermo Scientific). Es tracta d'un substrat quimioluminiscent que reacciona amb la peroxidasa de rave (HRP) present en l'anticòs secundari.

Es col·loca la membrana en una placa reveladora i s'exposa amb una pel·lícula 15 segons, 1 minut o 15 minuts depenent de la intensitat de la banda. Després es quantifiquen les bandes per anàlisi densitomètric.

## 2.5. siRNA (ARN de silenciament)

El siRNA (ARN de silenciament o de l'anglès *small interfering RNA*) és un ARN interferent de doble cadena d'uns 20-21 nucleòtids de longitud específic per la seqüència de nucleòtids del seu ARN missatger diana, interferint en l'expressió del gen. Els siRNA suprimeixen l'expressió dels gens diana tallant l'ARNm (ARN missatger).

Per seleccionar els siRNA de transfecció (Taula M.10) fem servir el programa que facilita Invitrogen (<https://rnaidesigner.invitrogen.com/rnaiexpress/>) pel seu disseny, *Block-iT™ iRNA designer*.

Treballarem amb plaques de 6 pous amb un volum d'1.5 ml per pou. Sembrem  $1.5 \cdot 10^5$  cèl·lules per pouet, depenent del tipus cel·lular. Passades 24 hores iniciarem la transfecció.

Diluïm el siRNA (20  $\mu$ M inicials) amb Optimem (*Invitrogen*) a 10 (20 pmols), 20 (40 pmols), o 50 nM (100 pmols) depenent del siRNA (Taula M.8.) en un volum de 250  $\mu$ l i deixem incubant 10 minuts a temperatura ambient.

Per altra banda preparam la *Lipofectamine2000* (*Invitrogen*). La Lipofectamina crea una coberta lipídica al voltant de l'ADN o ARN que es vol incorporar en la línia

cel·lular. Al incubar aquests “liposomes” són captats i incorporats per les cèl·lules. Treballem a una concentració de 3 µg/ml diluït en Optimem en un volum de 250 µl. Deixem incubant 10 minuts a temperatura ambient.

TAULA M.10. DISSENY DELS siRNA

siRNA	posició	Nucleòtid d'inici	Seqüència	Casa comercial	Dilució (pmols)	Temps (hores)
siRNA1	UTR3'	5057	GCAGCAAGCAAGTGCCTAT	Invitrogen	50	24
siRNA 2	UTR3'	7118	GCTATTGTTGCCCTATAT	Invitrogen	50	24
siRNA 3	ORF	840	CCACCTGGACTAAGACCAA	Invitrogen	50	24
siRNA 4	ORF	2299	GCTACAGTCCAACAAGAAT	Invitrogen	50	24
siRNAuniv	scrambled			Sigma	125	24

A continuació mesquem la solució de Lipofectamina amb la solució que conté el siRNA i incubem 20 minuts a temperatura ambient per permetre la formació del “liposomes”.

Afegim els 500 µl de la mescla en el pou corresponent.

Cada experiment tindrà varis controls: les cèl·lules sense transfectar, les cèl·lules amb Lipofectamina i les cèl·lules transfectades amb el siRNA control universal *MISSION® siRNA Universal Negative Control #1(Sigma)* (taula M.10).

Les cèl·lules es recullen a les 24, 48 i 72 hores per veure l'eficiència de transfecció.

Per veure si l'experiment ha funcionat correctament es comprovarà per diferents vies:

- Expressió mitjançant PCR quantitativa a temps real, seguint el protocol del apartat 2.2.3 d'aquesta secció i el disseny dels encebadors en l'annex I figura A3.
- Expressió mitjançant Western Blot, per veure si la proteïna es veu afectada, seguint el protocol de l'apartat 2.4. d'aquesta secció.

## 2.6. ESTUDI DE LA METILACIÓ

### 2.6.1. TRACTAMENT PER BISULFIT

Actualment hi ha una gran varietat de mètodes que proporcionen informació quantitativa i qualitativa sobre l'estat de metilació de l'ADN genòmic (Butcher and Beck, 2008; Laird, 2010).

Les primeres aproximacions es centraven en l'estudi de nivells globals de metil - citosina, actualment estan més centrats en l'estudi de l'estat de metilació de seqüències específiques de l'ADN. Concretament, l'optimització dels mètodes basats en el tractament de l'ADN amb bisulfit sòdic permet l'anàlisi d'un nombre limitat de CpGs basant-se en el seu estat de metilació i amb la caracterització de patrons específics de metilació (seqüenciació d'ADN modificat amb bisulfit sòdic)(Clark et al., 2006; Stirzaker et al., 2004) (Figura M.3).

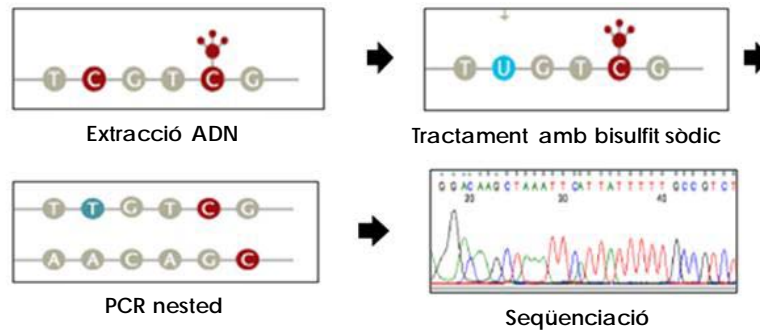


Fig. M.3 . Procés de transformació de l'ADN per mitjà del tractament per bisulfit sòdic.

### EXTRACCIÓ DE L'ADN

Partim d'un precipitat que conté aproximadament  $1 \cdot 10^6$  cèl·lules. Primerament afegim 1ml de SET1X (150 mM NaCl, 20 mM Tris-HCL pH7.8 i 1 mM EDTA) i 0.2  $\mu\text{g}/\text{ml}$  de RNAsa. Ho deixem incubant 1 hora a  $37^\circ\text{C}$ . Després afegim Proteinasa K (*Invitrogen*) (0.1 $\mu\text{g}/\text{ml}$ ) ho agitem i deixem incubant tota la nit a  $37^\circ\text{C}$ .

Passat aquest temps afegim 1 volum de fenol i 50 mM de NaCl. Agitem i centrifuguem a 16000 x g durant 15 minuts. Recuperem la fase aquosa i afegim 1 volum de Fenol cloroform isoamil. Centrifuguem a 16000 x g durant 15 minuts. Recuperem la fase aquosa i afegim 1 volum de Cloroform isoamil. Centrifuguem a 16000 x g durant 15 minuts. Recuperem la fase aquosa i afegim 2 volums de etanol absolut. Ho deixem durant tota la nit a  $-20^\circ\text{C}$ .

Passat aquest temps centrifuguem a 16000 x g durant 15 minuts a  $4^\circ\text{C}$  i descartem el sobrenedant. El precipitat el rentem amb etanol al 70%. Centrifuguem a 16000 x g durant 15 minuts a  $4^\circ\text{C}$  i descartem el sobrenedant. Sequem bé el precipitat i el resuspenem amb 25  $\mu\text{l}$   $\text{H}_2\text{O}$  bidestilada.

Quantifiquem l'ADN en el NanoDrop i comprovem el seu estat amb un gel d' agarosa al 1%. Portem totes les mostres a 200 ng/ $\mu\text{l}$ .

### TRACTAMENT PER BISULFIT SÒDIC

El bisulfit sòdic transforma les citosines no metilades en uracils, en canvi, les citosines metilades es mantenen com a citosines. Aquesta reacció permet diferenciar ADN metilat del no metilat (Figura M.3). Seguim el protocol del *kit* de modificació de l'ADN *EZ DNA Methylation KIT* (*Zymo-research, Orange, CA, USA*). Eluïm el producte amb 50  $\mu\text{l}$  d' $\text{H}_2\text{O}$  bidestilada fent que tinguem en cada vial 350 ng.

## PCR NESTED

### DISSENY DELS ENCEBADORS:

Per amplificar el fragment que es vol analitzar es realitza una PCR *nested* (dos reaccions de PCR consecutives). Primer, s'amplifica l'ADN transformat per bisulfit amb uns encebadors externs i a partir del producte obtingut es realitza una segona PCR (amb encebadors interns). Tal i com indica el nom, els encebadors interns estan situats més a prop del fragment d'interès que els encebadors externs.

Els encebadors tindran de 18 a 22 nucleòtids. Al dissenyar-los cal tenir en compte que les citosines no metilades seran timines i les citosines metilades seguiran sent citosines. Donat que la metilació es dona en el dinucleòtid CpG, cal evitar la presència d'aquests en l'encebador. En els casos en els que no és possible dissenyar un encebador sense el dinucleòtid CpG, es millor que estigui situat a l'extrem 5' de l'encebador i utilitzar encebadors degenerats, es a dir, que hibridin tant sobre l'ADN no metilat (que seria TpG un cop tractat amb bisulfit i amplificat per PCR) com el metilat (que continuaria sent CpG).

Es preferible que en l'extrem 3' hi hagin citosines (CC), que al no estar metilades, s'hauran transformat. D'aquesta forma, si hi ha amplificació es corrobora que el tractament de bisulfit ha funcionat.

El percentatge de CG ha de ser similar entre l'encebador directe i indirecte per tal que la temperatura de fusió ( $T_m$ , de l'anglès *melting temperature*) sigui similar.

Cal tenir en compte la següent formula:

$$C/G=4 \text{ y } A/T=2. \rightarrow T_m = \sum(A/T) + \sum(G/C)$$

Segons aquesta, és millor que la  $T_m$  dels encebadors externs sigui inferior o igual a la dels interns. Preferiblement les temperatures d'hibridació haurien de ser pròximes als 58°C.

Un cop s'han dissenyat les parelles de encebadors es determina la presència de *hairpins* i *dimers* utilitzant programes informàtics com el *PrimerSelect 3.10* ©1993-1998 DNASTAR Inc. A més també es fa un Blat per veure si la parella d'encebadors són específics pel nostre fragment.

En el tractament de bisulfit es degrada l'ADN, així que el producte amplificat amb la PCR externa pot ser, com màxim, de 600 pb i el de la interna ha de ser preferiblement entre 150 pb - 400 pb.

### REACCIÓ DE LA PCR:

La reacció de la PCR externa és fa amb una mescla de 2 µl d'ADN modificat per bisulfit sòdic amb 2.5 µl de dNTPs 1 mM, 0.2 µl de Magnesi 100 mM, 2.5 µl de tampó10x, 0.2 µl Taq polimerasa, 16.6 µl d'aigua bidestilada, 0,5 µl d'encebador invers extern 100 µM i 0,5 µl d'encebador directe extern 100 µM (seqüència i condicions dels encebadors en l'*annex I taula A1.6, A1.7 i A1.8*).



La reacció de la PCR interna és fa amb una mescla de 2 µl del producte de la PCR externa (bé de manera directa o bé diluint 1/10 depenent de les necessitats del fragment) amb 2.5 µl de dNTPs 1 mM, 2.5 µl de tampó10x, 0.2 µl Taq polimerasa, 16.8 µl d'aigua bidestilada, 0,5 µl d'encebador invers intern 100 µM i 0.5 µl d'encebador directe intern 100 µM (seqüència i condicions dels encebadors en l'annex I taula Al.6, Al.7 i Al.8) per duplicat.

En la taula M.1 és representa el programa de la PCR *nested*, on la Tm2 és la temperatura d'hibridació dels encebadors externs i la Tm4 és la temperatura d'hibridació dels encebadors interns, obtingudes a partir de la formula abans descrita. Cal tenir en compte que Tm2= Tm1+2 i que Tm4 = Tm3+2.

TAULA M.1. PROGRAMA DE LA PCR NESTED

PCR externa				
Programa		°C	temps	Cicles
Desnaturalització		95	4 min	1
Amplificació a	Desnaturalització	95	45 seg	5
baixa astringència	Hibridació	Tm <sub>1</sub>	2 min	
	Elongació	72	3 min	
Amplificació a	Desnaturalització	95	45 seg	30
Alta astringència	Hibridació	Tm <sub>2</sub>	1 min	
	Elongació	72	1m30s	
Elongació		72	5 min	1
		4	∞	1
PCR interna				
Programa		°C	temps	Cicles
Desnaturalització		95	4 min	1
Amplificació a	Desnaturalització	95	45 seg	5
baixa astringència	Hibridació	Tm <sub>3</sub>	1 min	
	Elongació	72	1min30s	
Amplificació a	Desnaturalització	95	45 seg	30
Alta astringència	Hibridació	Tm <sub>4</sub>	1 min	
	Elongació	72	1m30s	
Elongació		72	5 min	1
		4	∞	1

## SEQÜENCIACIÓ

El producte de la PCR interna és purifica amb el kit per columnes *Jetquick* (*Genomed, St Louis, MO, USA*). S'agafa d'1 a 4 µl del producte purificat, depenent de la quantitat que s'hagi obtingut, i s'afegeix 1 µl de *BigDye*, 1µl de tampó, 1 µl del encebador intern 3,2 µM invers o directe depenent de per quin extrem es seqüencii millor el fragment.

La seqüenciació s'ha fet utilitzant el Serveis Científico-Tècnics del *Parc Científic* i de l'Institut Germans Trias i Pujol (IGTP) amb el sistema *BigDye Terminator V3.1 Cycle Sequencing Kit* (*Applied Biosystems, Foster City, CA, USA*).

La lectura de les seqüències es fa utilitzant el programa *Finch TV versió 1.4.0* (*Geospiza Inc.*), en l'opció *Raw*. Segons amb quin encebador haguem seqüenciat, directe o invers, el fragment es llegirà de 3' a 5' o de 5' a 3' respectivament (Melki et al., 1999). Per tant, si llegim una seqüència feta amb l'encebador directe, les CG que apareguin com a CG estaran metilades i les que apareguin com a IG estaran desmetilades (figura M.4). Veient l'alçada relativa dels pics podem determinar si aquesta metilació és parcial o total. Per altra banda, si el fragment ha estat seqüenciat amb l'encebador invers, les CG que apareguin com a CG estaran metilades i les que apareguin com a CA estaran desmetilades (Figura M.5). Veient els pics veurem si aquesta metilació és parcial o total.

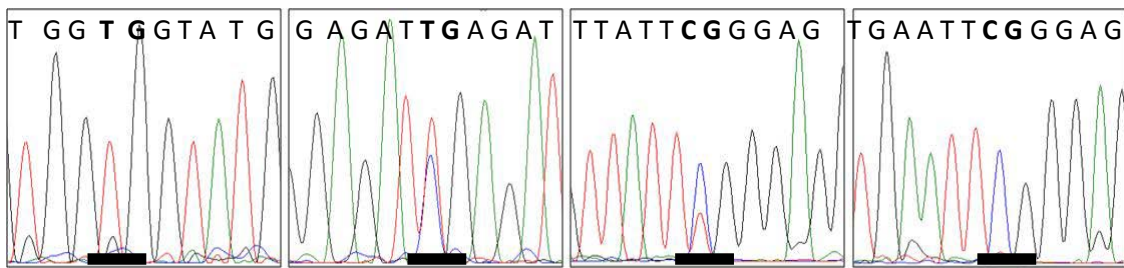


Fig. M.4. Representació d'un fragment seqüenciat amb l'encebador directe. Cada nucleòtid està representat per un pic de diferent color (C blau, G negre, T vermell i A verd). La franja en negre marca el dinucleòtid CG. En el primer gràfic veiem una seqüència desmetilada (IG), en el segon veiem que la CG està parcialment metilada (25-50%), en el tercer gràfic la CG també està parcialment metilada (75%) i en el quart gràfic la CG està totalment metilada (100%).

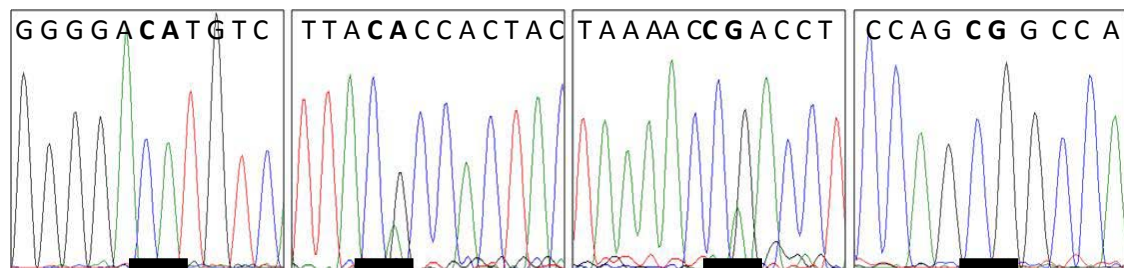


Fig. M.5. Representació d'un fragment seqüenciat amb l'encebador indirecte. Cada nucleòtid està representat per un pic de diferent color (C blau, G negre, T vermell i A verd). En el primer gràfic veiem una seqüència desmetilada (CA), en el segon veiem que la CG està parcialment metilada (25-50%), en el tercer gràfic la CG també està parcialment metilada (75%) i en el quart gràfic la CG està totalment metilada (100%).

### 2.6.2. QUMA I AUMA

Per quantificar i identificar les Alus desmetilades el nostre grup va posar a punt dues noves tècniques anomenades QUMA, de les sigles angleses *Quantification of Unmethylated Alu's* i AUMA, de l'anglès *Amplification of Unmethylated Alu's* (Figura M.6) (Rodríguez et al., 2008b).

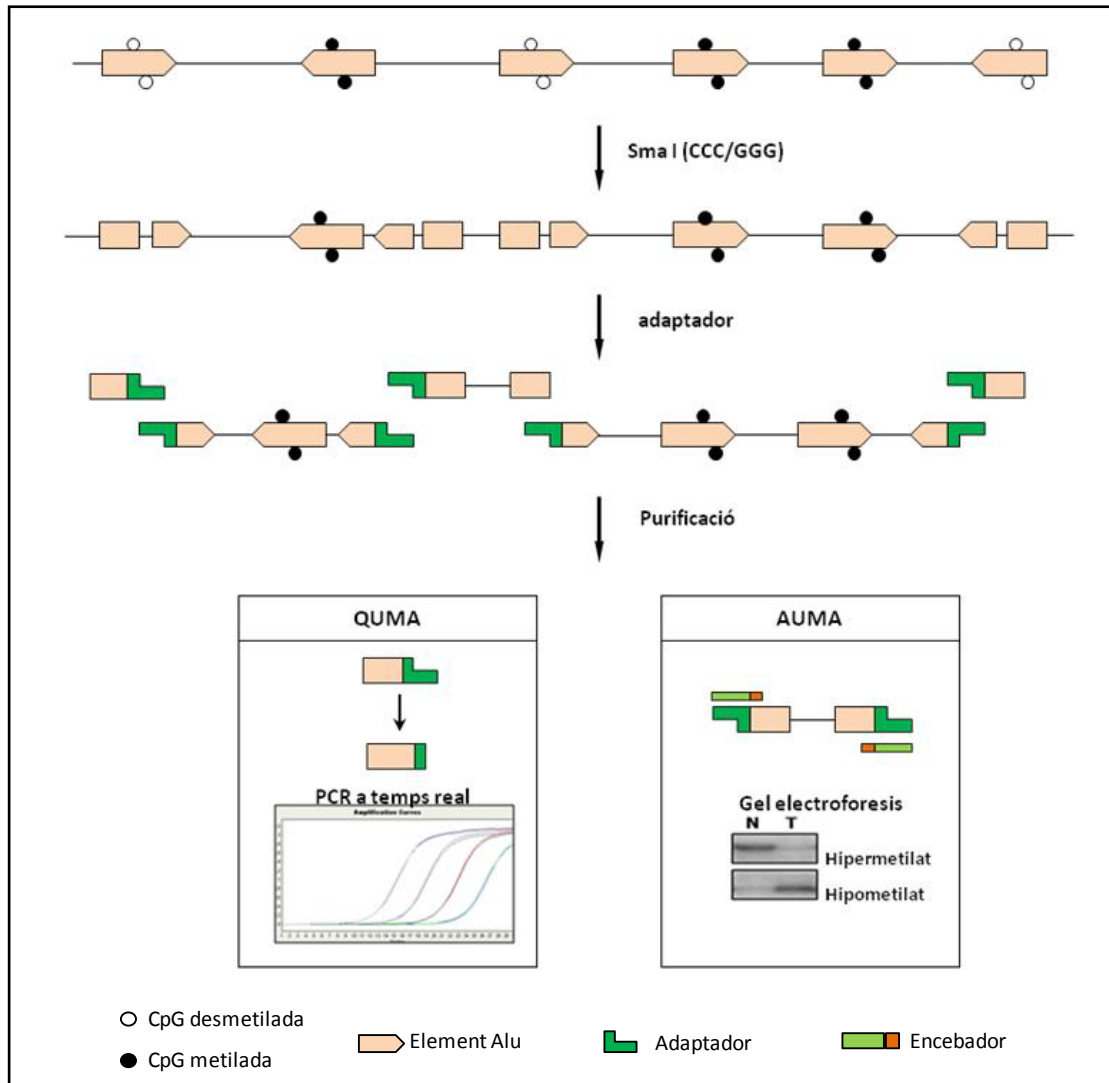


Fig. M.6. Esquema de les tècniques QUMA i AUMA. Metodologies aplicades per quantificar i identificar respectivament. Veure explicació detallada de les tècniques en el text.

#### QUMA

El lloc CpG dins de la seqüència consens de l'Alu, AACCCGGG, s'utilitza com a reporter de la metilació. Partim d'1 µg d'ADN i el digerim 16 hores a 25°C amb 20U de Sma I (CCCGGG) (New England Biolabs, Beverly, MA, USA), un enzim de restricció sensible a la metilació fent que les CpG no metilades es tallin.

Després de la digestió queden extrems roms (CCC/GGG) en els quals s'uneix un adaptador utilitzant 800U de T4 ADN lligasa i incubant 16 hores a 16°C.

Els adaptadors es preparen incubant els oligonucleòtids "blue" (CCGAATCG CAAAGCTCTGA) i el "5'MCF fosforilat" (<sup>P</sup>-TCAGAGCTTTGCGAAT) a 65°C durant 2 minuts i a temperatura ambient 30 minuts.

El producte que s'obté de la lligació es digereix 6 hores a 37°C amb Xma I (*New England Biolab*), enzim de restricció no sensible a la metilació que ens permetrà eliminar aquells elements Alu metilats sense digerir.

El producte final es purifica utilitzant el kit *GFX<sup>TM</sup>* (*GE healthcare*) i s'elueix amb 250 µl d'aigua bidestilada.

A partir d'1 ng del producte digerit i lligat obtingut en la primera etapa, es fa una PCR a temps real emprant un encebador consens a elements Alu que s'anella *upstream* a la regió Sma I (CCGTCTCTACTAAAATACA) i un altre encebador consens al adaptador més dues timines al final de l'extrem 3' (ATTCGCAAAGCTCTGAGGGTT).

Les magnituds s'expressen com a nombre d'elements Alu no metilats per genoma haploide després de normalitzar pel total d'ADN emprat.

Segons estudis previs, amb la tècnica QUMA es poden detectar 100 elements Alu no metilats per genoma haploide partint d'1 ng d'ADN per PCR.

### AUMA

La primera part d'aquesta tècnica (digestió amb Sma I, unió als adaptadors i purificació del producte final) és comú amb la QUMA, excepte que el producte de partida no pateix la digestió amb Xma I.

La segona etapa de l'AUMA consisteix en una amplificació per PCR (Taula M.2) utilitzant un encebador homòleg al adaptador amb els nucleòtids TT afegits al extrem 3' per enriquir en seqüències Alu (ATTCGCAAAGCTCTGAGGGTT). En cada PCR s'utilitza únicament un encebador.

El producte final és una col·lecció de fragments d'ADN de 100 a 2000 pb flauejats per la seqüència AACCCGGG desmetilada. El producte es resol en gels desnaturalitzants d'urea. Les bandes obtingudes es visualitzen després de dur a terme una tinció en plata.

Les seqüències generades per l'AUMA apareixen com un rastre (*fingerprint*) característic de cada mostra. Cada banda es pot identificar individualment. Donat que la astringència de l'AUMA es basa en una seqüència curta (AACCCGGG) que és troba preferentment, però no exclusivament en elements Alu, també es poden amplificar altres seqüències desmetilades. Com totes les Alu desmetilades són amplificades junt amb un fragment de seqüència única adjacent, és possible la seva identificació i el mapatge de l'element Alu.

Al comparar mostres aparellades tumor colorectal - mucosa de còlon normal es realitzen classificacions depenent del comportament del patró de bandes: hipometilació (si hi ha un increment d'intensitat en el tumor), hipermetilació (si decreix en el tumor) o sense canvi (si no hi ha canvi en la intensitat entre les dues mostres).

TAULA M.2. PROGRAMA DE PCR PER LA TÈCNICA QUMA

PCR				
Programa		°C	temps	Cicles
Desnaturalització		95	1 min	1
Amplificació	Desnaturalització	95	30 seg	34
	Hibridació	65	30 seg	
	Elongació	72	1min	
Elongació		72	5 min	1
		4	∞	1

### 2.3.3. NSUMA

Les sigles NSUMA venen de l'anglès *Next Generation Sequencing Unmethylated Alu's*.

Donat que la tècnica d'AUMA únicament detecta una petita part de les Alus del genoma humà, en el laboratori s'ha desenvolupat una variant de la tècnica per ampliar la cobertura d'elements Alu desmetilats.

El procediment de la tècnica NSUMA es mostra en la Figura 7.M.

El lloc CpG dins de la seqüència consens de l'Alu, AACCCGGG, s'utilitza com a reporter de la metilació. A l'igual que en l'AUMA, partim d'1 µg d'ADN i el digerim 16h a 25°C amb 20U d' Sma I (CCCGGG), un enzim de restricció sensible a la metilació. El producte que s'obté es torna a digerir 6h a 37°C amb Mse I (T/TAA) (*New England Biolabs*), enzim de restricció no sensible a la metilació, de manera que s'obtenen fragments més petits que podran ser amplificats per PCR. Després de la digestió queden extrems cohesius (T/TAA) i extrems roms (CCC/GGG) en els quals s'uneixen uns adaptadors utilitzant 800 U de T4 ADN lligasa (*New England Biolabs*) i incubant 16 hores a 16°C.

Els adaptadors es preparen incubant, per una banda els oligonucleòtids que s'uniran als extrems roms "S1 biotinilat" (biot-GATAGTATGCCCGGGTCA) i el "S2 fosforilat" (<sup>P</sup>-TCACCCGGGCATAC) i per altra banda els oligonucleòtids que s'uniran als extrems cohesius "M1" (CTGAGGCTGGATCCCTG) i el "M2 fosforilat" (<sup>P</sup>-TACAGGGATCCAGCCTCAG) a 65°C durant 2 minuts i 30 minuts a temperatura ambient.

El producte final es purifica utilitzant el kit GFX™ (*GEhealthcare*) i s'elueix amb 200 µl d'aigua bidestilada. Un cop purificat immobilitzem la biotina que contenen els fragments on s'ha unit l'adaptador "S1 biotinilat" (els fragments digerits amb SmaI i per tant desmetilats) amb boles magnètiques d'estreptavidina, *Streptavidine magna*

beads, Dynabeads® M-280, (Invitrogen). El producte s'elueix amb 100 µl d'aigua bidestilada.

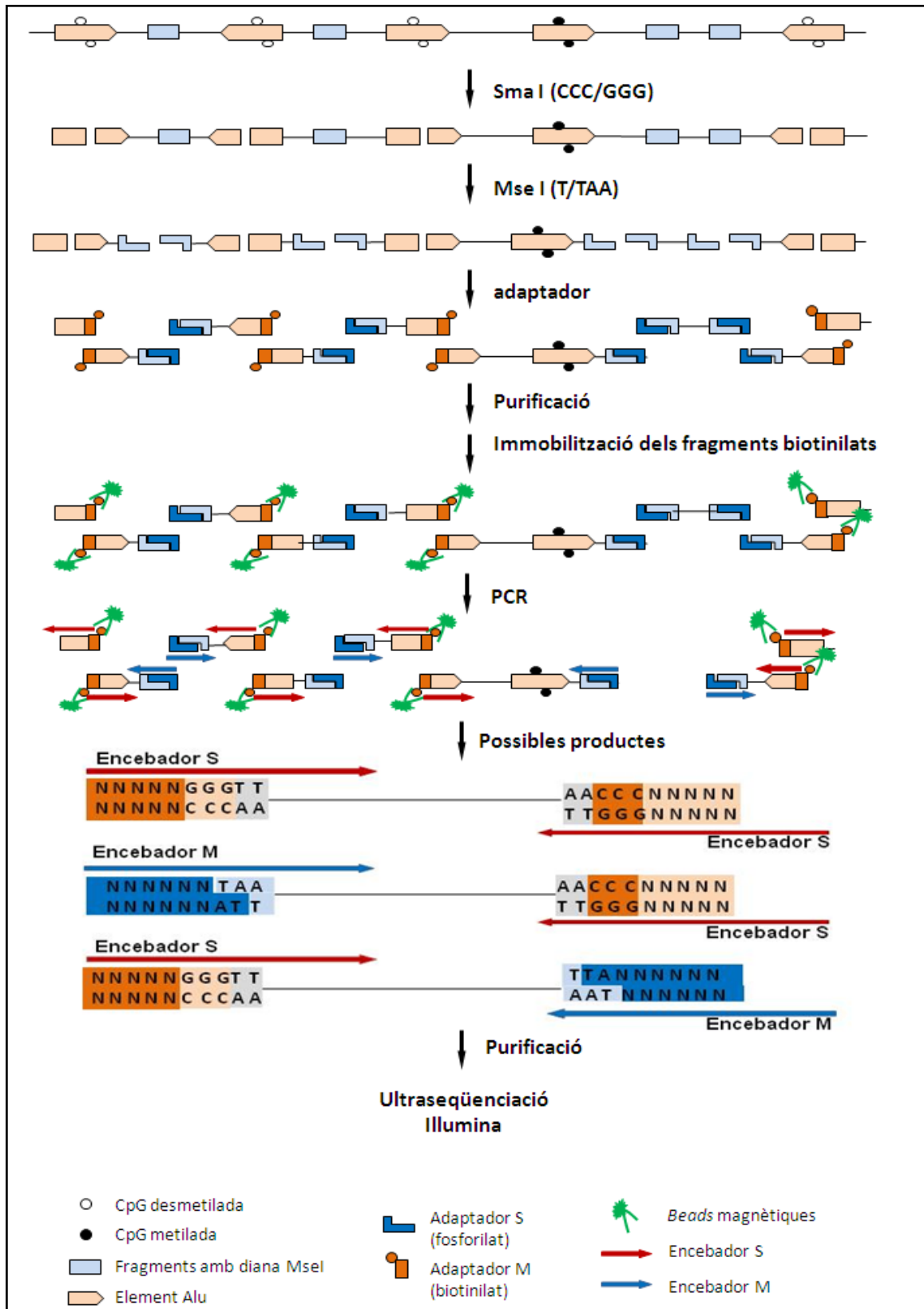


Fig. M.7. Esquema de la tècnica ZUMA

Posteriorment es porta a terme una amplificació per PCR utilitzant un encebador homòleg als adaptadors: "S-TT" (GATAGTATGCCCGGGTGAGGGTT) per una banda i "M" (CTGAGGCTGGATCCCTGTAA) per l'altra.

Per comprovar l'estat de la PCR inicial es fa una darrera PCR amb 2 µl del producte obtingut ja que el producte resultant d'aquesta seria un *smear* (Taula 3.M).

TAULA M.3. PROGRAMA PER LA PCR DE ZUMA

PCR				
Programa		°C	temps	Cicles
Desnaturalització		95	2 min	1
Amplificació	Desnaturalització	95	30 seg	30
	Hibridació	60	1 min	
	Elongació	72	1 min	
Elongació		72	5 min	1
		4	∞	1

S'utilitzen encebadors específics homòlegs als adaptadors: "S-TTCT" (GATAGTATGCCCGGGTGAGGGTTCT) per una banda i "M-CT" (CTGAGGCTGGATCCCTGTAACT) per l'altra.

Aquesta PCR es corre en un gel desnaturalitzant permetent veure un patró i assegurar-nos així que el producte que és seqüenciarà és l'adient.

El producte de la primera PCR és purifica amb el kit per columnes *Jetquick* (Genomed) i es quantifica en el Nanodrop.

5 µg de la mostra es porta a seqüenciar al PRBB on utilitzaran l'ultraseqüenciador Illumina (*HiSeq200*).

A grans trets, Illumina és una de les noves tecnologies conegudes com a seqüenciació d'alt rendiment, en les quals s'utilitzen mètodes que paral·lelitzen el procés de seqüenciació, produint mils o milions de seqüències al mateix temps. Degut a que els mètodes de detecció molecular sovint no són suficientment sensibles per seqüenciar una molècula, amb aquesta tècnica s'utilitza un pas de clonació *in vitro* per generar múltiples còpies de cada molècula individual. Hi ha diverses estratègies per portar a cap aquesta clonació. En el cas d'Illumina el que fa és una "PCR de pont", és a dir, els fragments s'amplifiquen a partir dels encebadors units a una superfície sòlida. Un cop les seqüències clonals es localitzen físicament en posicions separades de la superfície, es du a terme la seqüenciació per determinar les seqüències d'ADN de totes les localitzacions en paral·lel. En el cas d'Illumina s'utilitza el mètode de "terminadors reversibles", en el qual fan servir versions reversibles de terminadors marcats

fluorescentment. En anar afegint cada cop un nucleòtid es va detectant la fluorescència emesa corresponent a aquella posició.

## 2.7. ESTUDI DE LES MARQUES D'HISTONA: IMMUNOPRECIPITACIÓ DE CROMATINA

La Immunoprecipitació de Cromatina (*ChIP*) és un mètode bioquímic utilitzat per a determinar la localització en el genoma d'histones modificades i d'altres proteïnes *in vivo*. Aquesta tècnica consisteix en l'ús d'un anticòs que reconeix la proteïna d'interès no sols en dissolució sinó també en la cromatina (Figura M.8).

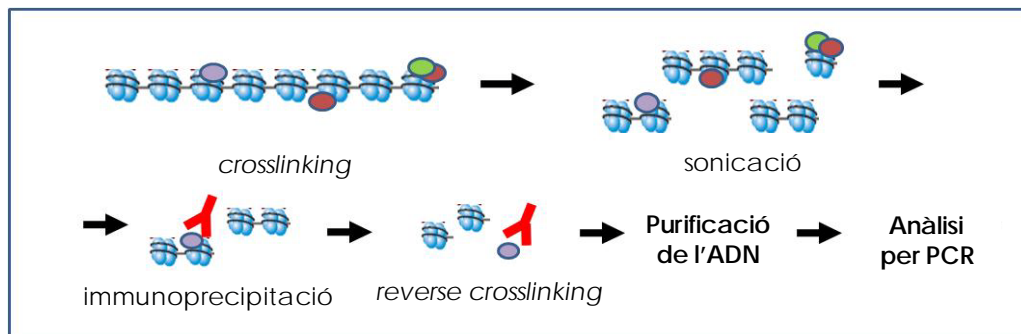


Fig. M.8. Esquema representatiu de la immunoprecipitació de cromatina pas a pas.

En la tècnica *ChIP* primerament es fa un entrecruament (*crosslinking*) amb formaldehid entre l'ADN i les proteïnes unides *in vivo*. D'aquesta manera es fixen les interaccions proteïna - proteïna i proteïna - ADN. Seguidament, a partir d'extractes que prèviament han estat sonicats per a fragmentar la cromatina es fa la immunoprecipitació dels complexos proteïna - ADN amb anticossos específics. A partir de l'ADN immunoprecipitat s'amplifica per PCR a temps real les regions d'interès, obtenint una mesura de la presència de les proteïnes o marques d'histones estudiades sobre aquesta seqüència.

Part dels experiments s'han fet amb un protocol manual, explicat a continuació, i la resta amb el kit *Chromatin Immunoprecipitation (ChIP) Assay kit* (Millipore, Billerica, MA, USA), que segueix el mateix fonament que el protocol manual, però resulta més pràctic, ràpid i reproducible.

### CROSSLINKING

Es parteix de precipitats cel·lulars que resuspenem amb PBS.

A fi d'encreuar la mostra s'afegeix formaldehid a una concentració final de 0.5%. Es deixa incubant 15 minuts a temperatura ambient, agitant. Per aturar la reacció s'afegeix Glicina a una concentració final de 0.125 M. Es centrifuga i es descarta el sobrenedant.

El precipitat es renta amb PBS complementat amb el "còctel" d'inhibidors de proteases *Complete mini*(Roche).



## SONICACIÓ

Preparem les mostres per que quedin a una concentració de  $3 \cdot 10^4$  cèl·lules /  $\mu\text{l}$  per immunoprecipitació.

El precipitat es resuspén amb un tampó de lisi (1% SDS, 5 mM EDTA, 50 mM TRIS, i amb un seguit d'inhibidors de proteases: 1 mM PMSF, 1  $\mu\text{g}/\text{ml}$  Aprotinina i 1  $\mu\text{g}/\text{ml}$  Pepstatina A (tots ells de *Sigma-Aldrich*). Ajustem el volum de mostra a 0.5 ml i es posen en tubs falcon de 15ml. Disgreguem la cromatina de les mostres amb el sonicador *Bioruptor*<sup>®</sup> (*Diagenode s.a. , Liege, Bèlgica*). Programem 60 segons mode encès, 60 segons mode apagat durant 10 cops a màxima potència. En aquest pas és molt important anar-se assegurant de que la temperatura de la mostra no puja, per tant anirem afegint gel de manera periòdica i es realitzarà tot en una cambra a 4°C.

Abans de continuar comprovarem l'estat de la cromatina, retirant 2  $\mu\text{l}$  de la mostra. Aquests 2  $\mu\text{l}$  els resuspendrem amb 18  $\mu\text{l}$  d'aigua i els descreuarem (de l'anglès *reverse crosslinking*) a 65°C durant 1 hora. Purificarem la mostra amb el Kit de Purificació *Jetquick* (*Genomed*) i analitzarem l'estat de la cromatina amb un gel d'agarosa l'1.5%. Si veiem que la mida de la cromatina està entre 300 i 500 pb continuarem el protocol.

El producte sonicat el centrifuguem a màxima velocitat durant 10 minuts i a 4°C. Recuperem el sobrenedant. Aquest s'haurà de diluir 1/10 amb el tampó de dilució (150 mM NaCl, 1% Tritó X-100, 2 mM EDTA, 20 mM TRIS amb un seguit d'inhibidors de proteases: 1 mM PMSF, 1  $\mu\text{g}/\text{ml}$  Aprotinina i 1  $\mu\text{g}/\text{ml}$  Pepstatina A) a fi de diluir l'alta concentració d' SDS de la mostra.

S' aliquota la mostra en 1ml per tub. En aquest punt guardarem la mostra a -80°C (la cromatina total obtinguda serà l'INPUT).

## IMMUNOPRECIPITACIÓ

Es renta la mostra (1ml) durant 2 hores rotant a 4°C amb 45  $\mu\text{l}$  de la suspensió al 50% de proteïna A i G (*GE Healthcare*), 2  $\mu\text{g}$  d'esperma de salmó i 10  $\mu\text{l}$  de sèrum normal de conill.

Passat aquest temps es centrifuga a 500 x g i es recupera el sobrenedant. El sobrenedant s'incuba tota la nit rotant a 4°C amb 5-10  $\mu\text{g}$  de l'anticòs d'interès (taula M.4).

Pels rentats s'afegeixen 45  $\mu\text{l}$  de la suspensió al 50% de proteïna A i G (*GE healthcare*) i s'incuba durant 1 hora a 4°C. Passat aquest temps descartem el sobrenedant i anem fent rentats amb quatre solucions diferents: TSE I (150 mM NaCl, 1% Tritó X-100, 2 mM EDTA, 20 mM TRIS, 0.1% SDS), TSE II (500 mM NaCl, 1% Tritó X-100, 2 mM EDTA, 20 mM TRIS, 0.1% SDS), TSE III (1% deoxycholate, 1 mM EDTA, 10 mM TRIS, 0.1% NP40, 0.25 M LiCl) i TE (10 mM TRIS, 1 mM EDTA) per duplicat.

TAULA M.4. ANTICOSSOS UTILITZATS PER LA IMMUNOPRECIPITACIÓ

ANTICÓS	Casa Comercial	Referència	Volum
AcK14H3	Upstate	07.053	5 µl
AcK9H3	Upstate	07.352	5 µl
H3K4me3	Upstate	17.614	3 µl
H3K9me3	Diagenode	pAb066.050	9 µl
H3K27me3	Upstate	07.449	5 µl
H3	Abcam	Ab1791	10 µl
IgG	Jackson Immunology	011.000.120	5 µl

Eluïm la cromatina unida al nostre anticòs afegint 75 µl de tampó d'elució (1% SDS, 0.1M NaHCO<sub>3</sub>). Aquest pas el fem dos cops quedant-nos al final amb 150 µl de la IP.

A continuació afegim 0.2M de NaCl tant a la IP (150 µl) com al INPUT (50 µl) per alliberar l'ADN (*reverse crosslinking*). L'IP representa la cromatina unida al nostre anticòs i l'INPUT la cromatina total.

Per eliminar les proteïnes afegim 8 µl EDTA 500 mM, 8 µl TRIS 2 M pH 6.2, 7.5 NaCl 5 M i 1ul de proteinasa K 20 µg/µl. Portem tant l'IP com l'INPUT a 0.5 ml amb aigua bidestilada i ho deixem incubar 1 hora a 55°C.

Un cop passat aquest temps fem una extracció convencional amb fenol –cloroform: Afegim 1 volum de Fenol cloroform isoamil. Centrifuguem a 16000 x g durant 15 minuts. Recuperem la fase aquosa i afegim 1 volum de fenol cloroform isoamil. Centrifuguem a màxima velocitat durant 15 minuts. Recuperem la fase aquosa i afegim 2 volums de etanol absolut i 20 µg de glicogen. Ho deixem en neu carbònica 20 minuts.

A continuació centrifuguem a 16000 x g durant 15 minuts a 4°C i descartem el sobrenedant. El precipitat el rentem amb etanol al 70%. Centrifuguem a 16000 x g durant 15 minuts a 4°C i descartem el sobrenedant i assequem bé el precipitat. Resuspenem els precipitats amb 50 µl amb aigua bidestilada en el cas de les IPs i amb 75 µl en el cas dels INPUTS. Les mostres es dilueixen ½ i es guarden a -20°C.

#### PCR QUANTITATIVA EN TEMPS REAL

La quantificació de la cromatina immunoprecipitada sobre les regions d'interès segons les modificacions d'histones seleccionades es fa per PCR a temps real. La tècnica, el disseny dels encebadors són les mateixes que les descrites en l'apartat 2.1.2.3. (seqüència i condicions dels encebadors en l'*annex I taula A1.9 i A1.10*).

A l'hora de quantificar les diferents marques s'utilitza la H3 total per normalitzar. La fórmula que s'aplicarà en aquest cas serà:

$$RATIO = Eficiència^{\Delta CP2}$$

on

$$\Delta CP2 = \Delta CP_{mostra} - \Delta CP_{H3}$$

La variància entre els duplicats es calcula:

$$Var(E^{\Delta CP2}) = \sqrt{((-RATIO * \ln Eficiència)^2 * (SD_{\Delta CP2})^2)}$$

on

$$SD_{\Delta CP2} = \sqrt{(SD_{mostra}^2 - SD_{H3}^2)}$$

En el cas de la representació de l'H3 total s'utilitza la següent normalització:

$$VALOR\ Absolut = Eficiència^{\Delta CP1}$$

on

$$\Delta CP1 = \Delta CP_{mostra} - \Delta CP_{input}$$

La desviació en aquest cas es calcula fent la desviació estàndard dels CPs de les mostra (feta per duplicat) o dels estàndards (fets per duplicat).

**RESULTATS**

## 1. CARACTERITZACIÓ A NIVELL DE SEQÜÈNCIA DELS ELEMENTS ALU DESMETILATS

Tal i com s'ha comentat en l'apartat d'antecedents, en estudis previs del grup es va analitzar una sèrie de 17 tumors colorectals i les seves parelles de teixit normal per mitjà de la tècnica QUMA (*Quantification of UnMethylated Alus*) i una sèrie de 50 parelles mitjançant AUMA (*Amplification of UnMethylated Alus*) (Rodríguez et al., 2008b). Aquestes tècniques ens van permetre estimar el nombre d'elements Alu desmetilats en teixit normal i tumoral (per QUMA) així com la identificació d'un gran nombre de seqüències hipermetilades i hipometilades en tumors colorectals, majoritàriament Alus (per AUMA).

Com a resultat d'aquests estudis es van trobar 94 elements Alu desmetilats total o parcialment en cèl·lules normals i/o tumorals de còlon. A més ens van revelar que les cèl·lules tumorals en còlon contenen aproximadament el doble d'elements Alu desmetilats que les cèl·lules epitelials normals.

Per determinar si les Alus desmetilades constitueixen una subpoblació particular i diferenciada dins de l'ALUOMA es van comparar a nivell bioinformàtic amb un grup de 100 elements Alu del genoma triats a l'atzar (que vam anomenar *random* i vam considerar representatius del conjunt d'elements Alu del genoma, la majoria dels quals estan metilats). A més, es van estudiar un conjunt de 100 Alus potencialment amplificables per la tècnica AUMA (que vam anomenar *virtuals*) per descartar qualsevol biaix tècnic ja que mitjançant l'AUMA només es poden estudiar aquells elements Alu que contenen una diana Smal, per tant només una fracció de totes les Alus del genoma (Rodríguez et al., 2008b).

Per fer l'estudi es van analitzar una sèrie de variables, obtenint les dades del *Genome Browser versió NCBI 38/hg18*.

En la taula R.1. veiem la comparació entre els elements Alu *random* (representatius del conjunt d'elements Alu del genoma) i els elements Alu *virtuals* (elements Alu amplificables per la tècnica AUMA). Observem que la tècnica AUMA presenta alguns biaixos tècnics ja que les Alus *virtuals* tenen algunes diferències significatives respecte l'ALUOMA (grup *random*): les Alus *virtuals* tendeixen tenir la cua poliA més curta, el percentatge de GC, així com l'índex de CpG Observades/Esperades, és més alt i la divergència tendeix a ser més baixa. Finalment també trobem menys Alus J del que caldria esperar i una sobrerrepresentació d' Alus Y. Totes aquestes característiques van associades a la presència de la diana Smal que determina la possibilitat d'una seqüència Alu d'estar representada o no en l'AUMA.

Donat el biaix en la representativitat de l'AUMA, vam considerar que els elements Alu *virtuals* eren el control més adient per comparar-los amb els 94 elements Alu parcial o totalment desmetilats identificats mitjançant AUMA i així poder definir trets genòmics característics i específics dels elements Alu desmetilats.

TAULA R.1. COMPARACIÓ DELS ELEMENTS ALU RANDOM I VIRTUALS

Elements Alu	Mida		Delecions		Insercions		cua poli(A) (pb) *	
	mitjana	sd	mitjana	sd	mitjana	sd	mitjana	sd
random	276,54	78,87	1,08	2,02	0,58	1,38	19,18	9,90
virtuals	293,16	34,51	0,86	1,64	0,93	1,75	8,79	10,71

Elements Alu	GC *		CpG O/E *		Divergència *	
	mitjana	sd	mitjana	sd	mitjana	sd
random	0,52	0,28	0,33	0,19	12,32	4,04
virtuals	0,54	0,00	0,53	0,16	9,02	3,13

Elements Alu	Orientació		Família *			Distància CpGi	
	5'→3'	3'→5'	J	S	Y	<3	>3
random	49	51	18	67	15	10	90
virtuals	48	52	3	66	31	11	89

Elements Alu	Context genòmic				
	UPSTREAM	UPSTREAM -promo-	INSIDE -intron-	INSIDE -exon-	DOWNSTREAM
random	16	0	40	5	39
virtuals	17	0	39	4	40

\*  $p < 0.05$  segons el test estadístic de Wilcoxon per mida, delecions, insercions, cua poli(A), contingut de GC, index CpG observades vers esperades i divergència, i  $p < 0.05$  segons el test de Fisher per orientació, família, distància a l' illa CpG més propera i context genòmic referent al gen més proper.

En la figura R.1 trobem els resultats de la comparació entre els elements Alu desmetilats (en vermell) i virtuals (en verd) de diferents variables.

Els paràmetres que no canvien al comparar ambdós grups són: la mida en parells de bases, les delecions, les insercions (Figura R.1.a) i l'orientació de l'element Alu (Figura R.1.e). El contingut de GC així com l'índex de CpG Observades/Esperades (Figura R.1.b i c) tampoc presenten canvis, sent el contingut de GCs majoritàriament del 50-55% en ambdós casos. També la divergència (Figura R.1.d) es manté igual entre elements virtuals i desmetilats, estant per sota del 10%. En aquest sentit, certs estudis han indicat que els elements Alu actius en el genoma presenten menys d'un 10% de divergència (Medstrand et al., 2002).

Altres paràmetres comparats presenten diferències significatives entre els elements Alu desmetilats i els virtuals:

La mida de la cua poli (A) (Figura R.1.a) veiem que és significativament més llarga que la dels elements Alu virtuals. Per altra banda, si observem la distribució segons la família a la que pertanyen (Figura R.1.f), veiem que les Alu desmetilades estan lleugerament enriquides en Alu S, aquesta família és la més freqüent i d'edat intermèdia. Pel que fa la família Alu Y (la família més jove) veiem que hi ha un lleuger empobriment en les Alu desmetilades. La família J està molt poc representada i per tant no es pot valorar.

Pel que fa a la distància de l'element Alu una illa CpG veiem que les Alu desmetilades normalment es troben més a prop d'una illa CpG (a menys de 3 kb) que els elements Alu virtuals (Figura R.1.g).

Finalment el context genòmic, és a dir la localització de l'element Alu en el genoma respecte al gen més proper, també és motiu d'estudi (Figura R.1.h). Veiem que els elements Alu desmetilats es localitzen en regions promotores (a menys de 1000 pb del TSS del gen més proper) en major freqüència que els elements Alu virtuals.

a)

Elements Alu	Mida		Deleccions		Insercions		cua poli (A) (pb) *	
	mitjana	sd	mitjana	sd	mitjana	sd	mitjana	sd
virtuals	293,16	34,51	0,86	1,64	0,93	1,75	8,79	10,71
desmetilats	276,44	55,70	1,39	3,38	0,87	1,55	18,85	10,19

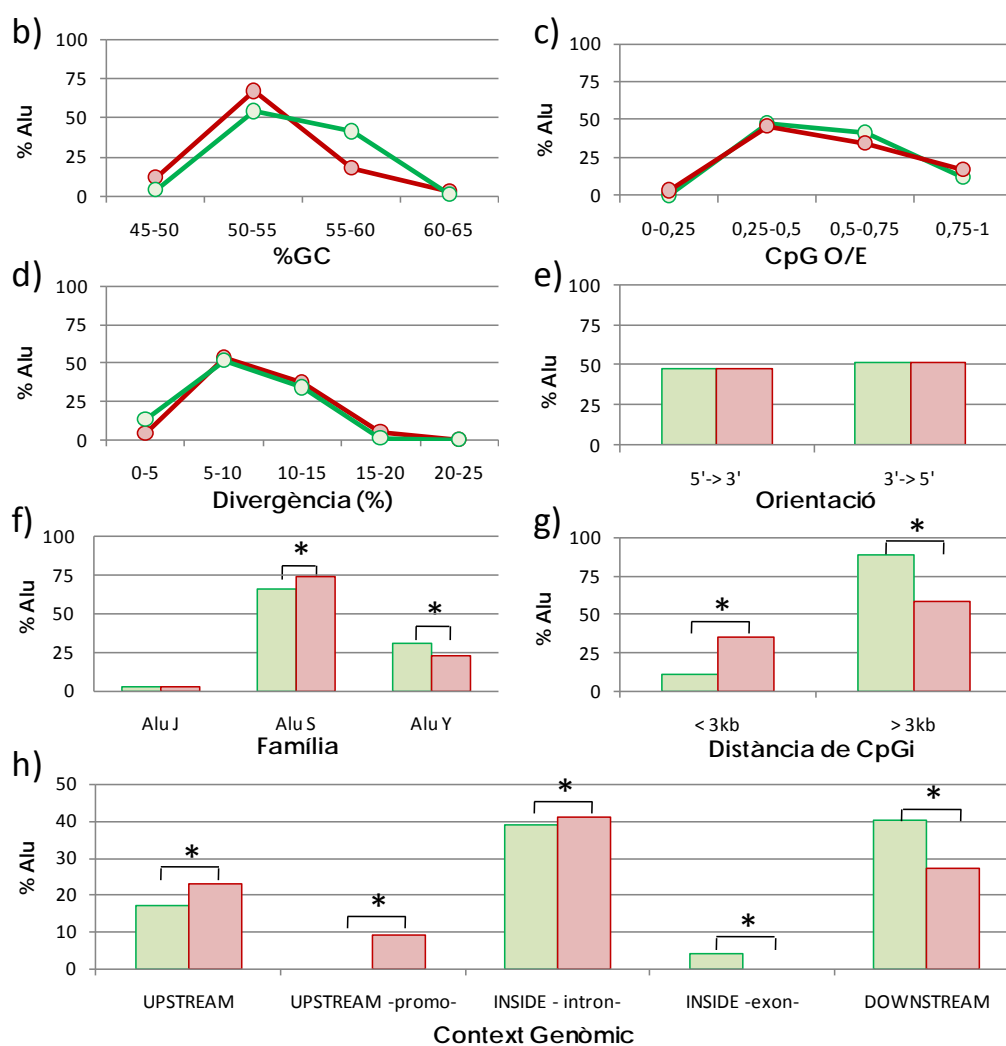


Fig. R.1. Trets genòmics específics dels elements Alu desmetilats. Comparació de 94 elements Alu desmetilats (en vermell) i 100 elements Alu virtuals (en verd). Amb un asterisc marquem  $p < 0.05$  segons el test estadístic de Wilcoxon per (a) mida, deleccions, insercions i cua poliA, (b) contingut de GC, (c) index CpG observades vers esperades, (d)divergència, i  $p < 0.05$  segons el test de Fisher per (e) orientació, (f) família, (g) distància a l'illa CpG més propera, (h) context genòmic respecte el gen més proper.

Així doncs, en teixit de còlon, tant normal com tumoral, veiem que els elements Alu desmetilats defineixen un compartiment diferencial dins de l'ALUOMA que es caracteritza per estar lleugerament enriquit en Alus S respecte Alus Y i J, per presentar una baixa divergència respecte la seqüència consens, i per situar-se prop d'inicis de transcripció associats a illes CpG.

La particularitat d'aquests elements Alu respecte al milió d'Alus del genoma humà va portar a un estudi més exhaustiu per investigar les seves possibles implicacions biològiques.



## 2. CARACTERITZACIÓ EPIGENÈTICA DELS ELEMENTS ALU

Per fer estudis en profunditat tant a nivell de metilació com de cromatina, es van seleccionar 20 elements Alu dels 94 parcial o totalment desmetilats en la mucosa de còlon normal i/o en cèl·lules de carcinoma colorectal. En base al seu perfil de metilació les podem classificar com a Alus hipometilades en tumor, Alus hipermetilades en tumor i Alus desmetilades en ambdues mostres (Taula R.2).

En la taula R.2 es detallen les característiques més importants de les Alus analitzades. L'estat de metilació segons els resultats de l'AUMA es mostren codificat en color: vermell significa hipermetilat en tumor, verd hipometilat en tumor i blau sense canvi de metilació al comparar tumor amb normal. La darrera columna mostra l'estat de metilació a partir de la tècnica de bisulfit sòdic i posterior seqüenciació, tècnica emprada per validar els resultats de l'AUMA, seguint el mateix codi de colors. Per problemes tècnics els 3 darrers elements Alu no es van poder analitzar (*nd* en la taula R.2).

TAULA R.2. ELEMENTS ALU SELECCIONATS

ID Banda	mida <sup>a</sup>	%GC <sup>b</sup>	CpG <sup>c</sup>	Família	Posició <sup>d</sup>	Distància del gen (bp) <sup>e</sup>	Localització dins del gen <sup>e</sup>	Gen <sup>e</sup>	Estat metilació AUMA	Estat metilació validació
Ar3 c7	301	56	16	Alu Y	chr5:167914086-167914386	INSIDE	intron 1	RARS	Verd	Blau
Lv165	304	56	10	Alu Sx	chr16:2010218-2010521	INSIDE	intron 1	NDUFB10	Verd	Blau
Lv183	209	58	14	Alu Sg/x	chr11:8931949-8932157	INSIDE	intron 1	ST5	Verd	Blau
Au104	306	53	14	Alu Sx	chr16:4527219-4527524	INSIDE	intron 1	HMOX2	Verd	Verd
Aj2 c1	321	55	18	Alu Sq	chr1:210000573-210000893	-440	upstream	C1orf107	Verd	Verd
Av2 c2	293	52	11	Alu Y	chr8:117814020-117814312	30115	downstream	None	Verd	Verd
Ar3 c5	309	49	4	Alu Sx	chr11:130588789-130589097	1850	downstream	C11orf44	Verd	Verd
Au60	287	52	9	Alu Sx	chr16:31193325-31193611	INSIDE	intron 1	FUS	Verd	Blau
At1 c3	311	59	14	Alu Sp	chr19:5322610-5322920	INSIDE	inside	PTPRS	Blau	Blau
Ah1 c1	297	49	14	Alu Sg	chr18:7502820-7503116	-64997	upstream	None	Verd	Verd
Aq3 c6	89	64	7	Alu Y	chr8:2020262-2020350	INSIDE	intron 6	MYOM2	Verd	Verd
Ac4 c1	251	54	6	Alu Sx	chr16:28038121-28038371	INSIDE	intron 1	GSG1L	Verd	Verd
Ao1 c4	288	50	8	Alu Sx	chr19:53550139-53550426	INSIDE	intron 3	TMEM143	Verd	Verd
Ah2 c10	294	48	8	Alu Sc	chr7:3120272-3120565	36802	upstream	None	Blau	Verd
Ao2 c1	138	46	4	Alu Sg/x	chr11:130394314-130394451	50598	downstream	None	Verd	Verd
Ag2 c5	301	55	15	Alu Y	chr9:38693243-38693543	116282	upstream	None	Blau	Verd
Ar3 c1	275	51	10	Alu Sx	chr2:180202550-180202824	53578	upstream	None	Verd	Verd
Ah2 c8	563	50	11	Alu Y	chr13:64681523 - 64682085	-1093233	downstream	None	Verd	nd
Ao1 c0	345	45	2	Alu Sx	chr9:81586350-81586688	50598	downstream	None	Verd	nd
LV9	338	49	12	Alu Sx	chr19:3521200-3521537	-2743	upstream	HMG20B	Blau	nd

a) mida de l'element en parells de bases (pb), b) contingut en tant per cent de GC, c) el nombre de CpGs, d) localització cromosòmica, e) distància a l'inici de transcripció (TSS) o context genòmic dins un gen segons el Genome Browser versió NCBI 38/hg18 i el nom del gen, segons UCSC (University of California, Santa Cruz) i l'Ensembl (EMBL-EBI and Sanger Institute).

Confirmant els resultats de l'apartat 1 podem observar que dins dels 20 elements Alu desmetilats escollits hi ha un gran nombre d'Alus S pròxims a l'inici de transcripció (Taula R.2).

## 2.1. ESTUDI DE LA METILACIÓ DELS ELEMENTS ALU

### 2.1.1. ESTUDI DE LA METILACIÓ DELS ELEMENTS ALU DURANT LA TUMORIGÈNESI

Com ja s'ha comentat els elements Alu contenen el 33% del total de CpGs del genoma i estan altament metilats en la majoria de teixits somàtics (Schmid, 1991). No obstant, una fracció d'Alus roman desmetilada en cèl·lules normals i aquesta proporció augmenta en cèl·lules tumorals (Rodríguez et al., 2008b). Segons les nostres hipòtesis, les Alu desmetilades estarien situades en dominis de cromatina activa. Aquesta circumstància els hi conferiria capacitat per intervenir en la regulació i en l'estabilitat genòmica i afectar diferents processos importants per al desenvolupament, l'edat, la diversitat fenotípica i algunes malalties, entre elles el càncer.

#### 2.1.1.1. **Estudi de la metilació dels elements ALU en tumors primaris i línies de càncer de còlon**

Per tal de confirmar els canvis observats en els *fingerprints* obtinguts amb la tècnica d'AUMA es va analitzar l'estat de metilació de 17 dels 20 elements Alus seleccionats, ja que 3 van descartar-se per problemes tècnics.

Es va seqüenciar l'ADN de les mostres prèviament modificades amb bisulfit sòdic de 10 parelles de tumors colorectals i les seves mucoses normals. Com ja s'ha comentat el bisulfit sòdic converteix les citosines no metilades en uracils mentre que les metilades romanen com a tals (tècnica explicada en el material i mètodes apartat 2.3.1.), permetent l'anàlisi de CpGs basant-se en el seu estat de metilació i la caracterització de patrons específics de metilació.

En la taula R.3 només es mostren 3 de les 10 parelles normal-tumor analitzades, a fi de simplificar els resultats i donat que totes les parelles mostraven el mateix comportament. A més, la representació mostra l'estat global de metilació de l'element Alu. En l'*annex II figura All.1* trobem representat l'estat de metilació de cada dinucleòtid CpG dels elements Alu en base als anàlisi de seqüenciació després del tractament amb bisulfit sòdic.

Els resultats d'AUMA es validen en la majoria dels casos, tal i com s'observa en la darrera columna de la taula R.2, i revelen un grau total o parcial de desmetilació de les Alus en les mostres tumorals en vers a les normals. A més s'observa que els canvis de metilació no només afecten a la CpG de la diana Sma I sinó que també afecten a les CpGs del voltant tal i com es pot observar en l'*annex II figura All.1*. És important remarcar que la majoria d'Alus que s'hipometilen en tumor ho fan en tots els casos analitzats, indicant la recurrència d'aquesta alteració en el càncer colorectal.

Segons l'estat de metilació en el teixit normal i tumoral de còlon, les Alus es podrien classificar en diferents grups (Taula R.3). Per una banda, tres dels elements Alu analitzats (Ar3c7, Lv165 i Lv183) no mostren variació entre el teixit normal i el teixit tumoral, estant en ambdós casos totalment desmetilats. Els altres elements Alu seleccionats per l'estudi, majoritàriament, estan hipometilats en la mostra tumoral al comparar-la amb la seva parella normal. Dins d'aquestes Alus trobem una metilació

parcial en els teixits normals i una desmetilació parcial en el teixit tumoral, és el cas de Aj2c1, Ar3c5, Au104 i Av2c2. Altres casos (Ah1c1, Ao2c1, Ar3c1, Ac4c1, Ah2c10, Ag2c5, Ao1c4 i Aq3c6) els trobem totalment metilats en el teixit normal i parcialment o totalment desmetilats en el tumor. Pel que fa a Au60, s'observa un patró de metilació variable entre les diferents mostres normal-tumor. En el cas de At1c3 veiem que tant el còlon normal com tumoral mostren un alt grau de metilació i el patró és variable (segons la tècnica AUMA estava hipermetilada en tumor).

TAULA R.3. ESTAT DE METILACIÓ DELS ELEMENTS ALU SELECCIONATS EN PARELLES DE CÒLON NORMAL (N) / TUMOR (T)

	Ar3c7			Lv165			Lv183			Au60			Aj2c1			Ar3c5		
N	0	0	0	0	0	0	0	0	0	25	43	20	53	36	32	50	60	60
T	0	0	0	0	0	0	0	0	0	40	10	30	18	16	16	30	40	50

	Au104			Av2c2			At1c3			Ah1c1			Ao2c1			Ar3c1		
N	62	56	62	61	59	86	72	56	71	75	86	63	80	80	80	83	92	67
T	4	10	0	5	34	39	61	76	56	45	66	75	20	72	80	67	88	56

	Ac4c1			Ah2c10			Ag2c5			Ao1c4			Aq3c6		
N	89	93	93	97	97	91	100	85	87	100	100	96	100	100	100
T	21	43	54	16	66	66	73	20	85	5	6	5	77	0	41

○ 0-25% metilat ○ 25-50% metilat ● 50-75% metilat ● 75-100% metilat

Un cop analitzat l'estat de metilació dels elements Alu en parelles de teixit de còlon normal-tumoral, es van analitzar diferents línies cel·lulars d'adenocarcinoma de còlon humà (CCR): algunes d'elles a prop de la diploidia i amb inestabilitat de microsatèl·lits (HCT116 i LoVo), i un altre grup de línies aneuploides i amb inestabilitat cromosòmica (Sw480, HT29 i CaCo2).

En la taula R.4. representem l'estat global de metilació de l'element Alu en cadascuna de les línies de CCR d'acord amb les dades obtingudes per tractament amb bisulfit sòdic i la posterior seqüenciació. En l'annex II figura All.2 trobem representat l'estat de metilació de cada dinucleòtid CpG dels elements Alu.

Els resultats indiquen que no hi ha cap correlació dins de les línies de CCR en funció de l'estabilitat cromosòmica o l'estat de diploidia/aneuploidia.

En general podríem dir que hi ha 3 grups diferenciats: els elements Alu totalment desmetilats en totes les línies, els parcialment metilats i els altament metilats. Els elements Ar3c7, Lv165 i AU104 mantenen un estat desmetilat en totes les línies de CCR la qual cosa és interessant ja que dues d'aquestes Alus (Ar3c7 i LV165) també estan totalment desmetilades en les parelles normal-tumor prèviament analitzades.

En canvi Lv183, que estava totalment desmetilada en les parelles normal-tumor, veiem que es troba totalment metilada en 3 de les 6 línies de CCR. Pel que fa als elements Aj2c1, Av2c2 i Ar3c5, Ao1c4, Ah1c1, Ah2c10, Ac4c1, Aq3c6, Au60 i At3c1 veiem que

mostren un estat parcialment metilat. Finalment trobem un grup altament metilat en totes les línies (Ag2c5, Ar3c1 i Ao2c1).

Per altra banda, i tal com es troba ordenat en la taula R.4, la línia més metilada és HCT116, seguida de CaCo2, HT29, Sw480, LoVo.

Aquest perfil de metilació heterogeni en les diferents línies cel·lulars de CCR representa una bona eina per altres tipus d'estudi, especialment a nivell funcional.

TAULA R.4. ESTAT DE METILACIÓ DELS ELEMENTS ALU EN LÍNIES CEL·LULARS DE CARCINOMA DE CÒLON HUMÀ

	Ar3c7	Lv165	Au104	Aj2c1	Av2c2	Ar3c5	Ao1c4	Ah1c1	Ah2c10
HCT116	0	0	6	29	27	60	16	64	93
CaCo2	5	0	2	21	57	70	4	77	71
HT29	9	0	23	39	25	40	81	55	29
Sw480	0	0	8	63	59	75	0	63	96
LoVo	11	0	0	29	0	40	100	45	0

	Ac4c1	Lv183	Aq3c6	Au60	At1c3	Ag2c5	Ar3c1	Ao2c1
HCT116	68	100	100	100	99	98	100	95
CaCo2	100	100	100	100	99	93	90	100
HT29	36	100	36	15	71	85	92	100
Sw480	75	0	75	60	46	100	79	100
LoVo	15	0	0	100	88	63	77	100

○ 0-25% metilat ○ 25-50% metilat ● 50-75% metilat ● 75-100% metilat

Veiem doncs l'existència d'elements Alu totalment desmetilats en teixit normal i tumoral de còlon així com en línies de carcinoma de còlon, fet que contrasta amb que la majoria d'Alus en el genoma estan metilades, almenys en teixit sa, convertint-les en dianes d'interès pels nostres estudis.

### 2.1.1.2. Efecte de drogues epigenètiques en la metilació dels elements Alu

Per entendre millor el possible paper de la desmetilació dels elements Alu en la regulació gènica, es va tractar la línia HCT116 amb l'agent desmetilant 5-aza-2'-deoxicitidina (5-Aza-dC), amb l'inhibidor de deacetilases d'histones (HDACs) trichostatina A (TSA) i amb un combinat d'ambdós agents (Taula R.5) (tractament amb drogues epigenètiques explicat en detall en la secció de material i mètodes 2.1). Mitjançant la tècnica de tractament amb bisulfit sòdic i posterior seqüenciació es va analitzar l'estat de metilació de diverses Alus tant en la línia HCT116 parental com en la línia cel·lular tractada.

En primer lloc, per comprovar que el tractament havia funcionat es va mirar l'expressió dels gens *CPLX2* i *EN1*. Com ja està descrit en treballs previs (Rodríguez et al., 2008b) (Mayor et al., 2009), *CPLX2* i *EN1* estan silenciats a nivell transcripcional i la seva illa CpG està hipermetilada en les cèl·lules HCT116. Amb el tractament amb 5-Aza-dC i TSA es produeix tant una reactivació de l'expressió del gen com una desmetilació de l'illa CpG (Figura R.2).

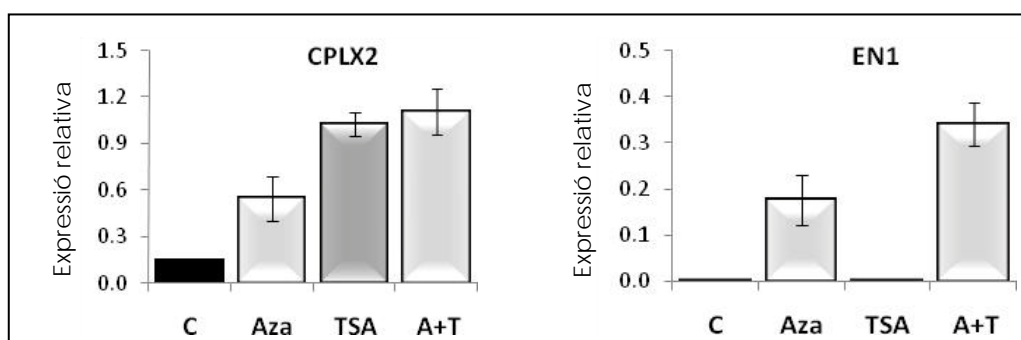


Fig. R.2. *EN1* i *CPLX2* utilitzats com a controls del correcte funcionament del tractament amb 5-Aza-dC, TSA i una combinació d'ambdós en la línia HCT116. El nivell d'expressió dels gens està normalitzat amb la mitjana de 5 estàndards. El color de les barres mostra el seu estat de metilació essent blanc 0-25% de metilació, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre 75-100% metilat.

Es va escollir la línia HCT116 ja que era la línia que contenia més elements Alu metilats (Taula R.4) i d'aquesta manera el tractament amb l'agent desmetilant 5-Aza-dC podria jugar un paper més rellevant. A part, també es va mirar una variant de la línia HCT116, HCT116-DKO la qual és deficient en les ADN metiltransferases DNMT1 i DNMT3b (Paz et al., 2003).

En aquest estudi no es van analitzar tots els elements Alus dels apartats anteriors sinó un subgrup representatiu. En el cas d'aquells elements Alu que en HCT116 control ja es troben desmetilats (Ar3c7, Lv165 i AU104), després del tractament continuen desmetilats (Taula R.5). En la resta dels elements Alu analitzats, parcial o totalment metilats sense tractar, no observem canvis en l'estat de metilació en tractar amb TSA, com es podria esperar ja que es tracta d'un inhibidor de deacetilases; en canvi, en

tractar amb 5-AzadC (agent desmetilant) i el combinat amb ambdós agents veiem una desmetilació parcial o total depenent dels casos. En les cèl·lules HCT116-DKO veiem una desmetilació en molts elements Alu, en la majoria de casos major que amb el tractament amb 5-AzadC, però també veiem que hi ha algunes Alus que són més resistents a la desmetilació. En l'annex II figura All.3 trobem representat l'estat de metilació de cada dinucleòtid CpG dels elements Alu, la línia DKO la trobem en l'annex II figura All.2.

TAULA R.5. ESTAT DE METILACIÓ DELS ELEMENTS ALU SELECCIONATS EN LA LÍNIA HCT116 i HCT116 TRACTADA AMB 5-AzadC, TSA i AZA+TSA I LA LÍNIA VARIANT DKO

	Ar3c7	Lv165	Au104	Aj2c1	Ar3c5	Ac4c1	Ah2c10	Ag2c5	Ar3c1	Au60	Lv183
HCT116	0	0	6	29	60	68	93	98	100	100	100
TSA	0	0	0	34	60	65	90	97	94	96	100
AZA	0	0	0	16	30	43	14	25	0	0	79
A+T	0	0	0	16	30	43	21	0	0	0	41
DKO	0	0	0	0	0	4	0	47	44	0	0

○ 0-25% metilat ○ 25-50% metilat ● 50-75% metilat ● 75-100% metilat

El fet que no totes les Alus metilades en HCT116 puguin desmetilar-se totalment al tractar-se amb agents desmetilants o en les HCT116-DKO suggereix la presència d'un o més factors diferencials que fa que alguns elements Alu siguin més o menys susceptibles a desmetilar-se.

El mateix es va fer amb la línia Sw480 i el comportament va ser similar. Les Alus desmetilades romanen desmetilades després dels tractaments mentre que les Alu parcial o totalment metilades es desmetilen al tractar amb 5-AzadC i amb el cotractament dels dos agents (dades no mostrades).

### 2.1.1.3. Estudi de la metilació de les regions flanquejants als elements Alu

Per esbrinar quin paper juguen els elements Alu desmetilats tant en cèl·lules normals com tumorals es van analitzar les regions flanquejants d'alguns elements Alu. Com ja s'ha comentat els elements Alu no estan distribuïts a l'atzar en el genoma humà sinó que sovint es troben en regions riques en gens (Lander et al., 2001), tendint-se a acumular en regions intròniques i intergèniques.

Per aquest estudi es van seleccionar sis elements Alu amb diferent estat de metilació en les diferents línies de CCR i propers a un inici de transcripció (Ar3c7, Lv165, Au104, Lv183, Aj2c1 i Au60) i les seves corresponents regions flanquejants pels extrems 3' i 5'. Les línies analitzades van ser HCT116, HCT116-DKO, LoVo, Sw480, HT29 i CaCo2 (Figura R.3). Els elements Alu Ar3c7, Lv165, Au104 i Au60 es troben dins de gens, en l'intró1, mentre que les Alu Aj2c1 i Lv183 es troben a menys de 500 pb de l'extrem 5' del gen

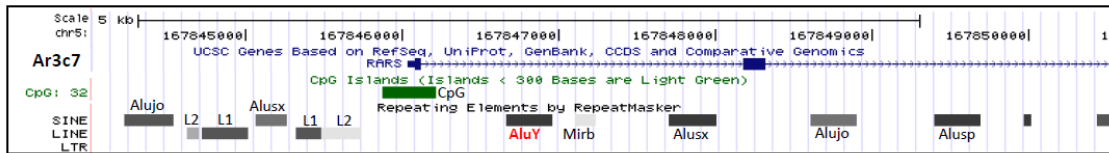
més proper, és a dir, en la regió promotora. A fi de simplificar la figura mostrem l'estat global de metilació de cadascun dels elements Alu i les seves regions flanquejants.

Observant l'estat de metilació de les regions flanquejants dels elements Alu Ar3c7, Lv165 i Au104 (desmetilats en totes les línies de CCR) veiem que hi ha un mateix patró en que l'Alu d'estudi es troba a l'extrem 3' d'una illa CpG situada a la zona promotora d'un gen. En tots aquests casos l'element Alu i l'illa CpG es troben completament desmetilats; en canvi, a mesura que ens allunyem d'aquesta, tant per l'extrem 3' com per l'extrem 5', els elements repetitius comencen a metilar-se. En el cas de l'element Lv165 no s'acaba de veure una metilació total en l'extrem 3', segurament degut al fet que molt a prop es troba la regió promotora d'un altre gen, *RPS2*.

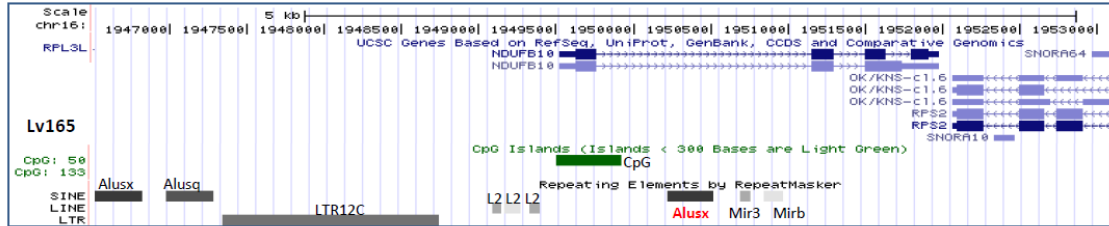
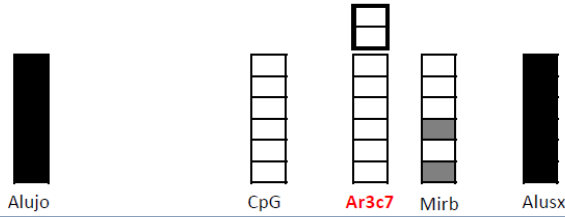
En el cas de Lv183 veiem que l'element Alu està situat al costat de l'illa CpG i, no obstant, presenta metilació en alguna línia de CCR tot i que s'havia trobat totalment desmetilada en les parelles de mucosa de còlon normal i tumoral. Aquest fet podria explicar-se pel cultiu *in vitro* que com s'ha descrit pot afectar a la metilació de l'ADN tendint a hipermetilar-se. L'extrem 5' de la regió no s'ha pogut determinar degut a la manca de CpGs de la regió i a la impossibilitat de dissenyar encebadors específics ja que tractem amb seqüències repetitives.

Pel que fa als elements total o parcialment metilats en línies de CCR (Aj2c1 i Au60). Aj2c1 mostra el mateix comportament que els altres casos i en Au60 veiem que l'element Alu que es troba entre l'illa CpG i l'Alu d'estudi, està totalment desmetilada. La resta d'elements repetitius analitzats, tant per l'extrem 3' com per l'extrem 5' mostren un estat metilat, mantenint doncs el comportament anterior.

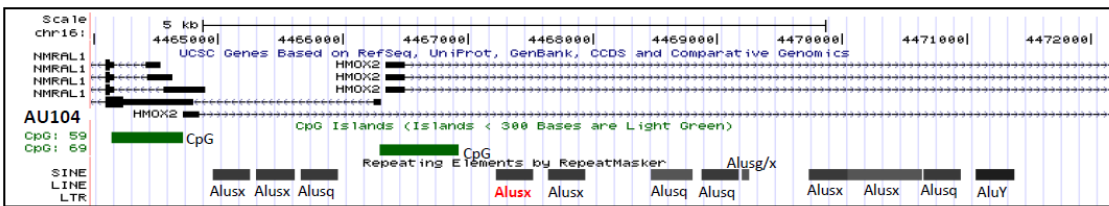
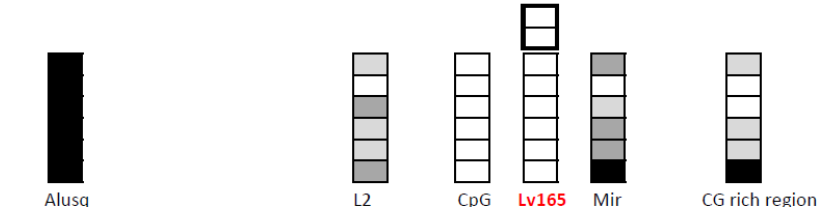
Així doncs, en totes les regions analitzades properes a un inici de transcripció (excepte la corresponent a l'Alu Lv183) hem trobat que l'Alu immediatament *downstream* de l'illa CpG i dins d'un intró està completament desmetilada en totes les línies de CCR, a l'igual que l'illa CpG, mentre que la resta de la regió està metilada, suggerint el possible paper de l'Alu com a barrera protectora de la metilació. A més a més, en alguns casos també hi ha una Alu diferencialment metilada segons la línia cel·lular.



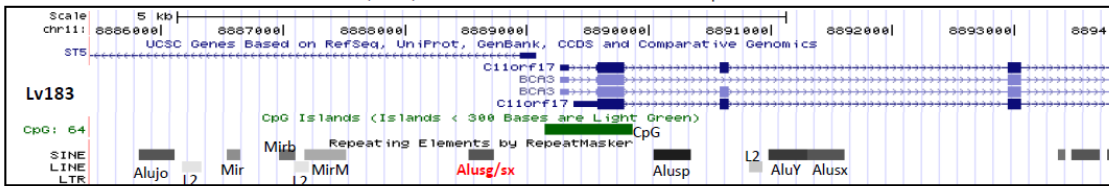
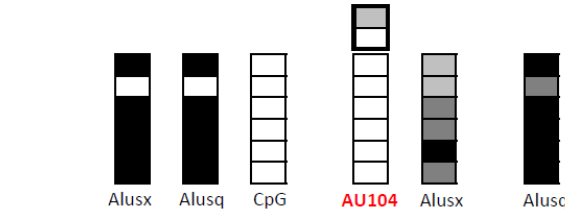
NORMAL  
TUMOR  
HCT116  
HCT116-DKO  
LoVo  
SW480  
HT29  
CaCo2



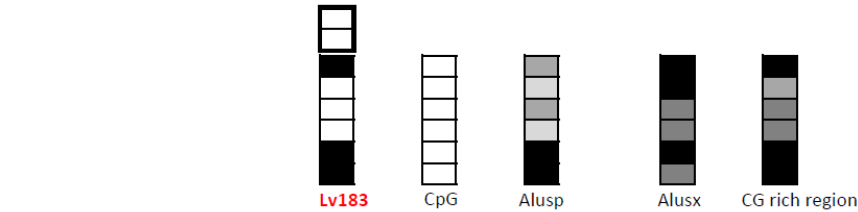
NORMAL  
TUMOR  
HCT116  
HCT116-DKO  
LoVo  
SW480  
HT29  
CaCo2



NORMAL  
TUMOR  
HCT116  
HCT116-DKO  
LoVo  
SW480  
HT29  
CaCo2



NORMAL  
TUMOR  
HCT116  
HCT116-DKO  
LoVo  
SW480  
HT29  
CaCo2





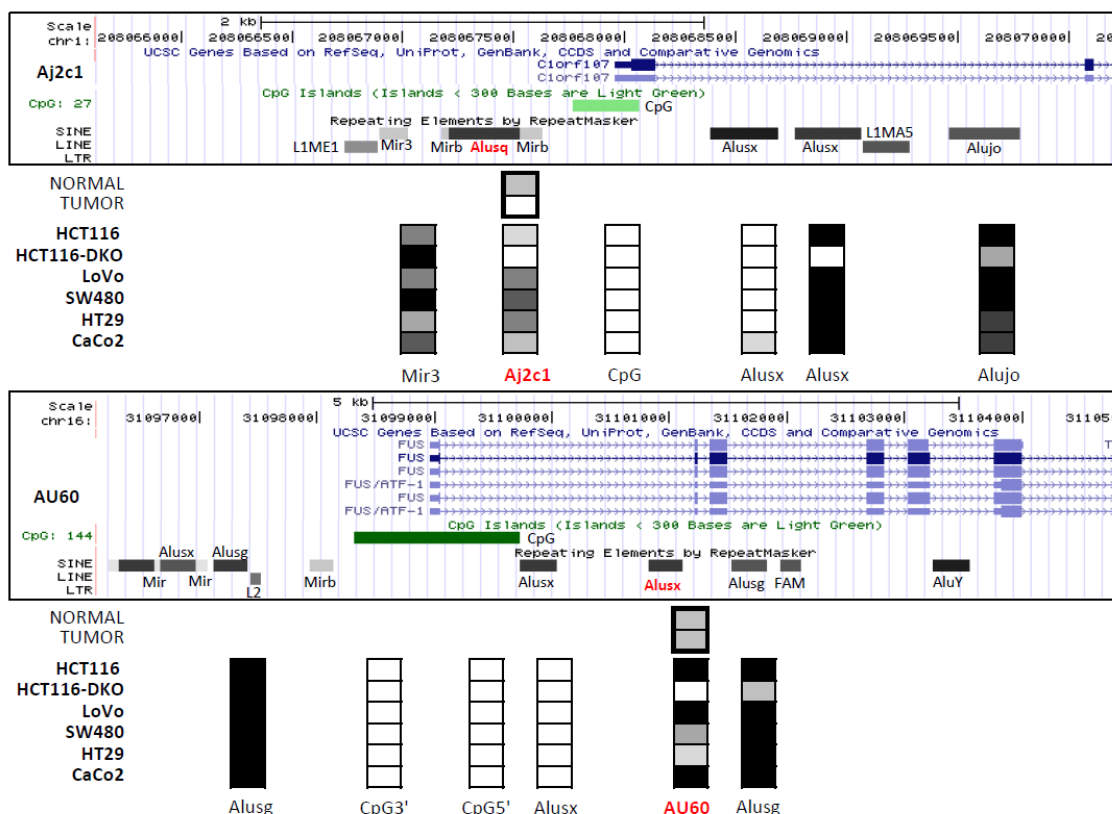


Fig. R.3. Estat de metilació de les regions flanquejants als elements Alu seleccionats (en vermell) en diferents línies de CCR. Regió obtinguda del Genome Browser versió Mar. 2006 (NCBI36/hg18). De cada regió es mostra el gen, les illes CpG i els elements repetitius. En blanc de 0-25% de metilació, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre 75-100% de metilació.

## 2.1.2. ESTUDI DE LA METILACIÓ DELS ELEMENTS ALU DURANT LA DIFERENCIACIÓ CEL·LULAR

La diferenciació cel·lular consisteix en la reprogramació gènica d'una cèl·lula mare, i suposa l'activació i repressió de milers de gens en un procés seqüencial, fins donar una cèl·lula plenament diferenciada. En aquesta regulació juga un paper important l'epigenètica, tant a nivell de metilació de l'ADN com a nivell de cromatina. La majoria d'estudis en aquest camp s'han centrat en regions genòmiques de seqüència única (Berdasco and Esteller, 2011; Hewitt et al., 2011).

Per esbrinar si l'estat de metilació dels elements Alu és específic de teixit i juga algun paper en la diferenciació cel·lular vam voler ampliar l'estudi a altres tipus cel·lulars (Taula R.6). Per això es va analitzar l'estat de metilació d'aquells elements Alu desmetilats en teixit normal i tumoral de còlon (Ar3c7, Lv165, Lv183 i Au104), els elements que es troben parcialment metilats depenent de les línies de CCR estudiades (Aj2c1 i Au60) i els elements altament metilats (Ag2c5 i Ao2c1).

Les línies que es van utilitzar van ser cèl·lules mare embrionàries humanes (hESC2, hESC4, hESC6), cèl·lules pluripotents induïdes originàries de queratinòcits (KIPS4F1,

KiPS4F8, [H]KiPS4F1), queratinòcits (K3, K18, K137, MMTA) i fibroblasts (F, HFF), mostres de sang (leucòcits) i tres parelles de teixit de còlon normal. En l'annex II figura All.4 trobem representat l'estat de metilació de cada dinucleòtid CpG dels elements Alu que es mostren en la taula R.6.

Els elements Ar3c7, Lv165 i Lv183 mantenen el seu patró de desmetilació en tots els tipus cel·lulars, la qual cosa és molt interessant i ens fa pensar en la importància biològica d'aquests elements que també es trobaven totalment desmetilats en les parelles normal-tumor.

Au60 presenta un estat de baixa o nul·la metilació en les línies de queratinòcits i fibroblasts, mentre que s'observa un increment en l'estat de la metilació en les cèl·lules pluripotents induïdes i aquest increment es fa més palès en les cèl·lules mare embrionàries suggerint un possible rol en la diferenciació.

TAULA R.6. ESTAT DE METILACIÓ DELS ELEMENTS ALU SELECCIONATS EN DIFERENTS LÍNIES CEL·LULARS

		Ar3c7	Lv165	Lv183	Au60	Au104	Aj2c1	Ag2c5	Ao2c1
CÈL·LULES MARE EMBRIONÀRIES	hES4	0	0	0	65	96	100	85	100
	hES2	0	0	0	38	96	100	85	100
	hES6	0	8	0	68	96	100	85	100
CÈL·LULES PLURIPOTENTS INDUÏDES	[H]KiPS4F1	0	13	0	48	98	99	95	100
	KiPS4F8	0	13	0	38	94	95	95	100
	KiPS4F1	0	0	0	33	98	87	88	100
QUERATINOCITS	K3	0	0	0	8	92	96	100	100
	K18	0	3	0	0	85	82	95	100
	K137	0	0	0	15	98	67	95	100
	MMTA	0	3	0	15	85	96	100	100
FIBROBLASTS	F	0	38	0	15	65	100	100	100
	HFF	0	23	0	15	58	100	93	100
LEUCÒCITS		0	10	0	0	68	75	97	95
TEIXIT CÒLON NORMAL		0	0	0	25	62	53	100	80
		0	0	0	46	56	36	85	80
		0	0	0	20	62	32	87	80

○ 0-25% metilat   ○ 25-50% metilat   ● 50-75% metilat   ● 75-100% metilat

L'element AU104 mostra un alt grau de metilació, molt major del que observàvem en la mucosa normal de còlon (apartat 2.1.1.1); a més s'observa que les cèl·lules mare i les cèl·lules pluripotents induïdes presenten més metilació que els fibroblasts i leucòcits, suggerint de nou una possible implicació a nivell de diferenciació.

L'Alu Aj2c1 presenta un alt grau de metilació en totes les línies estudiades tot i estar parcialment metilat en les línies de CCR i en la mucosa de còlon normal i tumoral. I finalment, Ag2c5 i Ao2c1 manté un alt grau de metilació en totes les línies, tal i com passava amb les línies de CCR i en les parelles normal-tumor.

Veiem doncs que hi ha certs elements Alu que mostren un patró de metilació específic de tipus cel·lular. Així doncs trobem alguns totalment desmetilats en totes les línies, altres totalment metilats i altres que es desmetilen o es metilen en el procés de diferenciació.

### 2.1.3. CONSERVACIÓ EVOLUTIVA DE LA METILACIÓ DELS ELEMENTS ALU EN PRIMATS

L'estudi amb genomes de mamífers suggereix que aproximadament el 5% del genoma humà s'ha conservat evolutivament en els últims 200 milions d'anys, la qual cosa inclou la gran majoria dels gens i seqüències reguladores. No obstant, els gens i les seqüències reguladores actualment conegudes suposen només el 2% del genoma, la qual cosa suggereix que la major part de la seqüència genòmica amb gran importància funcional és desconeguda. Un percentatge important dels gens humans presenta un alt grau de conservació evolutiva (Batzer and Deininger, 2002; Price et al., 2004).

Com ja s'ha comentat en la introducció la primera aparició de les seqüències Alu es va donar fa aproximadament 65 milions d'anys, coincidint amb l'origen i expansió dels primats. El seu estudi és útil per investigar la diversitat del patrimoni genètic de les poblacions humanes i l'evolució dels primats. La majoria de les insercions d'elements Alu són recents i per tant, sovint polimòrfiques, agafant diferents valors de freqüència depenent de la població. Es consideren esdeveniments mutacionals únics ja que és bastant improbable trobar la mateixa inserció Alu en el mateix lloc, per aquest motiu són tant interessants per l'estudi de l'evolució humana. Es tracta d'insercions estables del genoma (Kazazian, 2004). Però tot i haver molts estudis sobre la conservació i l'evolució genètiques dels elements Alu, no n'hi han dades sobre la seva conservació i l'evolució a nivell epigenètic.

En la taula R.7 es mostra l'estat de conservació de varis dels elements Alu d'estudi en diferents espècies de primats: *Homo sapiens* (humà), *Pan troglodytes* (ximpanzé), *Gorilla gorilla* (goril·la), *Pongo pygmeus* (orangutan) i *Macaca Mulata* (macaco). Per cada element Alu es presenta l'estat de metilació en mostres de sang humana i la família a la que pertany. Els paràmetres mostrats en la taula R.7 són el nombre de *hits*, és a dir, el nombre de còpies que apareixen en el genoma de cada espècie, el cromosoma al que pertany la seqüència i l'orientació del fragment i, finalment, el percentatge d'identitat respecte a la seqüència humana.

L'estudi es va fer *in silico*, utilitzant com a genoma de referència el genoma de l'*Homo sapiens* (Homo sapiens (GRCh37/hg19 versió del Febrer 2009)) i alineant-lo amb *Pan troglodytes* (Chimp (Oct. 2010 (CGSC 2.1.3/panTro3)), *Gorilla gorilla* (Gorilla (May 2011 (gorGor3.1/gorGor3))), *Pongo pygmeus abelii* (Orangutan (July 2007 (WUGSC 2.0.2/ponAbe2))), i *Macaca Mulatta* (Rhesus (Oct. 2010 (BGI CR\_1.0/rheMac3))).

Per cercar el tant per cent d'identitat, les seqüències de les diferents espècies de primats es van alinear amb la seqüència humana de referència fent servir alineaments locals (amb l'algoritme de Smith-Waterman) i globals (amb l'algoritme de Needleman-Wunsch) donant el mateix resultat en ambdós casos.

Per trobar el nombre de *hits* o còpies en el genoma es va agafar la seqüència de l'element Alu més 20pb en l'extrem 5' i 20 bases en l'extrem 3' i mitjançant un BLAT es va buscar si hi havia més còpies de la regió en el genoma de cada espècie.

TAULA R.7. ESTAT DE CONSERVACIÓ DELS ELEMENTS ALU EN DIFERENTS ESPÈCIES DE PRIMATS

		Element Alu																
		Ar3 c1	Ar3 c7	Lv165	Lv183	Au104	Aj2 c1	Av2 c2	Ar3 c5	Ah1 c1	At1 c3	Ac4 c1	Ah2c10	Au60	Ao2 c1	Ao1 c4	Ag2 c5	Aq3 c6
Espècie	Família	Alu Sx	Alu Y	Alu Sx	Alu Sg/x	Alu Sx	Alu Sq	Alu Y	Alu Sx	Alu Sg	Alu Sp	Alu Sx	Alu Sc	Alu Sx	Alu Sg/x	Alu Sx	Alu Y	Alu Y
	hits <sup>a</sup>	Humà	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ximpanzé		1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1
Goril·la		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Orangutan		1	1	2	1	1	1	2	1	1	1	1	1	1	2	1	3	1
Macaco		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
cromosoma <sup>b</sup> i orientació 5'-3' <sup>c</sup>	Humà	2 -	5 +	16 -	11 -	16 +	1 -	8 -	11 -	18 +	19 +	16 +	7 +	16 -	11 +	19 +	9 -	8 +
	Ximpanzé	2B -	5 +	16 -	11 -	16 +	1 -	8 -	11 -	18 -	19 +	16 +	7 +	16 -	11 +	19 +	9 -	8 +
	Goril·la	2B -	5 +	16 -	11 -	16 +	1 -	8 -	11 -	18 -	19 +	16 +	7 +	16 -	11 +	19 +	9 -	8 +
	Orangutan	2B -	5 +	16 -	11 +	16 +	1 +	8 -	11 -	18 -	19 +	16 +	7 +	16 -	11 +	19 +	9 +	8 +
	Macaco	12 -	1	20 -	8 -	20 +	1 +	1	14 -	18 -	19 +	20 +	3 +	20 -	14 +	19 +	1	1
% identitat <sup>d</sup>	Humà	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	Ximpanzé	99	94	98	93	98	98	95	95	94	97	100	96	98	99	98	93	73
	Goril·la	98	92	84	96	99	96	96	95	87	57	99	96	97	45	87	93	87
	Orangutan	94	88	91	96	94	95	92	92	91	89	98	94	91	94	89	91	89
	Macaco	88	88	90	93	72	83	84	84	87	87	88	85	45	91	85	85	85

El nom dels elements Alu es mostra en escala de grisos representant l'estat de metilació en sangs humanes (negre 100-75% metilat, gris fosc 75-50% metilat, gris 50-25% i blanc 25-0%).

<sup>a</sup> Nombre de cops que trobem la seqüència Alu més 20pb en l'extrem 5' i 20pb més en l'extrem 3' en el genoma. Destaquem amb un marró mig quan trobem més d'una còpia i en un marró fosc la manca de l'element.

<sup>b</sup> Cromosoma on es troba localitzada la seqüència. <sup>c</sup> Orientació 5' a 3' en + i Orientació 3' a 5' en -. Es ressalta en vermell l'orientació diferent respecte a la seqüència en humà. Destaquem amb un verd mig quan trobem un canvi de cromosoma i en un verd fosc la manca de l'element.

<sup>d</sup> Tant per cent d'identitat de la seqüència en les diferents espècies de primats. Destaquem amb taronges mitjos la pèrdua d'identitat i en un taronja fosc la manca de l'element.

Mirant els *hits* veiem que, majoritàriament, es troba una única còpia de la regió genòmica que inclou l'element Alu d'estudi, excepte Lv183 en Ximpanzé i Lv165, Av2c2, Ao2c1 i Ag2c5 en Orangutan, els quals presenten 2 o 3 còpies en el genoma. Cal destacar que aquestes múltiples còpies es troben en regions no mapades, pel que és podria tractar d'errors de l'assemblatge. En el cas de l'element Lv165 es troba una segona còpia dins el mateix cromosoma 16 d'orangutan.

Si ens fixem en el cromosoma on es troba situada cada Alu veiem que la localització es manté en totes les espècies respecte al genoma humà de referència. És interessant remarcar que el cromosoma 2 en humà es va originar de la fusió entre dos cromosomes presents en la resta de primats (d'aquí la terminologia 2B en els primats). Per altra banda, el macaco, que pertany a la família dels *Cercopithecidae* a diferència del humà, ximpanzé, orangutan i goril·la, presenta una distribució cromosòmica diferent a la resta de primats estudiats degut a reordenaments cromosòmics complexos i per tant no es mantenen les equivalències en la nomenclatura dels cromosomes (Gibbs et al., 2007).

El canvi d'orientació en alguns elements Alu segons l'espècie troba la seva explicació en el fet de senzills reordenaments per inversió, és el cas de Lv183, Aj2c1 i Ag2c5.

Fixant-nos amb el tant per cent d'identitat (Figura R.4) veiem que el tant per cent és major en ximpanzé i que a mesura que ens allunyem de l'escala evolutiva la identitat es redueix, sent en macaco on trobem menys tant per cent d'identitat i més *gaps*, entenent com a *gap* aquelles regions que no contenen un element Alu, és el cas de Ah1c1 i Aq3c6 en goril·la, de Aq3c6 en orangutan i de Ar3c7, Av2c2, Ag2c5, Aq3c6 en macaco (Taula R.7). El genoma del goril·la presenta alguns tants per cent d'identitat més baixos del que esperariem donada l'escala evolutiva, però donada la baixa qualitat del mapa genòmic del goril·la aquesta dada s'ha de prendre amb precaució.

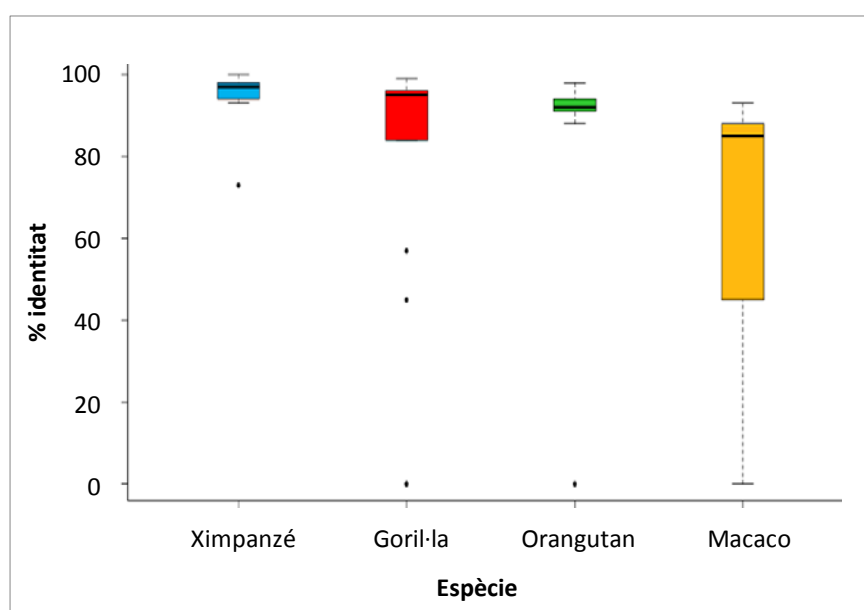


Fig. R.4. Comparació entre l'espècie i l'estat de conservació dels elements Alu respecte a l'humà.

Mirant per famílies (Figura R.5) no veiem una diferència significativa al comparar estat de conservació i família, recordem que dels 17 elements Alu analitzats cap pertany a la família J. Cal destacar que dels 4 elements Alu Y (la família més jove) analitzats cap es troba en Macaco, els 13 elements restants pertanyen a la família Alu S.

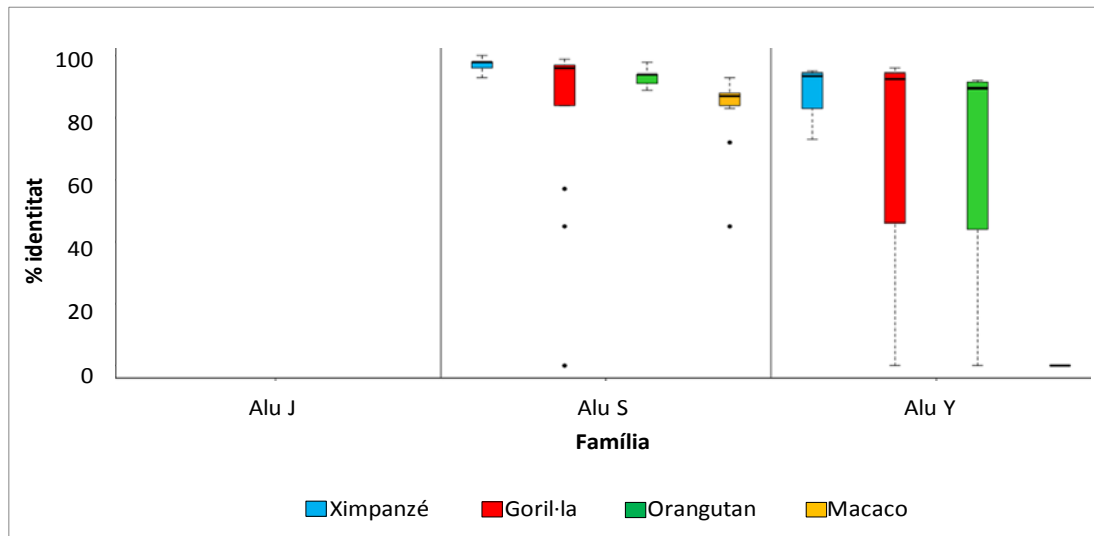


Fig. R.5. Comparació entre la família i l'estat de conservació dels elements Alu.

Pel que fa a l'estat de metilació veiem certa tendència a que els elements Alu desmetilats presenten un major grau de conservació de seqüència que els elements més metilats (Figura R.6). Caldria un major nombre d'elements per veure una tendència més clara.

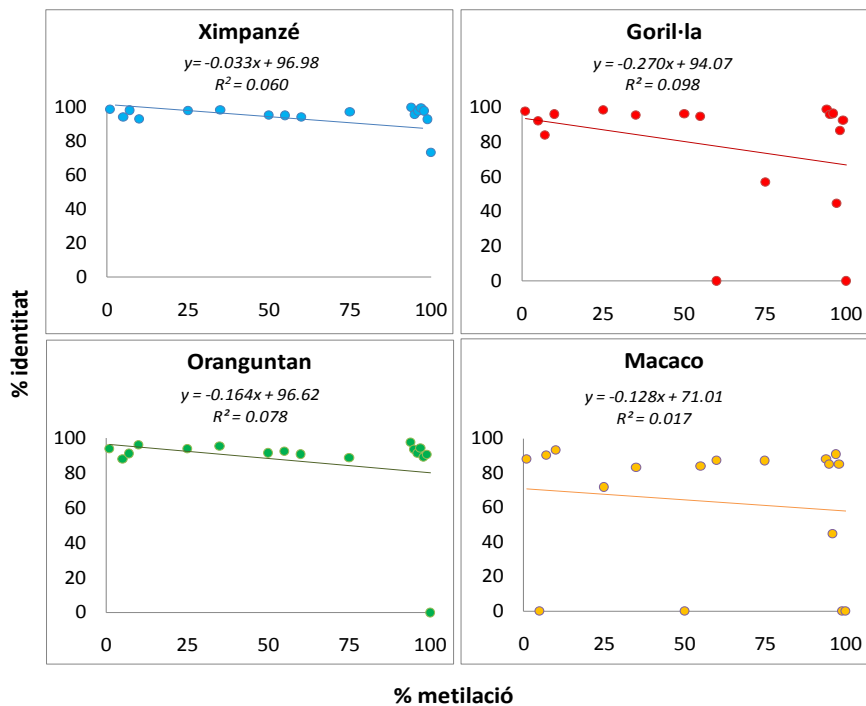
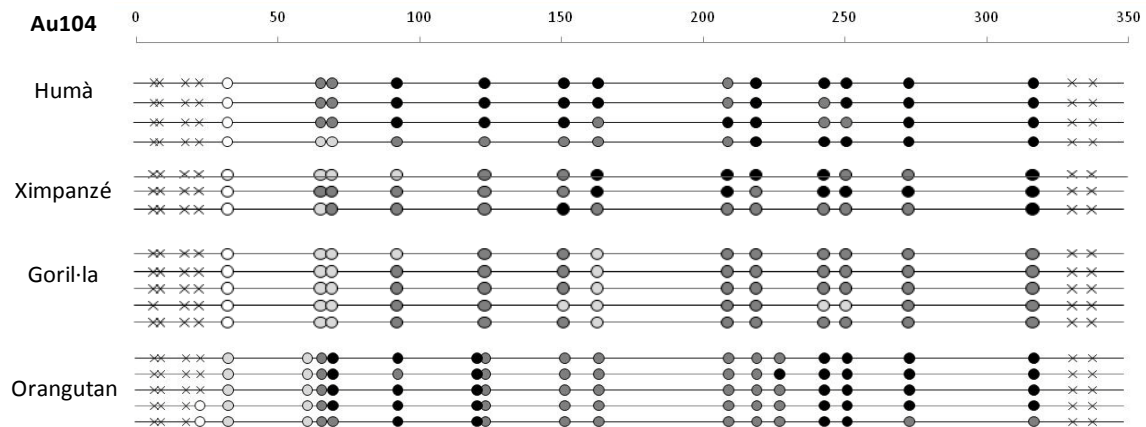
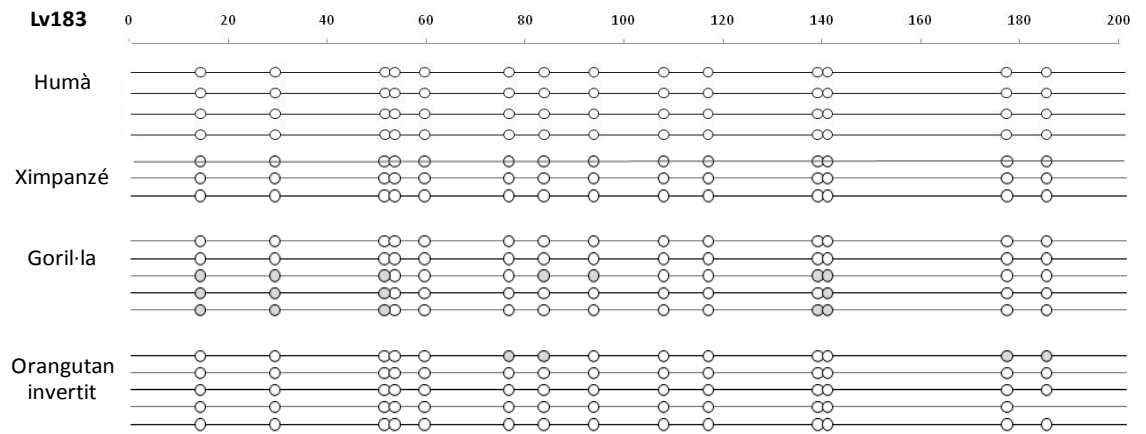
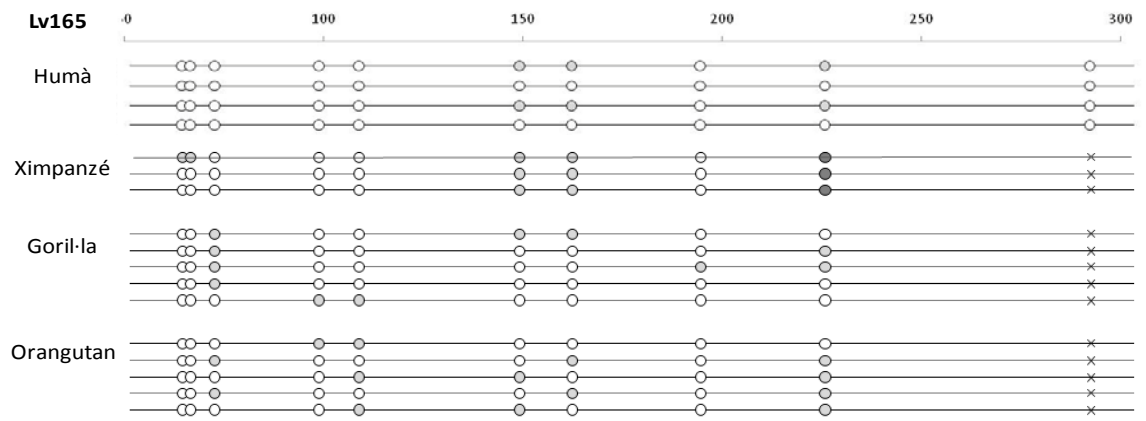
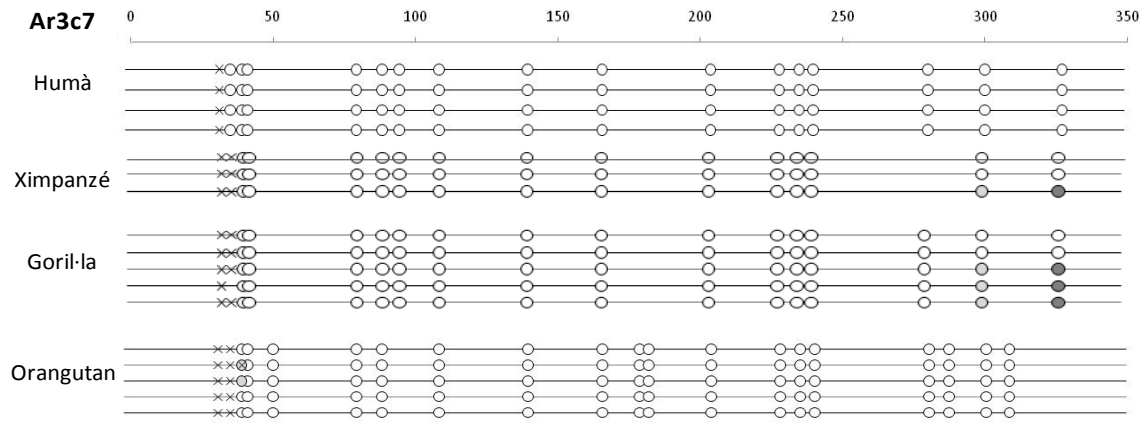


Fig. R.6. Comparació entre l'estat de metilació i l'estat de conservació dels elements Alu en ximpanzé, goril·la, orangutan i macaco. Grafiquem la línia de tendència així com la seva R i l'equació del gràfic.

Per veure l'estat de conservació de l'element Alu tant a nivell de seqüència (sobretot de les CpG) com epigenètic es van seleccionar alguns elements Alu i es va mirar el seu estat de metilació en 4 mostres de sang humana, 3 sangs de ximpanzé, 5 sangs de goril·la i 5 mostres de sang d'orangutan. Les seqüències seleccionades van ser tres elements Alu totalment desmetilats en sangs humanes (Ar3c7, Lv165 i Lv183), dos elements Alu parcialment metilats en sangs humanes (Au104 i Aj2c1) i un element Alu totalment metilat en sangs humanes (Aq3c6). En aquest cas representarem la metilació dels dinucleòtids CpG de cada element Alu estudiat per poder veure el seu grau de conservació (Figura R.7).

Podem observar que l'estat de metilació no varia, o si més no els canvis no són significatius, és a dir, les Alus desmetilades en humans es mantenen desmetilades en tots els primats mentre que les metilades en humans estan metilades en primats. Per altra banda, no acabem de veure una clara tendència per poder associar un major grau de d'identitat als elements Alu més desmetilats. Tot i tractar-se de pocs casos, semblaria que aquells elements que estan desmetilats, i probablement actius, estan més conservats en les diferents espècies de primats indicant doncs la seva importància biològica i els elements metilats, i probablement inactius, mostren més variació. No obstant caldria confirmar aquesta tendència analitzant més elements Alu.





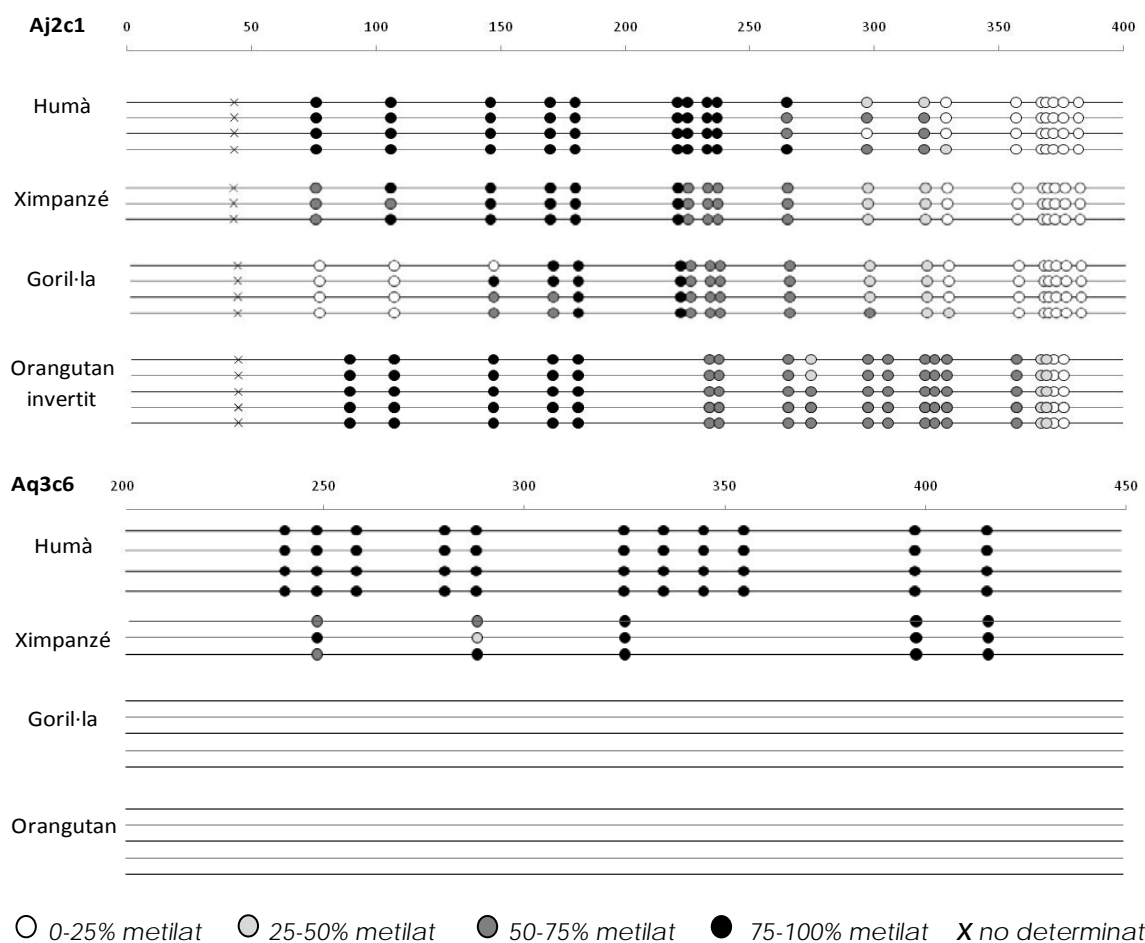


Fig. R.7. Estat de metilació del elements Alu seleccionats per l'estudi en mostres de sangs de diferents espècies de primats: humà, ximpanzé, goril·la i orangutan.

## 2.2. ESTUDI DE L'ESTRUCTURA DE LA CROMATINA ASSOCIADA ALS ELEMENTS ALU

Tal i com s'ha descrit en la introducció la metilació i acetilació d'histones, en particular de l'H3, actuen sobre la maquinària transcripcional. Mentre les marques de di/trimetilació de la lisina 4 estan associades a una transcripció activa, la di/trimetilació de la lisina 9 i la trimetilació de la lisina 27 estan associades a un estat de repressió (Felsenfeld and Groudine, 2003). La metilació de la lisina 4 s'associa a l'activació transcripcional i es troba en regions eucromàtiques, sobre gens actius o activables (Kouzarides, 2007). H3K4me3 sol trobar-se en regions promotores. Per altra banda l'acetilació de la lisina 9 i de la lisina 14 en la histona H3 són necessàries pel reconeixement de l'inici de transcripció (Kouzarides, 2007; Martin et al., 2005).

Veient els diferents estats de metilació dels elements Alu ens vam plantejar si aquests es podien relacionar amb l'estat de la cromatina. Per fer l'estudi es va utilitzar la tècnica de ChIP (Immunoprecipitació de Cromatina) (tècnica explicada en detall en la secció de material i mètodes apartat 2.7) amb diferents línies cel·lulars de CCR.

La immunoprecipitació es va fer amb anticossos contra tres marques d'activació, l'acetilació de la lisina 14 en la histona 3 (AcH3K14), l'acetilació de la lisina 9 en la

histona 3 (AcH3K9) i la trimetilació de la lisina 4 en la histona 3 (H3K4me3). Com a marques d'inactivació es van estudiar la trimetilació de la lisina 9 en la histona 3 (H3K9me3) i la trimetilació de la lisina 27 en la histona 3 (H3K27me3).

Es van utilitzar sèrum normal de conill (IgG) com a control negatiu de la immunoprecipitació per confirmar que la hibridació dels anticossos emprats era específica i com a control positiu es va utilitzar la histona 3 total (H3), ja que totes les marques analitzades pertanyen a la histona H3. Per altra banda, com a controls de la PCR, i per confirmar que la immunoprecipitació va ser correcte, vam utilitzar *GAPDH* (*Gliceraldehid-3-fosfat deshidrogenasa*) i *16CEN* (*regió centromèrica del cromosoma 16 humà*) com a controls positiu i negatiu, respectivament, de marca activa (AcH3K9 i H3K4me3). També es van utilitzar *GAPDH* com a control negatiu de la marca inactiva H3K9me3 i *16CEN* com a control positiu. A més, es va utilitzar *CPLX2* (*Complexina 2*) i *DRD1* (*Receptor de la dopamina D1*) com a controls positius de la marca inactiva H3K27me3 i *THOC3* (*Complexe 3 de THO*) com a control negatiu. Finalment es va quantificar el nombre total d'Alus immunoprecipitades (*qALU*) (Rodríguez et al., 2008a).

A l'hora de representar els gràfics de l'estat de cromatina s'ha utilitzat un mateix sistema. Així les gràfiques de marques actives i inactives són ràtios ja que estan normalitzades amb l'H3 total i les gràfiques d'H3 representen els valors absoluts (normalitzats per l'input) amb els quals hem normalitzat les diferents marques. Els gràfics de l'IgG no estan representats ja que els nivells estan per sota del límit de detecció, confirmant l'especificitat dels anticossos.

Les fórmules amb les quals s'han obtingut tant els ràtios, valors absoluts així com les desviacions estàndards i les variàncies estan explicades amb detall en el material i mètodes apartat 2.7.

## 2.2.1. ESTUDI DE LA CROMATINA ASSOCIADA ALS ELEMENTS ALU DURANT LA TUMORIGÈNESI

### 2.2.1.1. Estudi de la cromatina associada als elements Alu en línies de càncer de còlon

Es van analitzar en la línia HCT116 les diferents marques d'histona associades a 12 de les Alus seleccionades per l'estudi. La tria d'aquestes es va fer amb la intenció de representar els principals perfils de metilació: totalment metilades (representades en negre), parcialment metilades (representades en gris) i totalment desmetilades (representades en blanc) (Figura R.8).

Veiem que els elements Alu totalment desmetilats (Ar3c7, Lv165 i Au104) o amb nivells molt baixos de metilació (Aj2c1) presenten molta marca de AcH3K9 i H3K4me3 i en front dels elements Alu que es troben totalment metilats (Ao2c1, Ag2c5, Lv183, Aq3c6 i Au60) o amb nivells molt alts de metilació (Ar3c5, Ac4c1 i Ah2c10). En canvi, l'AcH3K14 es troba present en tots els elements Alu malgrat ser també una marca d'activació, la qual cosa indica que les marques actives H3K4me3 i AcH3K9 s'associen específicament a les Alus amb un alt grau de desmetilació.

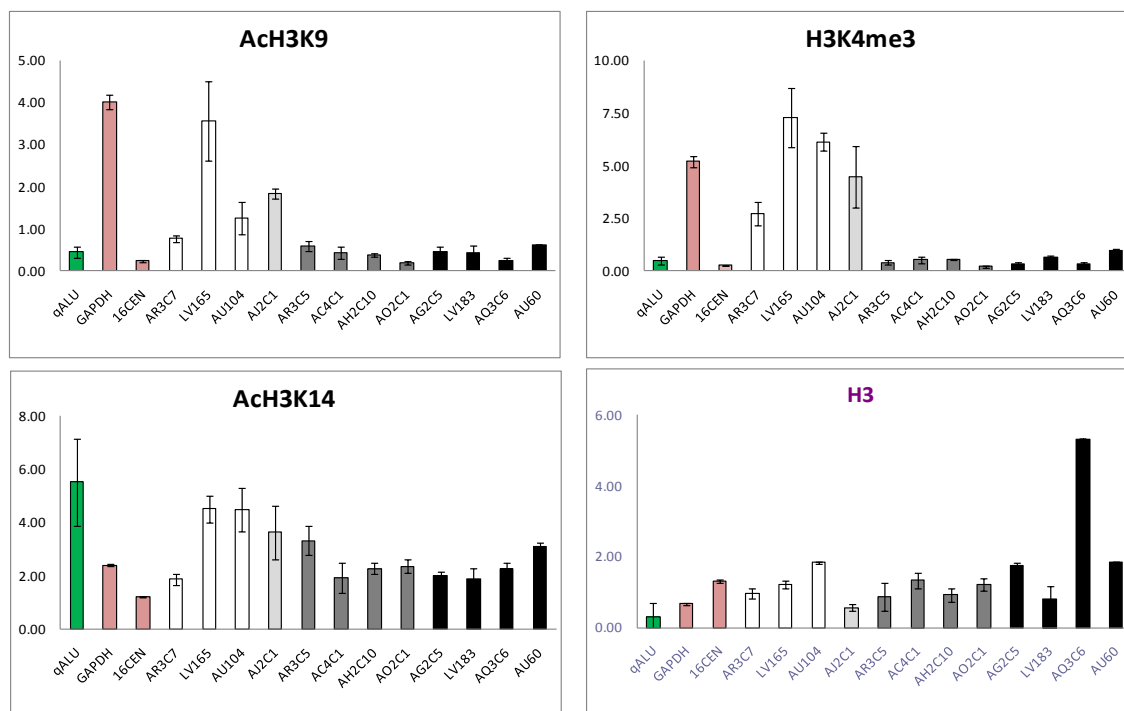


Fig. R.8. Anàlisi per ChIP dels elements Alu en la línia HCT116. Com a marques d'activació es troben representades ACh3K14, ACh3K9, H3K4me3. L'H3 total és el control. L'eix de les Y representa el ràtio de la marca d'histona analitzada respecte l'H3, a excepció d'H3 que està normalitzat respecte l'input. Cada columna representa un element Alu i el color mostra el seu estat de metilació essent blanc 0-25%, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre 75-100%. Per altra banda en vermell es representen els controls del ChIP, essent GAPDH control positiu de marques actives mentre que 16CEN és negatiu per marques actives. Per últim, la columna en verd representa ChIP-qALU, és a dir, el total d'elements Alu en la cromatina immunoprecipitada.

Tot i que s'han analitzat també marques d'inactivació (H3K27me3 i H3K9me3) no es van trobar presents en cap de les Alus analitzades, ni les metilades ni les desmetilades (dades no mostrades).

El global d'Alus (determinat mitjançant la PCR de ChIP-qAlu) presenta molta marca de ACh3K14, de forma similar a les Alus individuals analitzades, tant desmetilades com metilades, confirmant l'enriquiment d'aquesta marca observat en el grup d'Alus seleccionades. En canvi el ChIP-qAlu de marques actives (ACh3K9 i H3K4me3) va resultar en nivells molt baixos, el que s'interpreta com que els elements Alu desmetilats representen una petita fracció de totes les Alus.

Continuant amb l'estudi es va comparar l'estat de la cromatina dels 6 elements Alu (Lv165, Au104, Aj2c1, Lv183, Ao2c1 i Au60) que variaven més en l'estat de metilació entre les diferents línies de CCR (HCT116, Sw480 i LoVo) (Figura R.9).

Igual que en l'estudi anterior, els elements Alu desmetilats presenten nivells més elevats de les marques actives Ach3K9 i H3K4me3 que els elements metilats i les diferències més clares es donen en H3K4me3.

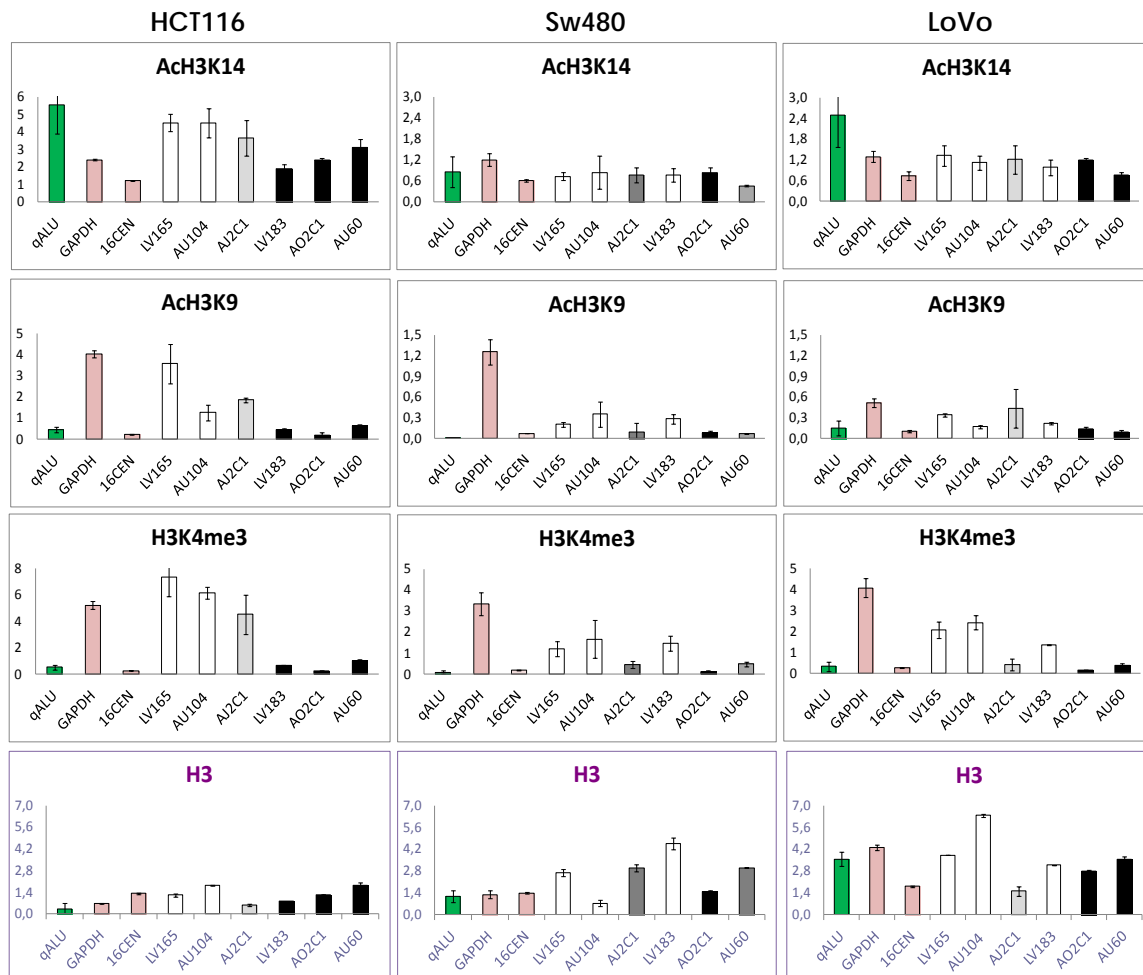


Fig. R.9. Anàlisi per CHIP dels elements Alu en les línies HCT116, Sw480 i LoVo. Com a marques d'activació es troben representades Ach3K14, Ach3K9, H3K4me3. L'H3 total és el control. L'eix de les Y representa el ràtio de la marca respecte l'H3, a excepció d'H3 que és el valor absolut. Cada columna representa un element Alu i el color mostra el seu estat de metilació essent blanc desmetilat, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre metilat. Per altra banda en vermell es representen els controls del CHIP, essent GAPDH control positiu de marques actives mentre que 16CEN és el negatiu per marques actives. Per últim, la columna en verd representa CHIP-qALU, és a dir, el total d'elements Alu en la cromatina immunoprecipitada.

Finalment es van analitzar les marques d'histona a la línia HCT116-DKO (Figura R.10). A nivell de metilació ja s'ha vist que el dèficit en metiltransferases provoca una acusada desmetilació dels elements Alu analitzats (dades mostrades en l'apart 2.1.1.2) però no sabem com això afectava l'estructura de la cromatina. Per això, es van analitzar les marques d'histones de 12 de les Alu que mostraven diferent estat de metilació.

Els elements Alu desmetilats en HCT116 es mantenen desmetilats en HCT116-DKO i per tant presenten alts nivells de marques actives. En canvi les Alus que es troben metilades en HCT116 i que es desmetilen total o parcialment en HCT116-DKO no mostren canvi en les marques d'histona, el que suggereix que no canvia l'estat d'activitat de la cromatina. L'únic increment de marca activa que veiem és en el cas de l'element Alu Lv183. Aquest element Alu, tal i com es mostra en l'apartat 2.1.1.3., on miràvem l'estat de metilació de les regions flanquejants, es troba immediatament a la regió *upstream* d'una illa CpG molt a prop de l'inici de transcripció del gen *ST5* i del gen *C11orf17*, a diferència dels altres elements Alu metilats en HCT116 i que es desmetilen en HCT116-DKO, els quals es troben a grans distàncies d'un inici de transcripció. Això suggereix de nou el paper regulador de les Alus situades prop de l'extrem 5' d'un gen.

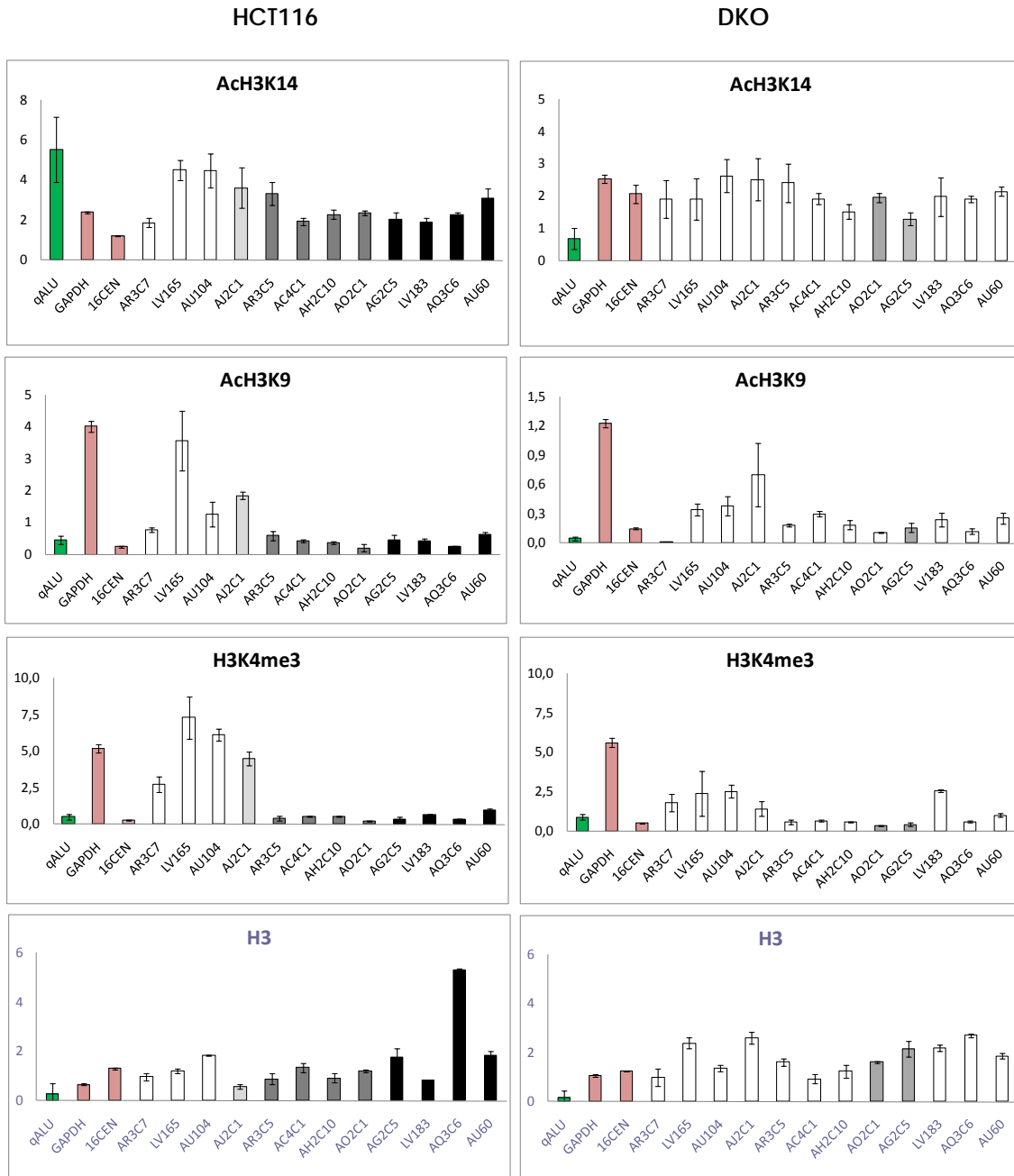


Fig. R.10. Anàlisi per ChIP dels elements Alu en les línies HCT116 i HCT116-DKO. Com a marques d'activació es troben representades ACh3K14, ACh3K9, H3K4me3. L'H3 total és el control. L'eix de les Y representa el ràtio de la marca respecte l'H3, a excepció d'H3 que és el valor absolut. Cada columna representa un element Alu i el color mostra el seu estat de metilació essent blanc desmetilat, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre metilat. Per altra banda en vermell es representen els controls del ChIP, essent GAPDH control positiu de marques actives mentre que 16CEN és el negatiu per marques actives. Per últim, la columna en verd representa ChIP-qALU, és a dir, el total d'elements Alu en la cromatina immunoprecipitada.

### 2.2.1.2. Estudi de la cromatina associada a les regions flanquejants als elements Alu

Per determinar l'estructura de la cromatina de les regions flanquejants als elements Alu estudiats es van seleccionar Alus properes a l'inici de transcripció d'un gen ja que podrien tenir un paper regulador de l'expressió més fàcilment demostrable (Figura R.11 i R.12).

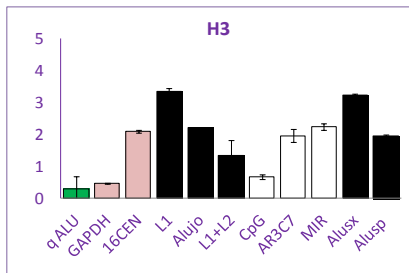
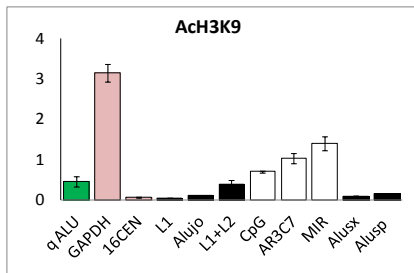
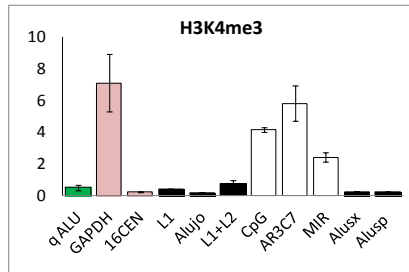
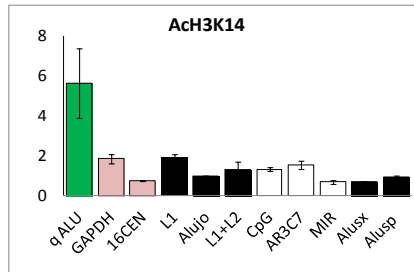
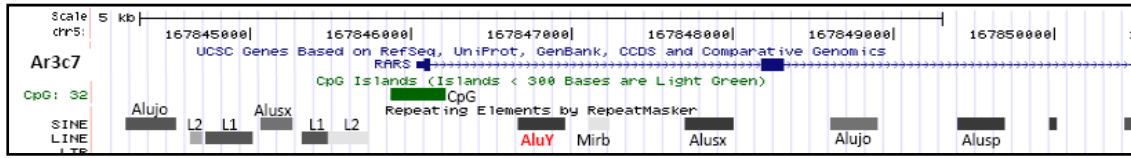
Els elements Alu que romanen desmetilats (Ar3c7, Lv165 i Au104) es troben a l'extrem 3' d'una illa CpG situada a la zona promotora d'un gen (*RARS*, *NDUFB10* i *HMOX2* respectivament) (Figura R.11). En tots aquests casos l'element Alu, l'illa CpG i la seqüència *downstream* de l'Alu estan desmetilats i presenten molta marca activa Ach3K9 i H3K4me3 de forma específica ja que la resta d'elements flanquejants que estan altament metilats tenen nivells molt baixos o nuls d'aquestes marques. En canvi, la marca activa Ach3K14 és ubiqüa i està associada a tota la regió independentment de l'estat de metilació.

Després vam analitzar aquells Alu que es trobaven parcial o totalment metilats: Aj2c1, Lv183 i Au60 (Figura R.12). En aquests elements veiem que l'illa CpG apareix desmetilada i presenta alts nivells de marca d'activació així com els elements repetitius flanquejants. En canvi, a mesura que ens allunyem, tant per l'extrem 3' com 5', els elements repetitius apareixen metilats, incloses les Alus d'estudi, i presenten nivells baixos de marca activa. Veiem doncs que en aquests casos també l'estat de desmetilació està específicament associat a les marques actives Ach3K9 i H3K4me3 però no a Ach3K14.

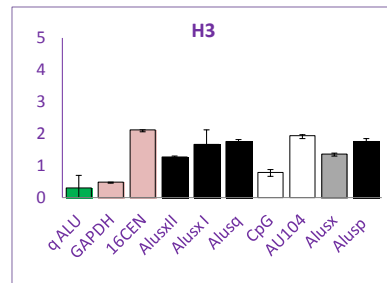
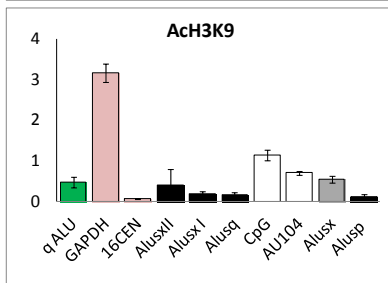
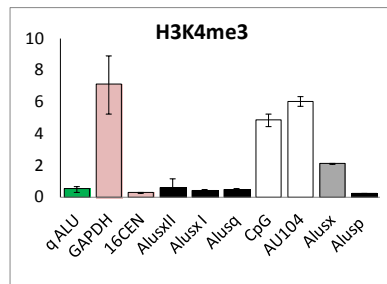
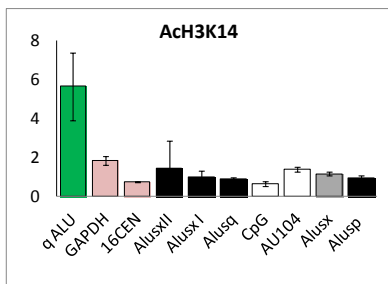
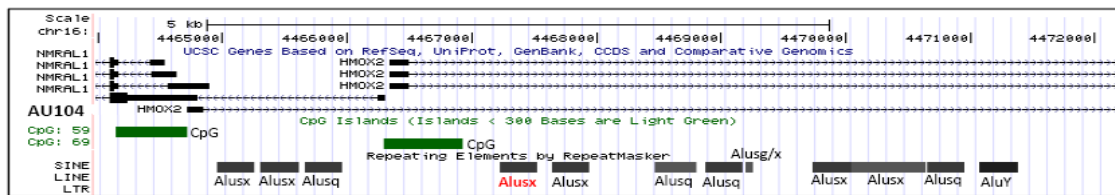
Observant les H3 totals de la figura R.11 i R.12 veiem que l'illa CpG normalment presenta menys H3 que la resta de la regió. Aquest fet està d'acord amb la literatura ja que la cromatina sol estar més descompactada en regions promotores actives, a prop de l'inici de transcripció (on estan situades aquestes CpGi), i per tant acumula menys histones, és a dir, menys nucleosomes, que altres regions.

En resum veiem un patró i una concordança entre l'estat de metilació de l'element Alu (estudiat en detall en l'apartat 2.1.1.3) i les marques d'activació. Aquest comportament es fa notable observant les marques d'activació H3K4me3 i Ach3K9, que estan específicament associades a les illes CpG i a les Alus desmetilades. De nou veiem que el perfil epigenètic dels elements Alu podria jugar un paper en la regulació transcripcional. Els resultats suggereixen que l'estat de metilació dels elements Alu estudiats definiria dominis de cromatina (marcant la frontera entre actiu/inactiu). Així doncs, segons la seva localització dins el genoma podrien participar, juntament amb altres elements, en el manteniment de l'estat de la cromatina oberta facilitant la unió de la maquinària transcripcional.

Ar3c7



Au104





## Lv165

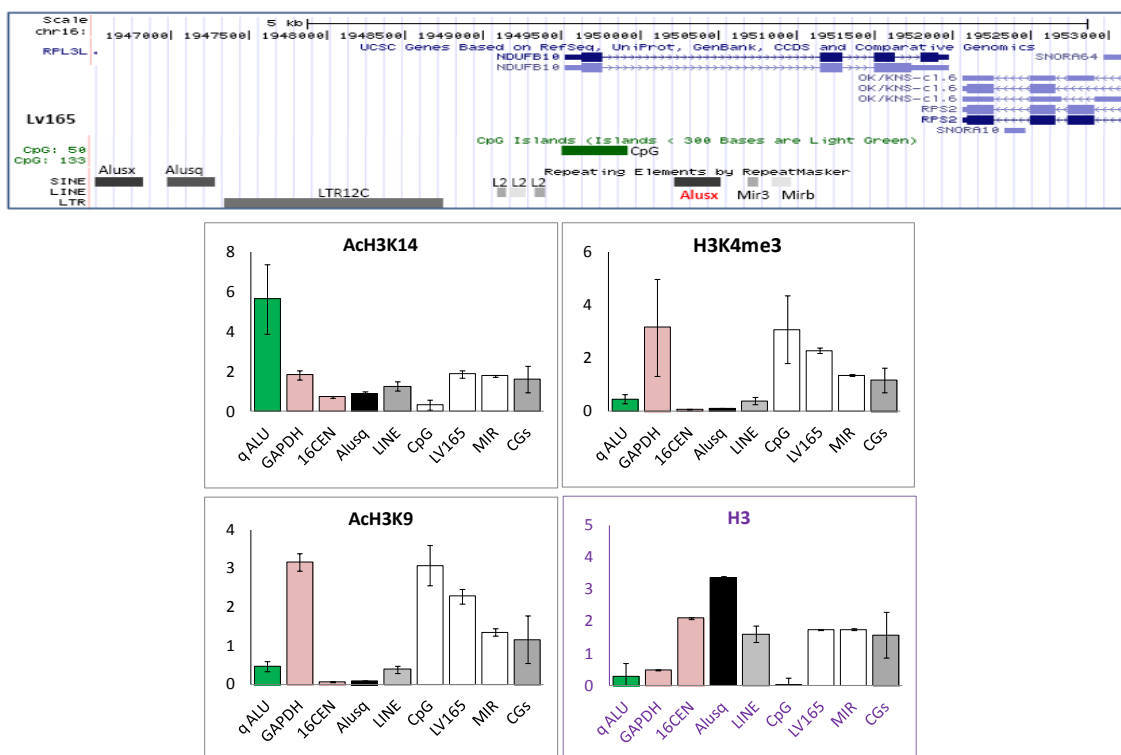
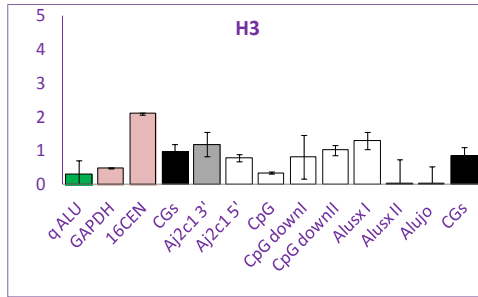
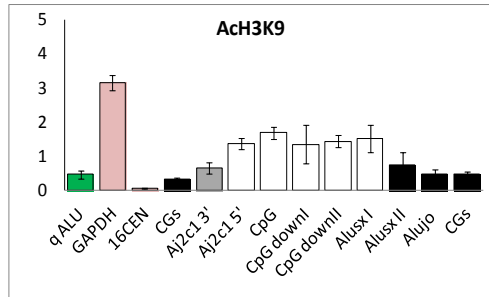
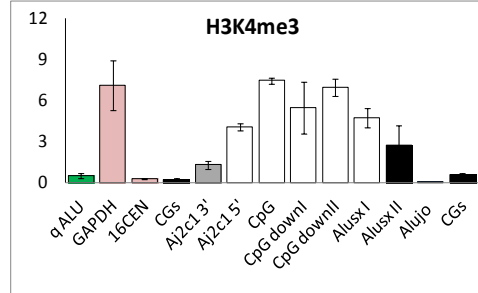
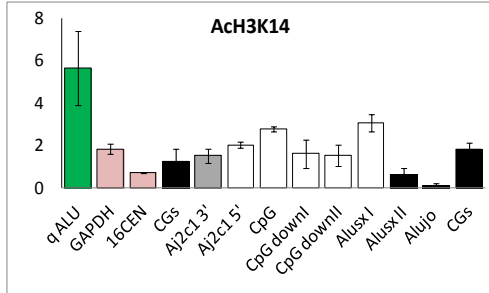
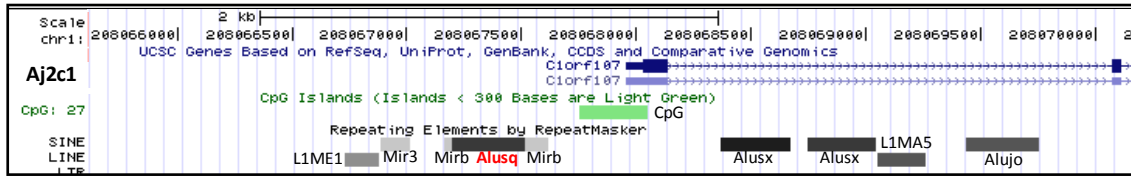
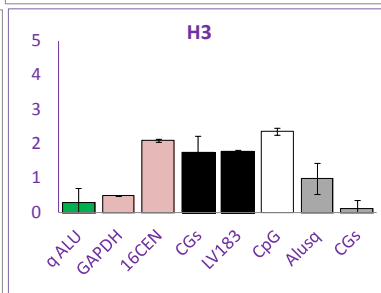
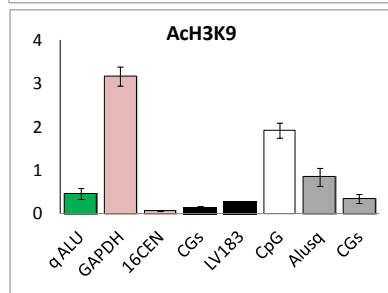
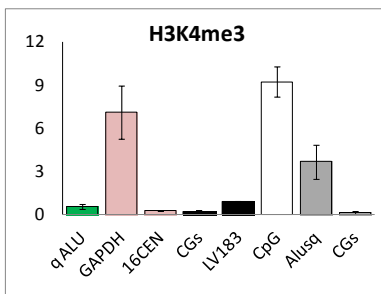
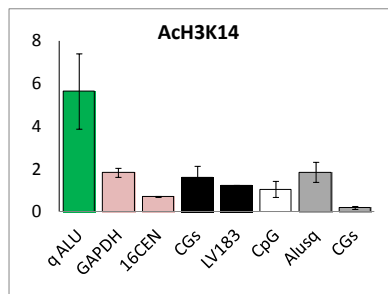
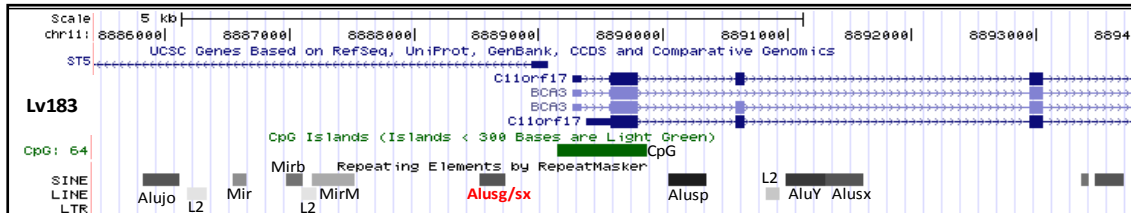


Fig. R.11. Anàlisi per ChIP de les regions flanquejants als elements Alu desmetilats (Ar3c7, Au104, Lv165) de la línia HCT116. Com a marques d'activació es troben representades AcH3K14, AcH3K9 i H3K4me3. L'H3 total és el control. L'eix de les Y representa el ràtio de la marca respecte l'H3, a excepció d'H3 que és el valor absolut. Els amplicons representats en les columnes mostren el seu estat de metilació sent blanc desmetilat, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre >75% metilat. Per altra banda en vermell es representen els controls del ChIP, essent GAPDH control positiu de marques actives mentre que 16CEN és el negatiu de marques actives. Per últim, la columna en verd representa ChIP-qALU, és a dir, el total d'elements Alu en la cromatina immunoprecipitada.

Aj2c1



Lv183



## AU60

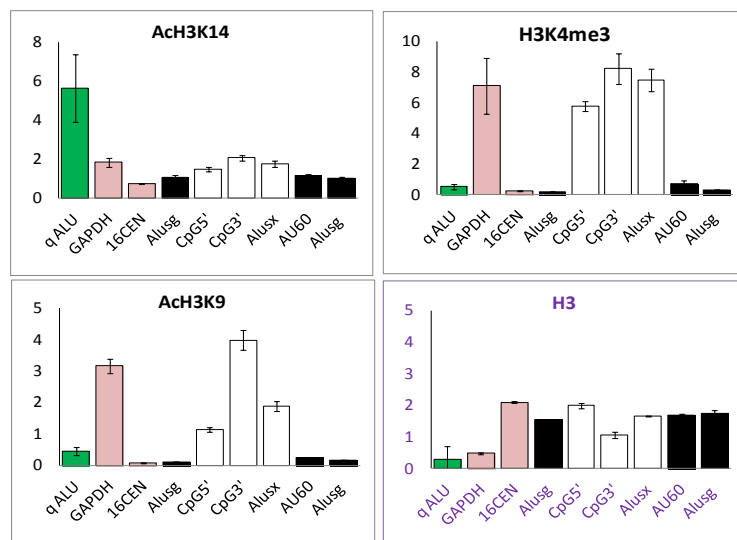
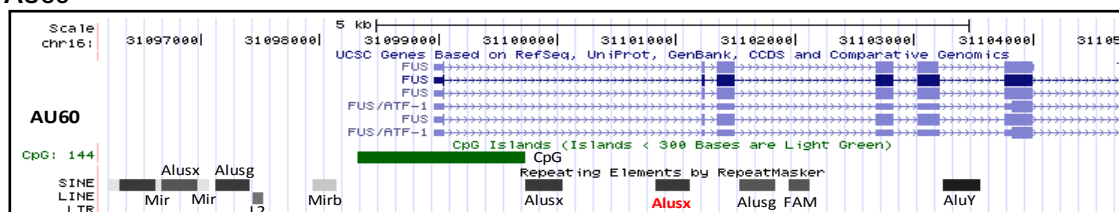


Fig. R.12. Anàlisi per ChIP de les regions flanquejants als elements Alu parcialment metilats (Aj2c1, Lv183, Au60) de la línia HCT116. Com a marques d'activació es troben representades ACh3K14, ACh3K9 i H3K4me3. L'H3 total és el control. L'eix de les Y representa el ràtio de la marca respecte l'H3, a excepció d'H3 que és el valor absolut. Els amplicons representats en les columnes mostren el seu estat de metilació sent blanc desmetilat, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre >75% metilat. Per altra banda en vermell es representen els controls del ChIP, essent GAPDH control positiu de marques actives mentre que 16CEN és el negatiu de marques actives. Per últim, la columna en verd representa ChIP-qALU, és a dir, el total d'elements Alu en la cromatina immunoprecipitada.

## 2.2.2. ESTUDI DE LA CROMATINA ASSOCIADA ALS ELEMENTS ALU EN LA DIFERENCIACIÓ

Com ja s'ha comentat, la diferenciació cel·lular suposa l'activació i repressió de milers de gens en un procés seqüencial; es basa en la reprogramació gènica d'una cèl·lula mare cap a una cèl·lula diferenciada. L'epigenètica juga un paper important en aquest procés tant a nivell de metilació, com ja s'ha vist en l'apartat 2.1.2 taula R.6, com a nivell de cromatina.

Per l'estudi a nivell de cromatina vam comparar dues línies cel·lulars: les cèl·lules mare embrionàries humanes (hESC4) i els fibroblasts (HFF) (Figura R.13). Vam triar diferents elements Alu depenent del seu estat de metilació: un grup que es manté desmetilat en ambdues línies (Ar3c7, Lv165 i Lv183), un altre que es desmetila parcial o totalment en fibroblast (Au60 i Au104) i un darrer grup que es manté metilat en ambdues línies (Aj2c1, Ao2c1 i Ag2c5).

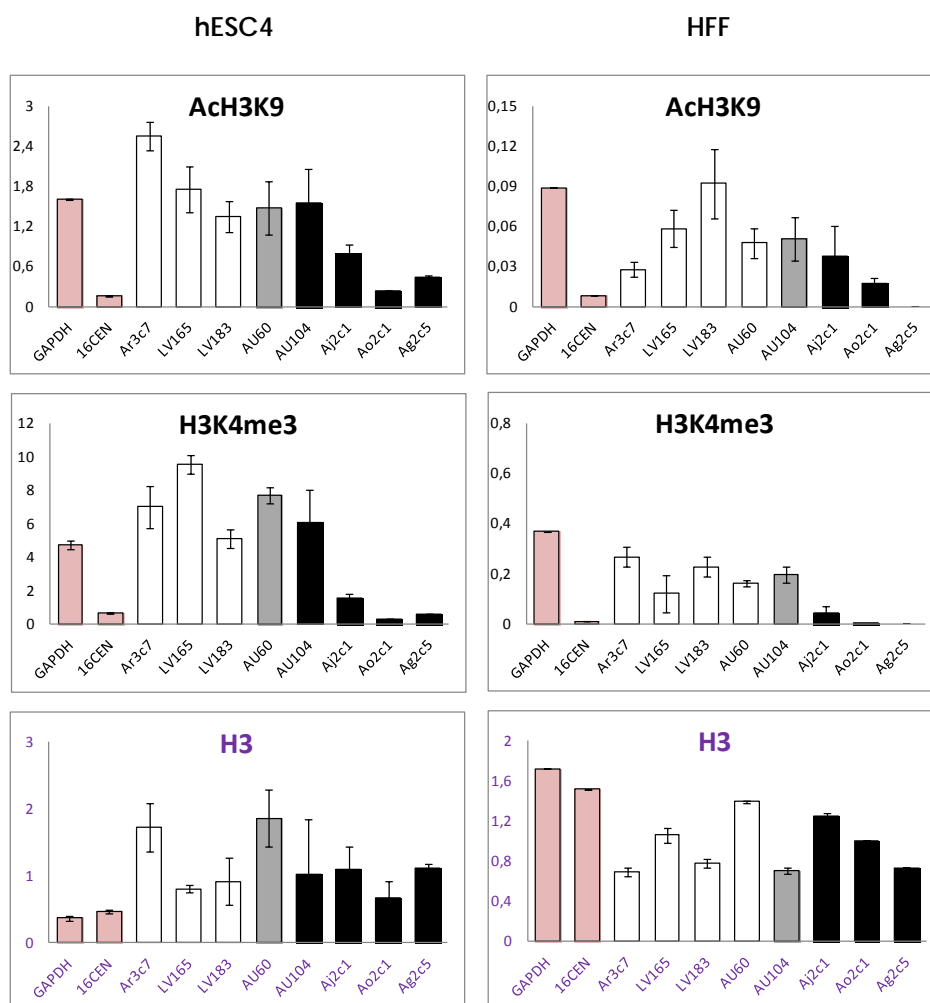


Fig. R.13. Anàlisi per CHIP dels elements Alu en les línies hESC4 i HFF. Com a marques d'activació es troben representades AchH3K9 i H3K4me3. L'H3 total és el control. L'eix de les Y representa el ràtio de la marca respecte l'H3, a excepció d'H3 que és el valor absolut. Cada columna representa un element Alu i el color mostra el seu estat de metilació essent blanc 0-25%, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre 75-100%. Per altra banda en vermell es representen els controls del CHIP, essent GAPDH control positiu de marques actives mentre que 16CEN és el control negatiu de marques actives.

Veiem que els elements que es mantenen desmetilats en els dos tipus cel·lulars (Ar3c7, Lv165 i Lv183) mostren marca en particular si ens fixem amb la marca activa H3K4me3. En canvi, Ao2c1 i Ag2c5 es mantenen metilats al llarg de la diferenciació i no presenten marques actives; recordem que aquests dos elements es troben lluny d'inicis de transcripció. Per altra banda elements com Au104 i Au60, metilats en cèl·lules mare embrionàries i que es desmetilen en les línies diferenciades presenten marca activa en ambdós línies cel·lulars.

Més endavant, i quan ens centrem en l'estudi d'un element Alu en particular (apartat 3), farem un anàlisi tant de metilació de l'ADN com de l'estat de la cromatina de tota la regió així com l'expressió dels gens propers tant en cèl·lules mare embrionàries com en fibroblasts a fi d'elucidar el paper dels elements Alu durant la diferenciació.

### 3. ELEMENTS ALU COM A POTENCIALS REGULADORS EPIGENÈTICS DE L'EXPRESSIÓ GÈNICA

Veient la relació entre l'estat de metilació de l'ADN, l'estat de la cromatina i la regió on es troben els elements Alu desmetilats vam trobar interessant aprofundir en l'estudi de dos elements Alu que podrien jugar un paper important en la regulació transcripcional.

El primer d'ells serà l'element **Aj2c1**, obtingut a partir de la tècnica AUMA al comparar una sèrie de 50 tumors colorectals i les seves parelles de teixit normal. Al llarg de l'apartat 2 ja s'ha anat estudiant aquest element. L'element Aj2c1 es troba hipometilat en tumor respecte al teixit normal. Es tracta d'un Alu Sq, situat a -440pb del gen *C1orf107* (*digestive-organ expansion factor homolog, DIEXF*), un gen relacionat amb la via de p53 i l'aparell digestiu (Chen et al., 2008).

El segon element d'estudi serà **ZALU3**. Aquest element es va obtenir més recentment, a partir de la tècnica NSUMA (tècnica explicada en l'apartat 2.3.3 de material i mètodes), una variant de la tècnica AUMA la qual permet igualment identificar elements Alu desmetilats. Amb la tècnica NSUMA es van analitzar diferents línies cel·lulars de CCR, cèl·lules mare embrionàries, cèl·lules pluripotents induïdes, queratinòcits, fibroblasts i 3 parelles normal-tumor. El producte resultant es va analitzar per ultraseqüenciació (fent servir la plataforma Hi-Seq d'Illumina). Amb aquesta tècnica és va obtenir un ampli mapa d'elements Alu desmetilats que posteriorment es van analitzar bioinformàticament, confirmant que els elements Alu constituïen un compartiment definit i amb propietats específiques (Jordà M. et al. en preparació). L'element d'estudi es va trobar altament desmetilat en cèl·lules mare embrionàries i el vam considerar interessant degut a la seva possible vinculació en la diferenciació. ZALU3 es tracta d'un Alu Sc, situat a -1064pb del gen *GLDC* (*glycine dehydrogenase*), un gen relacionat amb el metabolisme de la glicina (Zhang et al., 2012).

### 3.1. ALU AJ2C1

#### 3.1.1. ESTUDI EPIGENÈTIC DE L'ELEMENT AJ2C1

L'Alu Aj2c1 es localitza a prop d'una illa CpG associada al gen *C1orf107* o *Digestive organ expansion factor homolog* de les sigles en anglès *DIEXF* (Figura R.14). Aquest gen ens va resultar interessant ja que en *Zebrafish* està descrit que regula la via de *p53* controlant el creixement dels òrgans digestius (Chen et al., 2005), fet que el fa molt "atractiu" ja que estem treballant en càncer colorectal.

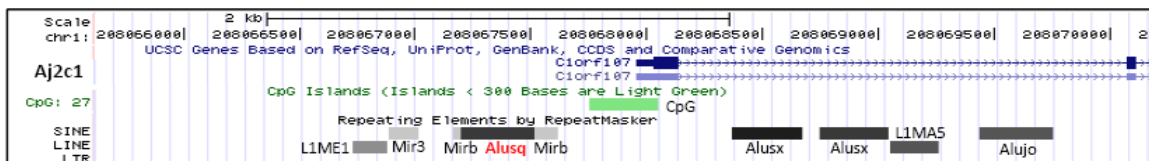


Fig. R.14. Esquema de la regió d'estudi obtinguda del Genome Browser versió Mar. 2006 (NCBI36/hg18). En vermell l'element Alu d'estudi, Aj2c1 (Alu sq).

Aquest element es va trobar hipometilat en tumors corresponents a les parelles Normal-Tumor utilitzades en l'AUMA i es van validar els resultats en 12 parelles Normal-Tumor mitjançant la tècnica de bisulfit sòdic i la posterior seqüenciació (Figura. R.15 ). A més en l'apartat 2 ja vam veure que presentava una bona correlació entre la metilació de l'ADN i l'estat de la cromatina (Figura R.12).

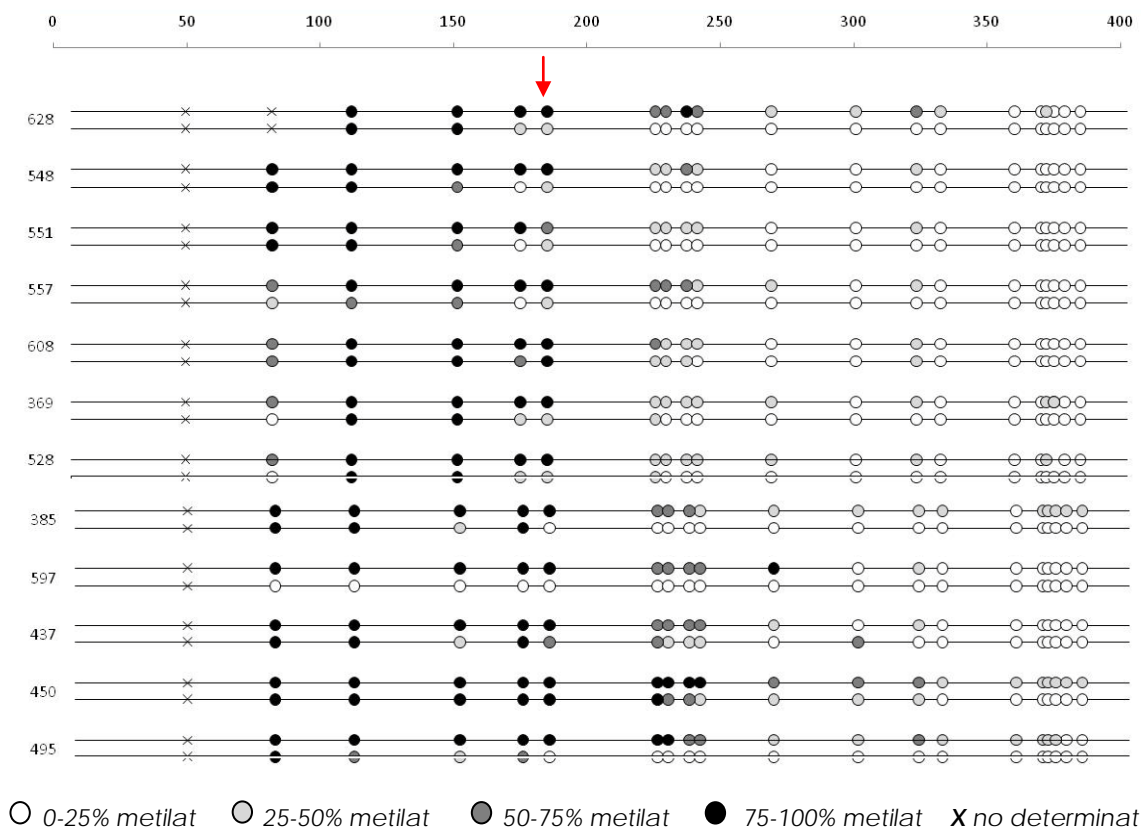


Fig. R.15. Estat de metilació de l'element Alu Aj2c1 en 12 parelles Normal-Tumor. La fletxa vermella indica la CpG corresponent a la diana Smal, utilitzada en la tècnica AUMA.

Per altra banda aquest element Alu està parcialment metilat, de manera que s'observa una zona intermèdia metilada-desmetilada o barrera de metilació que varia entre els diferents tipus cel·lulars (Figura R.16), suggerint un perfil de metilació específic del tipus cel·lular. En línies de CCR veiem un patró heterogeni, sent Sw480 la més metilada i HCT116, incloent la seva variant DKO, la més desmetilada.

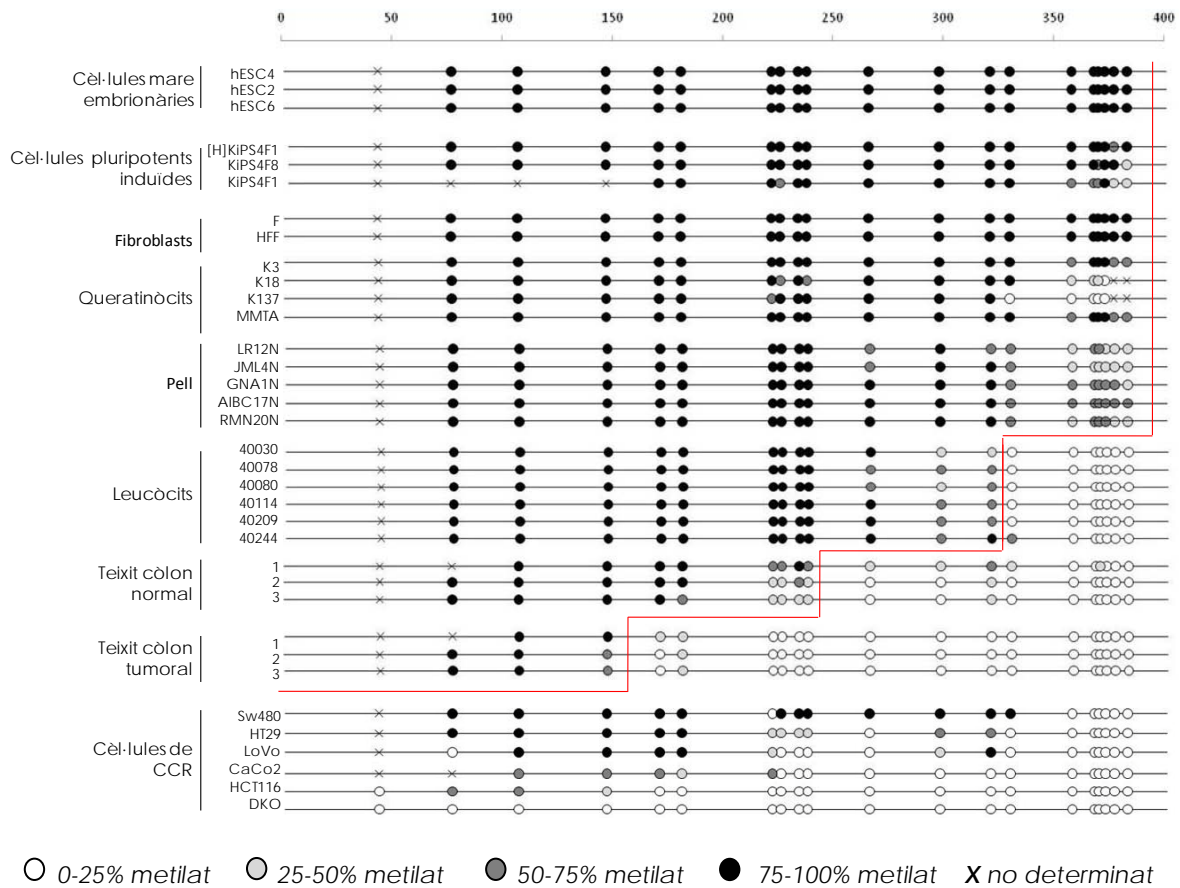


Fig. R.16. Estat de metilació de l'element Alu Aj2c1 en mostres normals (cèl·lules mare embrionàries, cèl·lules pluripotents induïdes, fibroblasts, queratinòcits, pell, leucòcits i teixit de còlon normal) i mostres tumorals (teixit de còlon tumoral i línies de CCR).



### 3.1.2. ESTUDI EPIGENÈTIC DE LA REGIÓ FLANQUEJANT A AJ2C1

Primerament ens vam centrar en l'estat de metilació de la regió de l'element repetitiu Aj2c1 en diferents línies cel·lulars així com parelles Normal-Tumor (figura R.17). En la figura R.17 veiem que "la barrera de metilació" dins de l'Alu Aj2c1 varia segons el tipus cel·lular. Així doncs, Aj2c1 marcaria l'inici de desmetilació de la regió que continua direcció 3' en l'illa CpG i la primera Alu Sx. També veiem que l'element Aj2c1 mostra un gradient regional en el perfil de metilació, el que és consistent amb una regulació dinàmica de la barrera entre dominis actius/inactius en la cromatina.

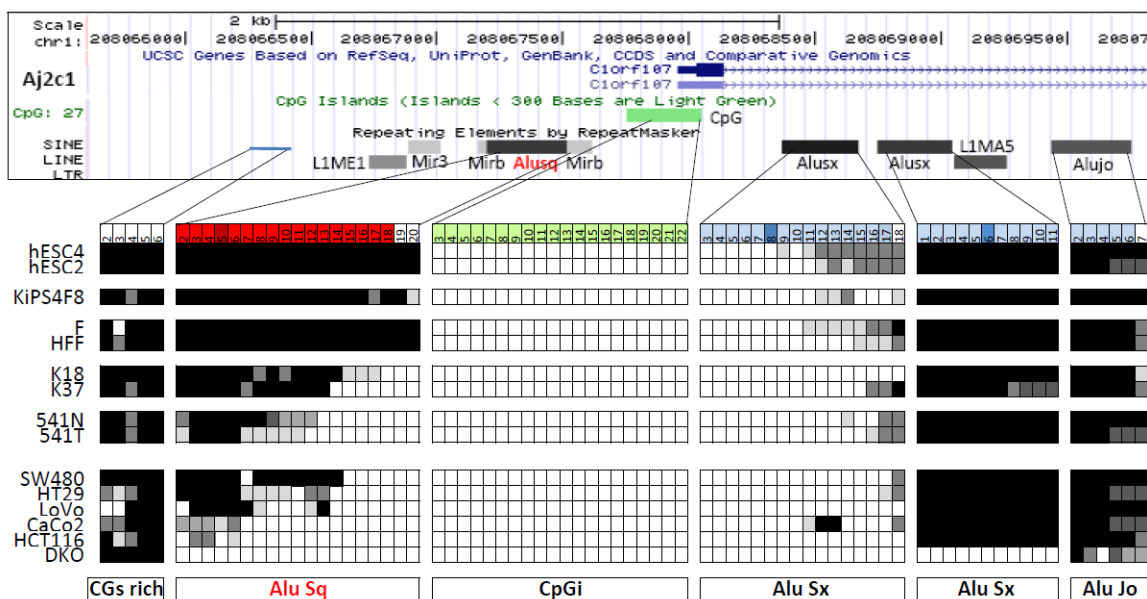


Fig. R.17. Estat de metilació de la regió flanquejant a l'element Alu seleccionat Aj2c1 (en vermell) en cèl·lules mare embrionàries (hESC4 i hESC2), cèl·lules pluripotents induïdes (KiPS4F8), fibroblasts (F, HFF) queratinòcits (K18 i K37), una parella normal-tumor (541N i 541T) i diferents línies de CCR (SW480, HT29, LoVo, CaCo2, HCT116 i DKO). Regió obtinguda del Genome Browser versió Mar. 2006 (NCBI36/hg18). Cada requadre indica una CpG. En vermell trobem l'Alu Aj2c1, en verd l'illa CpG i en blau diferents elements repetitius. En blanc de 0-25% de metilació, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre 75-100% de metilació.

També es van mirar diverses marques d'histona i mostrem, com en els altres casos, les marques actives de Ach3K14, H3K4me3, Ach3K9 i l'H3 total. Com que l'Alu Aj2c1 està parcialment metilada en algunes mostres, es van dissenyar encebadors per analitzar les marques d'histona associades específicament a cada meitat de l'Alu (Aj2c1 3' i Aj2c1 5'). La localització dels encebadors de la regió és troben en l'annex III AIII.1.

En la figura R.18 mostrem únicament la marca H3K4me3 i H3 en les línies de hESC2, HFF, Sw480, LoVo, HCT116 i DKO. Veiem que en totes les mostres analitzades, la correlació entre estat de metilació i marques actives és elevada, observant més marca quan la regió està més desmetilada. Al comparar les diferents mostres, el que resulta més interessant és que el perfil de marques actives segueix perfectament els dominis desmetilats: quan més metilada està l'Alu més petit és el domini de cromatina activa (seria el cas de hESC2 i HFF) i a mida que la regió es va desmetilant es fa més ampla la regió marcada per cromatina activa (HCT116 i DKO).

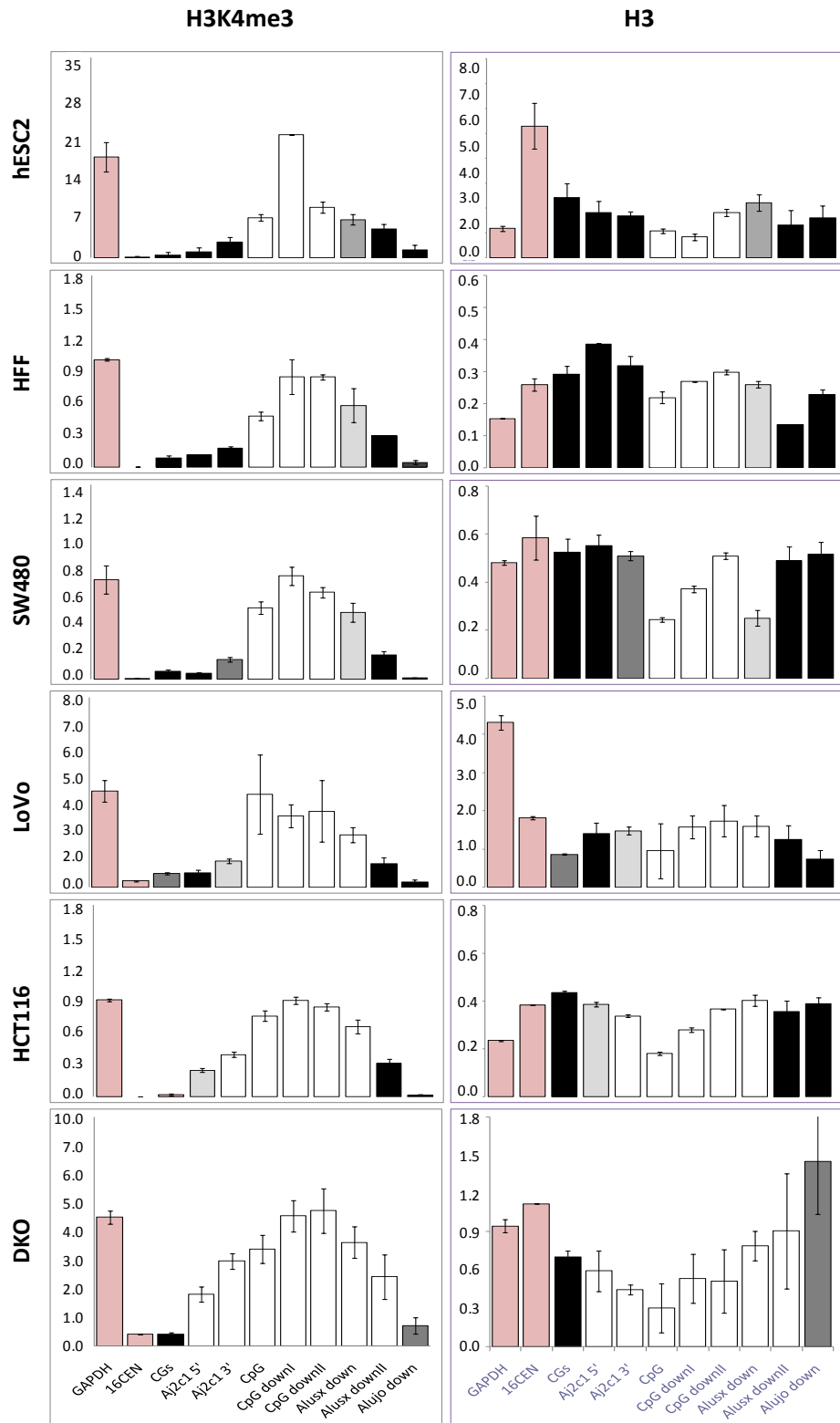


Fig. R.18. Anàlisi per CHIP de les regions flanquejants a l'element Alu parcialment metilat Aj2c1 en les línies hESC2, HFF, Sw480, LoVo, HCT116 i DKO. Es representen només la marca d'activació H3K4me3 i l'H3 total. L'eix de les Y representa el ràtio de la marca respecte l'H3 i l'H3 és el valor absolut. Els amplicons representats en les columnes mostren el seu estat de metilació sent blanc desmetilat, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre >75% metilat. Per altra banda en vermell es representen els controls, sent GAPDH control positiu de marca activa i 16CEN control negatiu de marca activa.

### 3.1.3. ESTUDI DE L'EXPRESSION DE *DIEXF* A NIVELL D'ARNm

Centrant-nos ja en el gen proper a l'element Aj2c1 vam observar que hi havia diferents transcrits del mateix gen segons les dades obtingudes de l'Ensembl/Havana (Figura R.19). Els transcrits 1 i 3 són idèntics a nivell de seqüència a excepció de l'UTR5', que comença abans en el transcrit 1, i l'UTR3', que és major en el transcrit 3. Ambdós transcrits contenen 12 exons i tenen una mida de 3094 pb en el cas del transcrit 1 i de 8422 pb en el cas del transcrit 3. El transcrit 2 només conté 3 exons i una mida de 2896 pb.

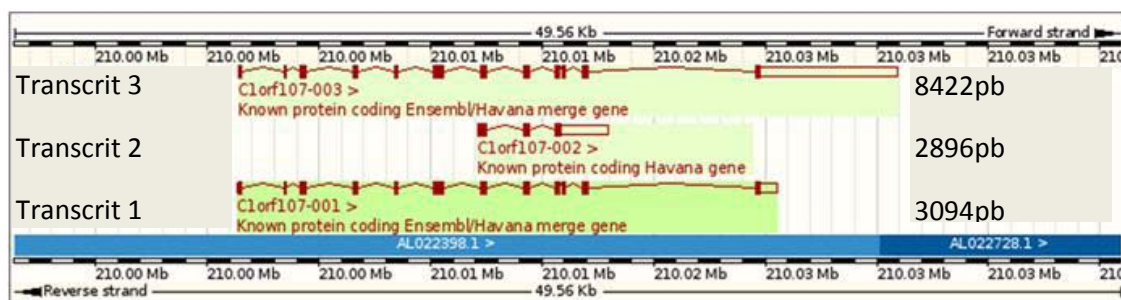
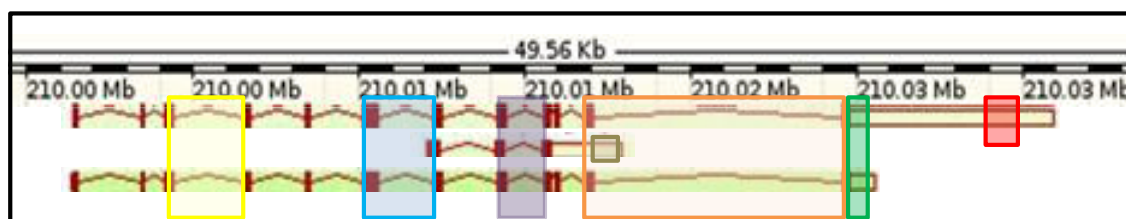


Fig. R.19. Transcrits alternatius del gen *DIEXF*. Segons l'Ensembl versió March 2006 (NCBI36/hg18)

L'anàlisi de l'expressió es va fer mitjançant RT-qPCR i es van dissenyar diversos encebadors a fi de cobrir els diferents transcrits, alguns d'ells específics de cada transcrit i altres comuns per poder fer un estudi amb detall de l'expressió (Taula R.8).

TAULA R.8. DISSENYS DELS ENCEBADORS PER L'ESTUDI D'EXPRESSION DELS DIFERENTS TRANSCRITS DEL GEN *DIEXF*

NOM ENCEBADORS	LOCALITZACIÓ	TRANSCRITS COBERTS
C1orf107 3-4	Exó 3 i 4	1+3
C1orf107 6-7	Exó 6 i 7	1+3
C1orf107 8-9	Exó 8 i 9	1+2+3
C1orf107 11-12	Exó 11 i 12	1+3
C1orf107 12-UTR003	Exó 12 i regió UTR3' comú als transcrits 1 i 3	1+3
C1orf107 UTR003	Regió UTR3' específica del transcrit 3	3
C1orf107 UTR002	Regió UTR3' específica del transcrit 2	2



Per determinar si l'estat de metilació juga un paper en la regulació transcripcional del gen *DIEXF* es van mirar l'estat de metilació de l'element Alu Aj2c1 així com l'expressió del gen *DIEXF* en diferents línies de CCR (HCT116, HT29, LoVo i Sw480), cèl·lules mare embrionàries humanes (hESC2, hESC4, hESC6), cèl·lules pluripotents induïdes ([H]KiPS4F1, KiPS4F1 i KiPS4F8), una línia de queratinòcits (K1) i una de fibroblasts (HFF) (Figura R.20).

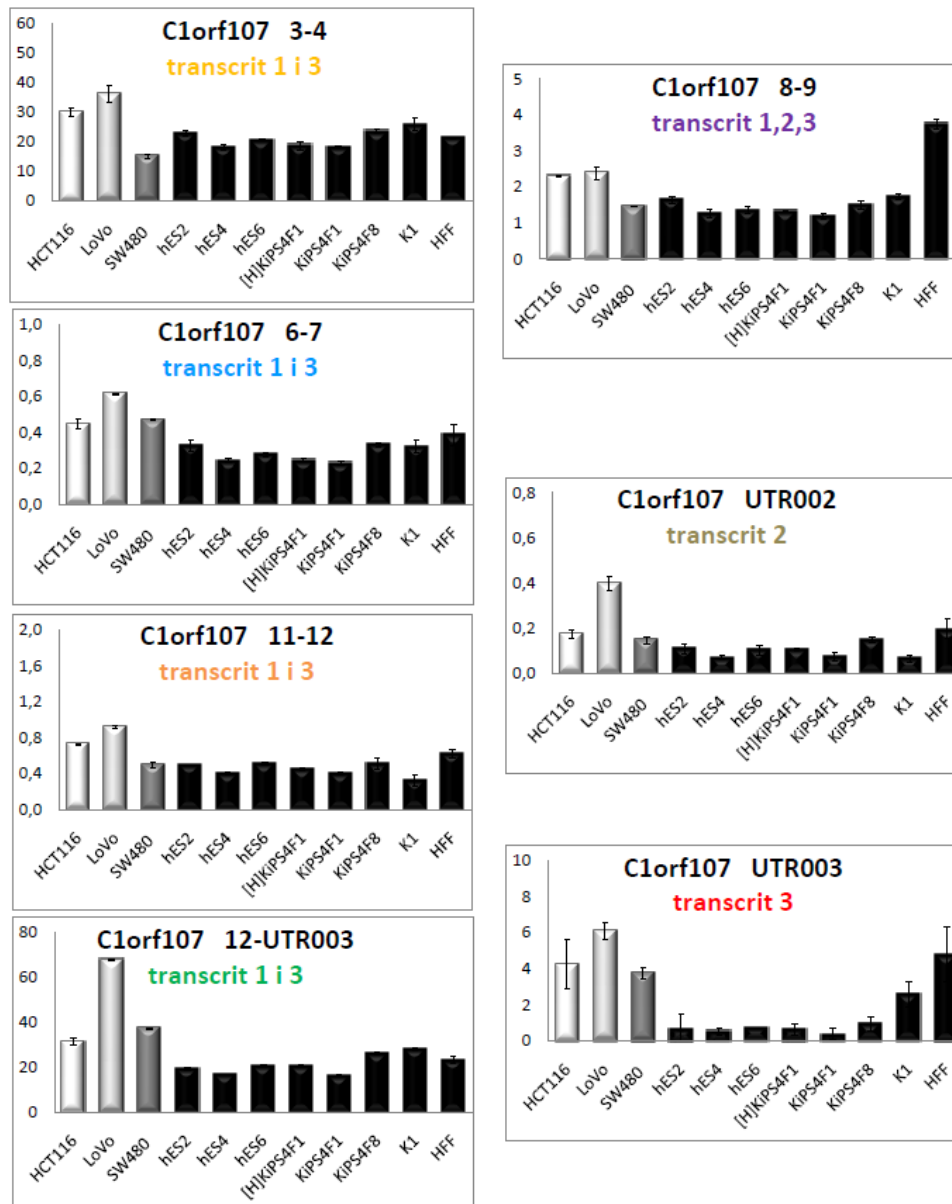


Fig. R.20. Nivells d'expressió de *DIEXF* en diferents línies de CCR (HCT116, HT29, LoVo i Sw480), cèl·lules mare humanes (hES2, hES4, hES6), cèl·lules pluripotents induïdes ([H]KiPS4F1, KiPS4F1 i KiPS4F8), queratinòcits (K1), fibroblasts (HFF). L'eix de les Y representa l'expressió relativa. La normalització està feta amb la mitjana dels estàndards. L'estat de metilació de l'element Aj2c1 es representa en escala de grisos, sent el blanc totalment desmetilat i el negre totalment metilat.

Pel que fa a l'expressió del gen *DIEXF* veiem que les regions 3-4, 6-7, 11-12 i 12-UTR003 (comuns als transcrits 1 i 3) presenten una expressió similar en totes les línies essent major en aquelles que es troben parcialment desmetilades. En la regió 8-9 (comuna als tres transcrits) veiem un patró similar a excepció de la línia HFF en el qual no hi ha gaire correlació entre metilació de l'Alu i expressió gènica.

Quan s'analitzen transcrits específics: concretament la regió UTR002 (específica del transcrit 2) i UTR003 (específica del transcrit 3) veiem una clara disminució de l'expressió en les línies totalment metilades (cèl·lules mare i cèl·lules pluripotents induïdes), a excepció de la línia HFF.

Per confirmar els resultats obtinguts per RT-qPCR, vam analitzar el nivell d'ARNm dels diferents transcrits per Northern Blot (tècnica explicada en detall en l'apartat material i mètodes 2.2.4.). Es van dissenyar diverses sondes, una que hibridava amb els 3 transcrits, concretament en la regió dels exons 7-8-9 (sonda *C1orf107* 7-9) i una altra que hibridava amb l'UTR3' del transcrit 3 (sonda *C1orf107* utr003) (Figura R.21).

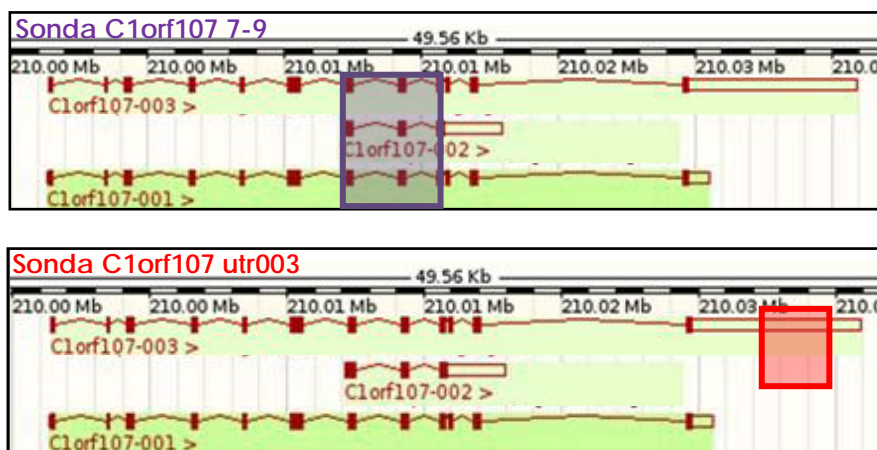


Fig. R.21. Localització de les sondes dissenyades pel Northern Blot.

L'experiment es va fer amb línies cel·lulars de diferent llinatge (Figura R.22). Es van mirar cèl·lules mare humanes (hESC4), cèl·lules pluripotents induïdes (KiPS4F1 i KiPS4F8), *embryoid bodies* (EB), diferents línies de CCR (HCT116, DKO, Sw480), una línia de queratinòcits (K1) i una de fibroblasts (HFF).

En hibridar amb la sonda *C1orf107* 7-9 (Figura R.22), comú als 3 transcrits, ens apareixen 3 bandes que es corresponen, segons el pes molecular, als 3 transcrits descrits. Així doncs el transcrit 3 és el de major pes, amb 8422 pb i és la banda superior i els transcrits 1 i 2 corresponen al doblet de la imatge, i de menor pes, 3094 i 2896 pb respectivament. Veiem una correlació entre estat de metilació i d'expressió. S'observa que les línies de CCR, parcialment desmetilades, presenten els tres transcrits, donant una forta senyal en totes tres. En canvi, les cèl·lules mare, les línies pluripotents induïdes, els queratinòcits i els fibroblasts, que tenen l'Alu *Aj2c1* altament metilada, veiem la desaparició gairebé absoluta de la banda corresponent al transcrit 3, similar al que veiem en les RT-qPCRs. En HFF, no obstant, tot i haver-hi més expressió gènica segons la RT-qPCR veiem uns nivells baixos d'ARN en el Northern, corresponent-se amb l'estat de metilació.

Després de deshibridar la membrana es va hibridar amb la sonda *C1orf107* utr003 observant una única banda (Figura R.22) corresponent al transcrit 3, i confirmem que la seva expressió gènica presenta una correlació negativa amb l'estat de metilació de l'Alu.

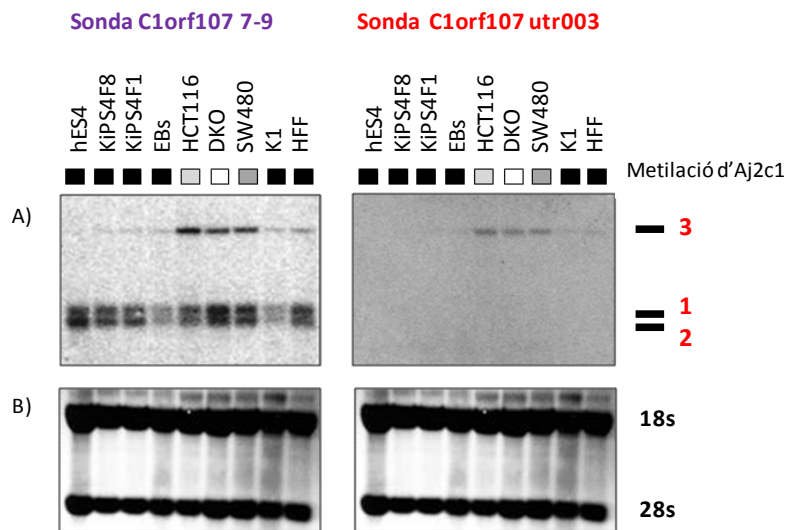
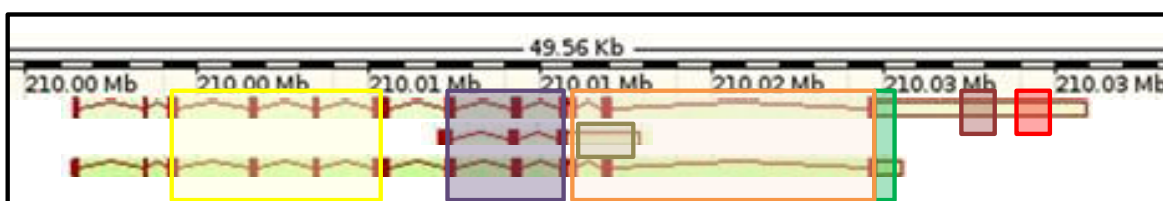


Fig. R.22. Estudi de DIEXF a nivell d'ARN per Northern Blot de diferents línies cel·lulars. A) Sonda *C1orf107* 7-9 que cobreix els 3 transcrits unint-se en la regió dels exons 7-8-9. I la sonda *C1orf107* utr003 que hibrida amb el transcrit 3. En vermell es marcar en el transcrit al que correspon la banda. B) Tinció amb blau de metilè per visualitzar l'ARN ribosòmic (18S i 28S). Els quadrats superiors mostren l'estat de metilació global de l'Alu Aj2c1 de les mostres, essent negre 100% metilat i blanc 0% metilat.

Per aprofundir en l'estudi es va analitzar per Northern blot el perfil de transcrits expressats en la línia Sw480 (Figura R.23) utilitzant diferents sondes al llarg de tots els exons, tant específics com comuns (Taula R.9).

TAULA R.9. SONDES PER L'ESTUDI D'EXPRESSION DELS DIFERENTS TRANSCRITS DEL GEN DIEXF PER NORTHERN BLOT

NOM SONDES	LOCALITZACIÓ	TRANSCRITS COBERTS
C1orf107 3-6	Exó 3 a 6	1+3
C1orf107 7-9	Exó 7 a 9	1+2+3
C1orf107 10-12	Exó 10 i 12	1+3
C1orf107 12-UTR003	Exó 12 i regió UTR3' comú als transcrits 1 i 3	1+3
C1orf107 UTR003	Regió UTR3' específica del transcrit 3	3
C1orf107 UTR003 BIS	Regió UTR3' específica del transcrit 3	3
C1orf107 UTR002	Regió UTR3' específica del transcrit 2	2



La sonda *C1orf107* 7-9 dóna 3 bandes corresponents als 3 transcrits descrits. Les sondes *C1orf107* 3-6, *C1orf107* 10-12, *C1orf107* 12-UTR003, que segons el descrit en el *Genome Browser* cobreixen els transcrits 1 i 3, donen 3 bandes. La sonda *C1orf107* UTR003 bis es va dissenyar per confirmar la hibridació específica al transcrit 3 que observàvem amb la sonda *C1orf107* UTR003. Pel que fa a la sonda *C1orf107* UTR002, esperàriem que hibridés amb el transcrit 2 però no observem cap banda.

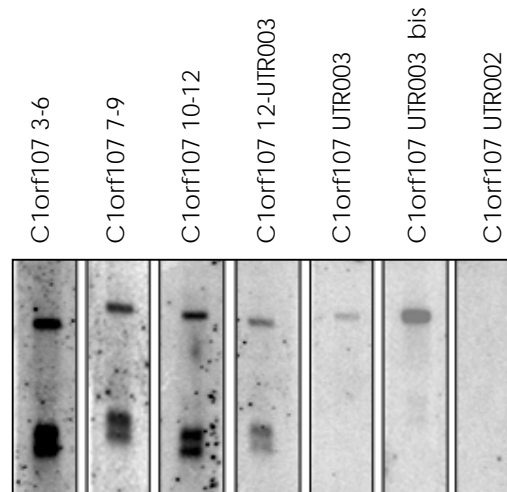


Fig. R.23. Estudi de diferents sondes dissenyades al llarg del gen *DIEXF* hibridades en la línia SW480.

Les dades obtingudes per Northern Blot ens van suggerir la possible existència d'altres transcrits no descrits.

A fi de poder caracteritzar millor el gen *DIEXF* i els seus transcrits vam optar per emprar la tècnica RACE (de l'anglès *Rapid Amplification of cDNA Ends*) (Zhumabayeva et al., 2004) (tècnica explicada en detall en l'apartat material i mètodes 2.3.). La tècnica RACE és un mètode que permet clonar la seqüència completa d'un ADNc d'estudi a partir d'una seqüència parcial prèviament coneguda basant-se amb la tècnica de PCR. Aquesta tècnica permet realitzar una amplificació dels ADNc extrems tant en 5' com en 3'.

Així doncs vam voler amplificar diferents fragments a fi d'identificar possibles nous transcrits. Es van dissenyar encebadors específics per l'ADNc d'estudi tant en 3' (8F, 12F4, 12F3 i 12F2) com en 5' (9R, 7R i 3R) (Figura R.24).

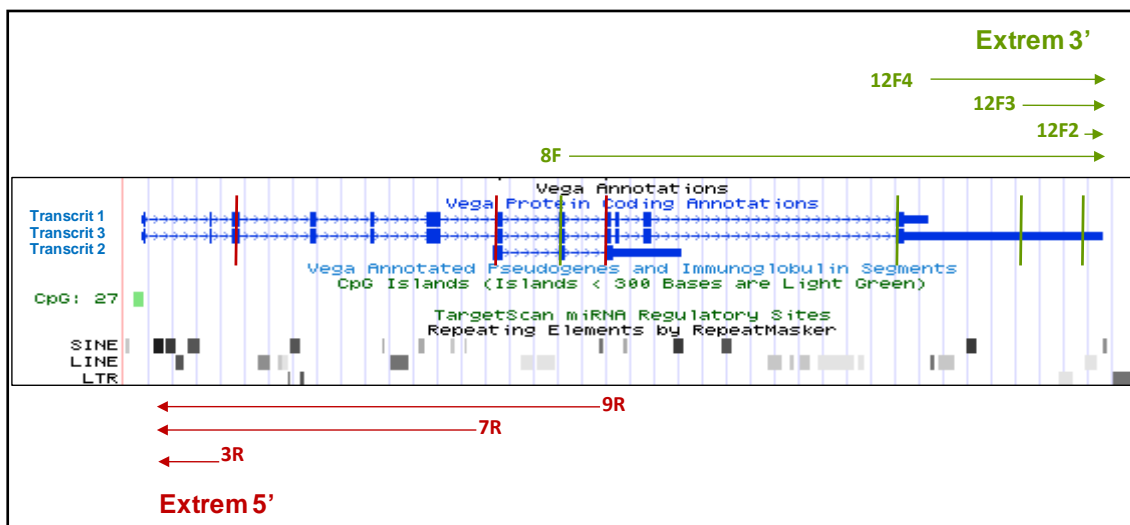


Fig. R.24. Posició dels encebadors de la tècnica RACE. En vermell la regió 5' i en verd la 3'

Es va fer l'amplificació amb encebadors específics de la nostra seqüència obtenint fragments de diferents mides segons els transcrits descrits en el *Genome Browser* (Taula R.10). A la mida esperada de la taula R.10 cal sumar-li els pb dels encebadors emprats en cada cas, així com la seqüència "extra" que s'incorpora segons el protocol del kit.

TAULA R.10. MIDA ESPERADA DE LES AMPLIFICACIONS AMB ELS ENCEBADORS DEL RACE

ISOFORMES DEL GEN	MIDA ESPERADA						
	DIRECCIÓ 3'				DIRECCIÓ 5'		
	12F2	12F3	12F4	8F	3R	7R	9R
Transcrit 1	-	-	902	1579	406	1173	1688
Transcrit 3	631	2199	6254	6931	382	1149	1664
Transcrit 2	-	-	-	2395	-	158	673



Les línies escollides per l'estudi van ser la línia de CCR Sw480, els fibroblasts (HFF) i les cèl·lules mare embrionàries (hESC4).

A l'hora d'estudiar l'extrem 3' vam amplificar 4 fragments (Figura R.25). Per l'estudi específic del transcrit 3 es van emprar els encebadors 12F2 i 12F3. Com era d'esperar i d'acord amb la taula R.10, en amplificar amb 12F2 veiem una banda tènue únicament en Sw480 corresponent al transcrit 3 ( $\approx 700$ pb) i amb 12F3 veiem una única banda en Sw480 d'aproximadament 2200pb corresponent també al transcrit 3. Ambdues bandes es van clonar i seqüenciar (explicat a continuació en la figura R.26). Aquests resultats confirmen que Sw480 expressa el transcrit 3 però no HFF i hESC4 com mostrava el Northern Blot.

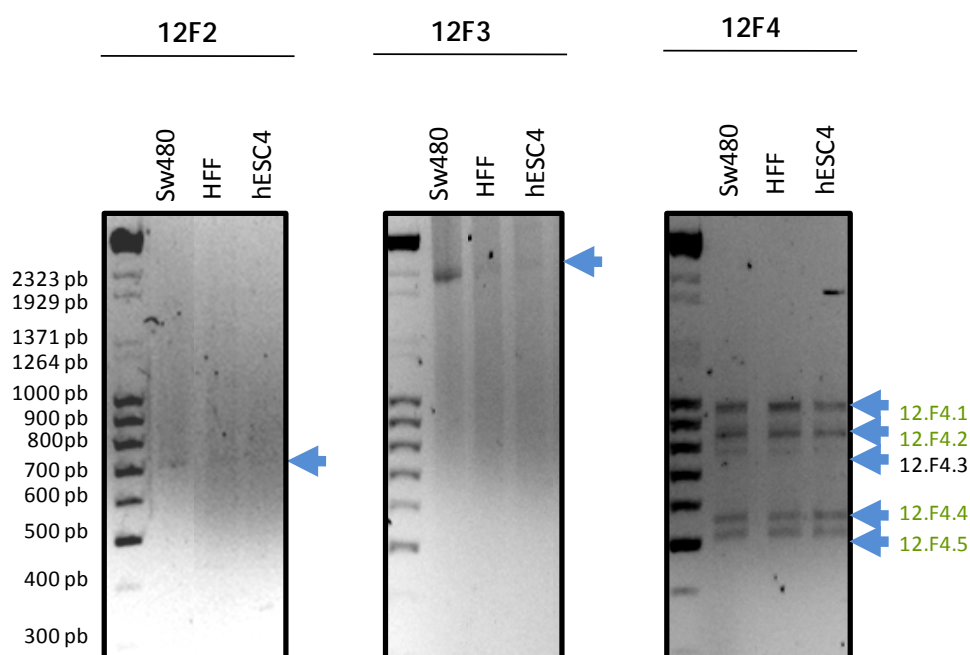


Fig. R.25. 3'-RACE de les línies Sw480, hESC4 i HFF. Resultat de la PCR dels encebadors específics del nostre ADNc d'interès en 3' (12F2, 12F3 i 12F4). El marcador de pes molecular és BstIII i el DNA ruler de 100pb. Les fletxes blaves ens indiquen els fragments amplificats.

L'amplificació obtinguda amb 12F4 ens va donar cinc bandes tot i que caldria esperar, segons el descrit (Taula R.10), únicament dues bandes corresponents al transcrit 3 (de 6254pb) i al transcrit 1 (de 902pb). Les cinc bandes del gel corresponen als fragments 12.F4.1 ( $\approx 900$ pb), 12.F4.2 ( $\approx 750$ pb), 12.F4.3 ( $\approx 650$ pb), 12.F4.4 ( $\approx 450$ pb) i 12.F4.5 ( $\approx 410$ pb). 12F4.1 semblaria correspondre al transcrit 1 descrit però la resta de fragments no havien estat prèviament descrits, fent pensar en nous possibles transcrits. Pel que fa al fragment corresponent al transcrit 3 no s'ha amplificat, probablement degut a la seva mida (6254pb), però amb les sondes 12F2 i 12F3 ja confirmem que existeix aquest transcrit. Vam clonar i seqüenciar 4 dels 5 fragments obtinguts, descartant 12F.4.3 per manca de material.

El fragment 8F ens va donar bandes inespecífiques (dades no mostrades) i no es va seguir el procés de clonació.

Es van clonar els amplicons de línies Sw480 i hESC4 i de cada mostra es van seqüenciar de 4 a 8 clons. En la figura R.26 representem la seqüència majoritària obtinguda. Així doncs, si mirem les seqüències dels clons de 12F2 i 12F3 veiem que corresponen al transcrit 3 tal i com caldria esperar, ja que els encebadors serien específics per dit transcrit. Aquest transcrit no s'expressa en hESC4.

Amb 12F4 només vam analitzar Sw480 ja que les bandes que apareixien eren iguals en totes les línies cel·lulars analitzades. Al clonar i seqüenciar les 4 bandes provinents del fragment 12F4 veiem que el fragment 12F4.1 correspondria al transcrit 1. En canvi els altres fragments correspondrien a nous transcrits, que vam anomenar per 12F4.2, el transcrit 4, per 12F4.4 el transcrit 5 i pel 12F4.5 el transcrit 6.

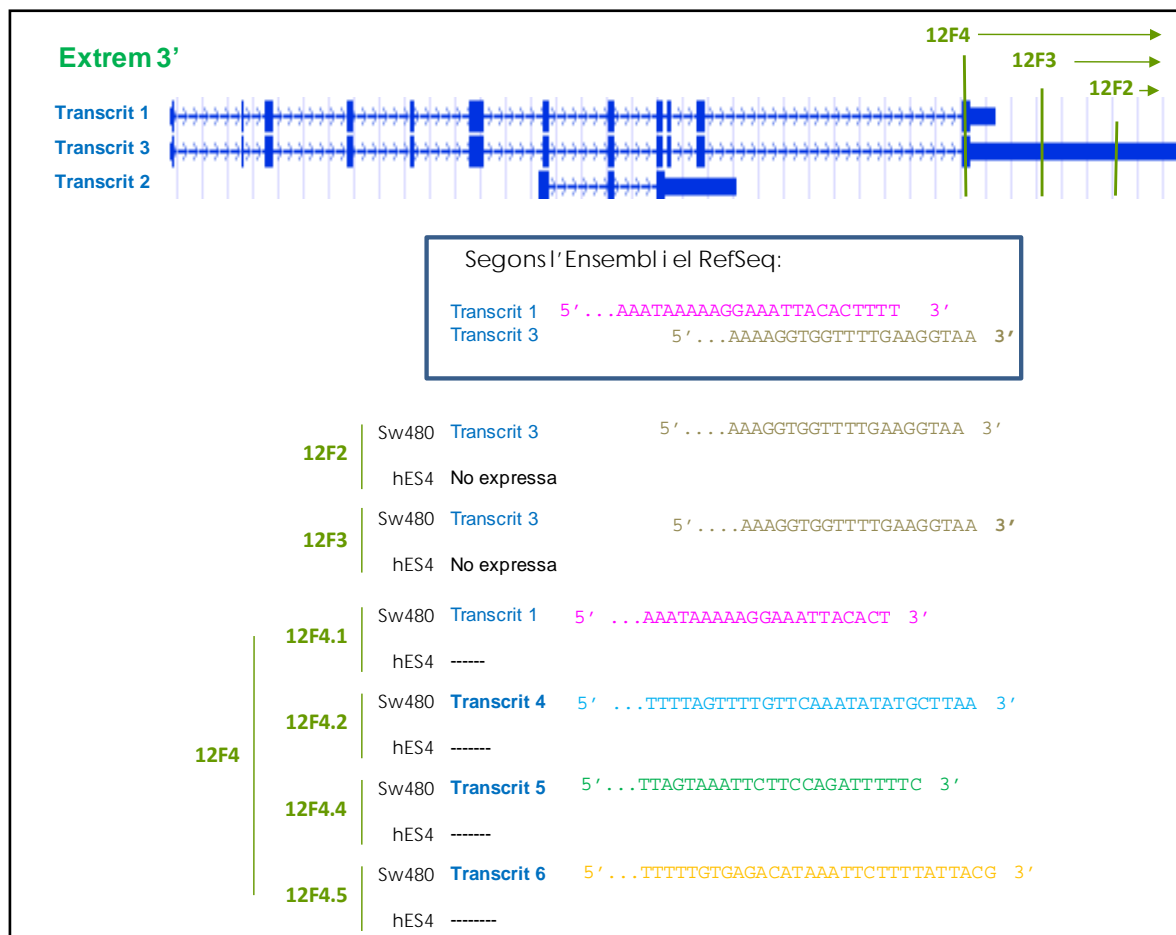


Fig. R.26. UTR3'-RACE. La imatge del Genome Browser mostra la posició dels encebadors. Es mostra també la seqüència de l'UTR3' segons l'Ensembl i el RefSeq i per últim les seqüències dels clons amplificats per PCR corresponents a les bandes obtingudes amb els encebadors 12F2, 12F3 i 12F4. Amb 12F4 analitzem 4 de les 5 bandes obtingudes (12F4.1, 12F4.2, 12F4.4 i 12F4.5). Les mostres analitzades van ser Sw480 i hESC4.

Aquest nous transcrits tindrien el seu final en diferents llocs poli(A) descrits, tal i com s'observa en la figura R.27, on el track de *Poly(A) sites* del *Genome Browser* (versió GRCh37/hg19) ens mostra aquells llocs Poli(A) descrits i registrats en la base de dades polyA\_DB. Aquesta base de dades és una col·lecció de llocs de poliadenilació d'ARNm humà basat en evidències EST/ADNc. Es defineix com a "lloc de poliadenilació" una sola base que indica l'inici de la cua poli(A) en el nou RNAm transcrit i es troba normalment 10-30 nucleòtids *downstream* de la senyal de poliadenilació (majoritàriament AAUAAA).

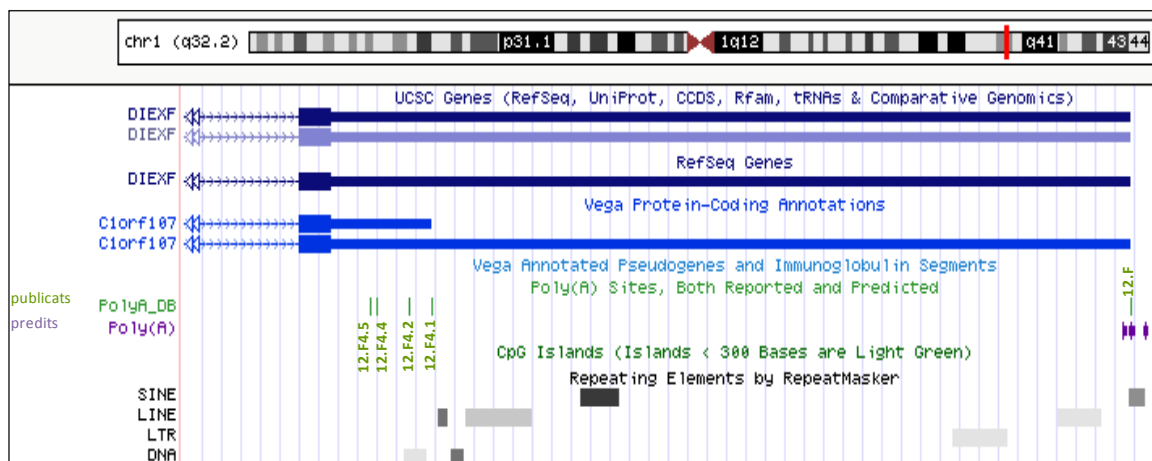


Fig. R.27. Cada línia verda correspon als diferents llocs poly(A) que estan descrits a en la Poly(A)\_DB del *Genome Browser* versió GRCh37/hg19.

Per tant, pel que fa a l'UTR3' veiem que el transcrit 3 s'expressa en la línia Sw480 (que té l'Alu parcialment metilada) i no en hESC4 ni en HFF (línies amb l'Alu Aj2c1 totalment metilada).

El transcrit 1 i els nous transcrits (4, 5 i 6) s'expressen en Sw480 i també en hESC4 i en HFF segons els patrons de bandes de la figura R.25.

En la següent figura veiem l'extrem UTR3' (en lletres blaves) del gen *DIEXF* per esquematitzar tota la informació obtinguda de l'extrem UTR3' (figura R.28).

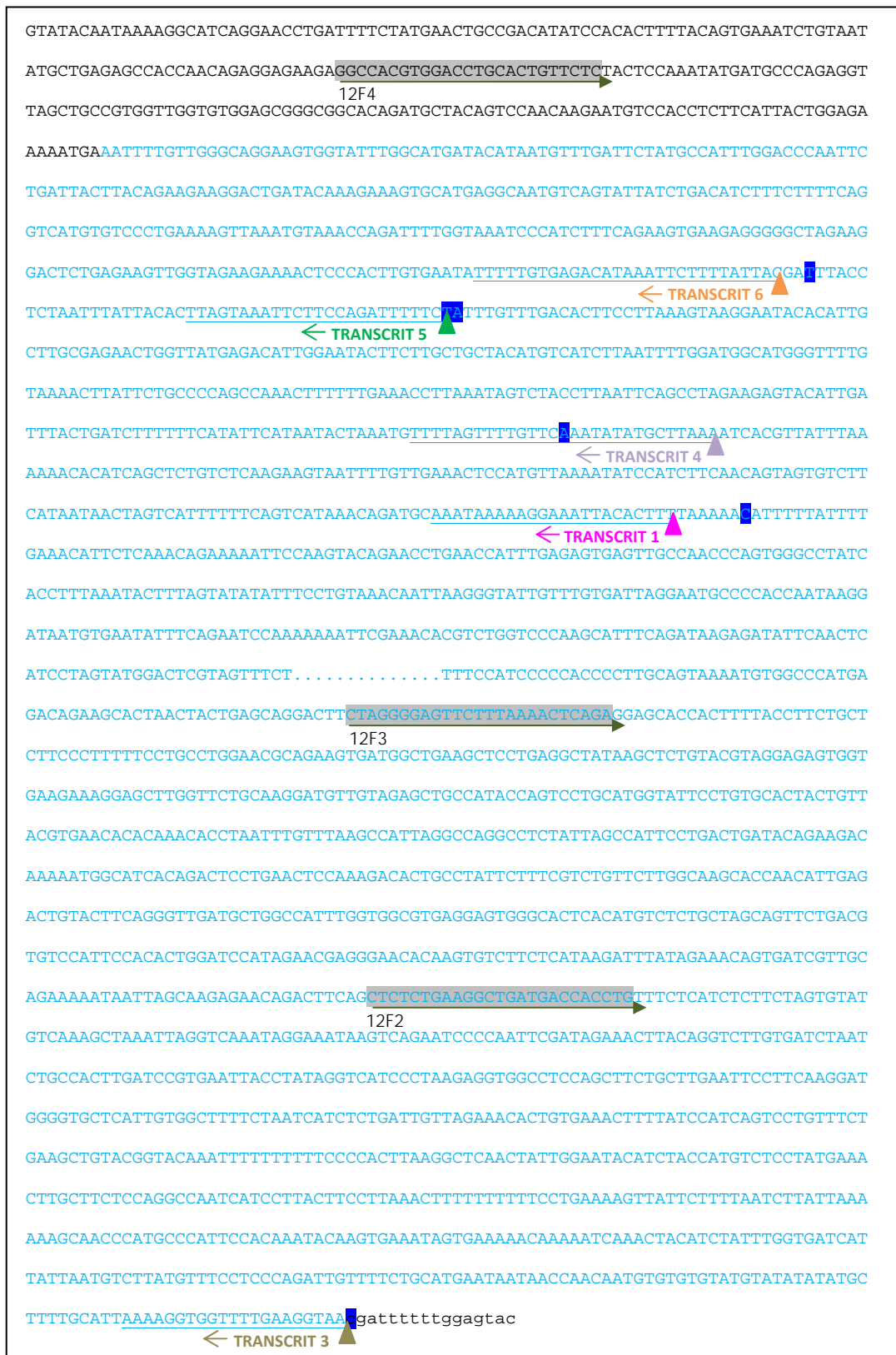


Fig. R.28. Seqüència UTR3' del gen DIEXF. En blau es representa la seqüència UTR3'. Subratllat en gris es representen els encebadors 12F2, 12F3 i 12F4. Resaltat en blau fosc veiem els llocs Poli(A) descrits en la base de dades poly\_DB. El final de cadascun dels transcrits identificats per RACE es representa amb un triangle. Marcats en marró i rosa veiem els transcrits fins ara anotats (transcrit 3 i 1 respectivament). La resta de transcrits identificats també es troben representats: en morat el transcrit 4, verd el transcrit 5 i en taronja el transcrit 6.

També es va fer RACE de l'extrem 5', amplificant 3 fragments (Taula R.10 i Figura R.29).

En amplificar amb 3R observem en Sw480 una doble banda a 450-500pb que podria correspondre als transcrits 3 (382pb) i 1 (460pb)descrits. La resta de línies analitzades només mostren una banda d'uns 500pb que equivaldria al transcrit 1 (460pb). Amplificant amb 7R, així com ocorre amb 3R, i malgrat que el gel no permet apreciar-ho ja que no està prou separat, podria existir una doble banda en Sw480 equivalent als transcrits 3 (1149pb) i 1 (1173pb); en canvi, a HFF i hESC4 veiem una banda menys gruixuda i lleugerament més elevada, indicant-nos que només hi hauria el transcrit 1 (1173pb). Ambdues bandes es van clonar (explicat a continuació en la figura R.30).

L'amplificació amb 9R va donar una única banda d'aproximadament 1700pb corresponent als transcrits 3 i 1 descrits però no detectem la banda a 673pb corresponent al transcrit 2 (dades no mostrades). No obstant no es van clonar ni seqüenciar donant per suposat que no ens aportaria més informació de la ja descrita.

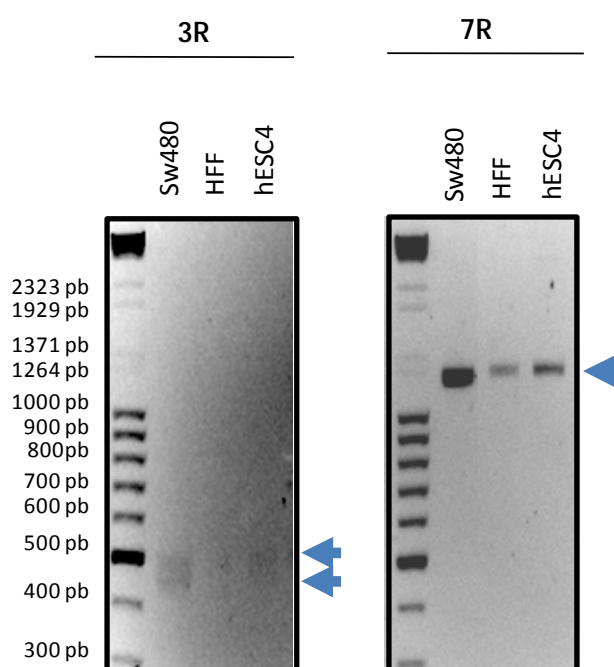


Fig. R.29. 5'-RACE de les línies Sw480, hESC4 i HFF. Resultat de la PCR dels encebadors específics del nostre ADNc d'interès en 5' (3R i 7R). El marcador de pes molecular és BstIII i el DNA ruler de 100pb. Les fletxes blaves ens indiquen els fragments amplificats.

Vam clonar i seqüenciar les bandes 3R i 7R (Figura R.30). Igual que amb la regió UTR3' també es van seqüenciar de 4 a 8 clons i representem aquelles seqüències majoritàries. Els resultats en UTR5' no són tant clars com en 3'. La complexitat de les seqüències obtingudes en 5' potser es deu a l'existència de 6 transcrits diferents, els quals, possiblement, tenen alguna variació en la seqüència corresponent a l'inici de transcripció o TSS (de l'anglès *Transcription Start Site*).

Mirant la banda 3R veiem que al seqüenciar els clons de Sw480 apareixen dues seqüències majoritàries, que al comparar-les amb les seqüències descrites de l'Ensembl o del RefSeq, tot i no ser del tot iguals, suggereixen que una correspondria al transcrit 1 (i possiblement als transcrits 4, 5 i 6) i l'altre al 3. En canvi quan seqüenciem hESC4 només trobem una seqüència majoritària que correspondria als transcrit 1, 4, 5 i 6 (Figura R.30).

Pel que fa a la banda 7R vam tenir problemes a l'hora de seqüenciar els clons de Sw480 fent que no tinguem resultats d'aquesta línia, però en hESC4 veiem que la seqüència que s'obté es correspondria amb els transcrit 1, 4, 5 i 6.

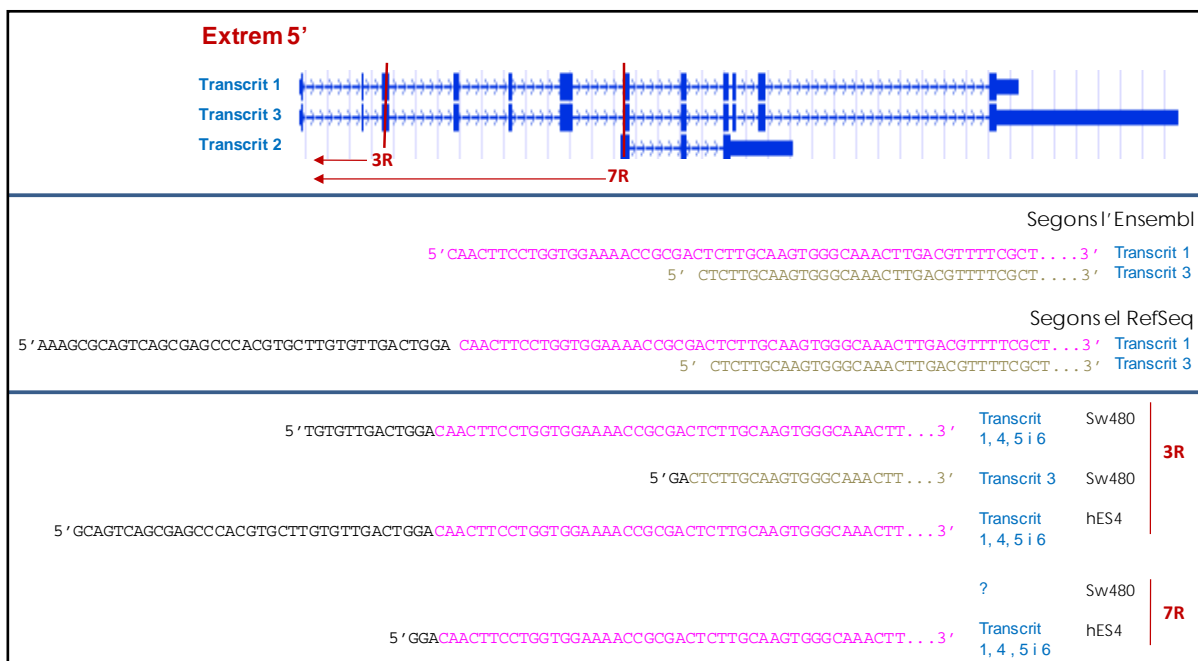


Fig. R.30. 5'-RACE. La imatge del Genome Browser mostra la posició dels encebadors. Es mostra també la seqüència del UTR5' segons l'Ensembl i el RefSeq i per últim les seqüències dels clons amplificats per PCR corresponents a les bandes obtingudes amb els encebadors 3R i 7R. Les mostres analitzades van ser Sw480 i hESC4.

Així doncs, pel que fa al UTR5' veiem que la línia hESC4 només presentaria el transcrit 1, 4, 5 i 6. El mateix passaria amb HFF si ens fixem en la figura R.29. En canvi Sw480 presenta també el transcrit 3.

Així com s'ha fet amb l'extrem 3', es mostra un esquema de l'extrem 5' del gen *DIEXF* (Figura R.31).



Figura R.31. Seqüència UTR5' del gen *DIEXF*. En blau representa la seqüència UTR5'. Subratllat en gris es representen els encebadors 7R i 3R. L'inici de cadascun dels transcrits es representa amb un triangle. Els transcrits 1, 4, 5 i 6 en rosa i el transcrit 3 en marró.

En resum, analitzant els resultats del RACE podem concloure que existeixen nous transcrits del gen *DIEXF* que difereixen en l'extrem UTR3'. A més també veiem que la línia Sw480 presenta els transcrits descrits 1 i 3 així com els nous (4, 5 i 6).

Pel que fa a la línia hESC4 veiem que presenta el transcrit 1 així com els nous transcrits descrits però no el transcrit 3. Sense confirmar per seqüenciació però amb un patró de bandes similar podríem dir que HFF també presentaria tots els transcrits excepte el 3.

Per tant, en analitzar l'expressió a nivell d'ARN per tres tècniques diferents (Taula R.11): RT-PCR a temps real (RT-qPCR), Northern Blot (NB) i RACE veiem un alt grau de correlació amb totes les tècniques confirmant l'expressió gènica (en verd) o manca d'aquesta (en vermell). En el transcrit 3 trobem una incongruència en HFF i queratinòcits ja que per RT-PCR trobàvem un increment en l'expressió que no s'ha vist en cap de les dues tècniques posteriors. Pel que fa al transcrit 2 no hem trobat cap evidència en les nostres mostres.

Fixant-nos en la correlació de l'expressió del gen *DIEXF* i l'estat de metilació d'Alu Aj2c1 semblaria que la metilació regularia d'expressió del transcrit 3.

TAULA R.11. EXPRESSIÓ DELS TRANSCRITS DEL GEN DIEXF

		estat metilació Aj2c1		EXPRESSIÓ DIEXF		
		Aj2c1 5'	Aj2c1 3'	TRANSCRIT 3		
				RT-qPCR	NB	RACE
	Cèl·lules mare embrionàries	■	■	■		
	cèl·lules pluripotents induïdes			■ ND		
	fibroblasts	■	■	■	■	
	queratinocits			■ ND		
CCR	Sw480	■	■	■		
	LoVo			■	■ ND	■ ND
	HCT116			■	■	■ ND
	DKO			■ ND	■	■ ND

		estat metilació Aj2c1		EXPRESSIÓ DIEXF		
		Aj2c1 5'	Aj2c1 3'	TRANSCRIT 1, 4, 5 i 6		
				RT-qPCR	NB	RACE
	Cèl·lules mare embrionàries	■	■	■		
	cèl·lules pluripotents induïdes			■ ND		
	fibroblasts	■	■	■		
	queratinocits			■ ND		
CCR	Sw480	■	■	■		
	LoVo			■	■ ND	■ ND
	HCT116			■	■	■ ND
	DKO			■ ND	■	■ ND



### 3.1.4. ESTUDI DE L'EXPRESSIONI DE *DIEXF* A NIVELL DE PROTEÏNA

Per tal d'esbrinar si els canvis d'expressió a nivell d'ARN es tradueixen també en canvis en els nivells de proteïna es va analitzar per Western Blot l'expressió de la proteïna *DIEXF* en diferents línies cel·lulars. Primerament es va utilitzar l'anticòs Anti-C1orf107 (Sigma), però degut a la seva inespecificitat, es va utilitzar l'anticòs Anti-*DIEXF* (Abcam). Els detalls de la tècnica es poden trobar en l'apartat 2.4. de la secció material i mètodes.

Segons *l'Ensembl/Havana* els transcrits 1 i 3 generen una proteïna de 756 residus i el transcrit 2 de 248 residus (Figura R.32). L'anticòs Anti-*DIEXF* està dirigit contra un pèptid sintètic de la regió N terminal que reconeix únicament el producte de 756 residus (89kDa), corresponent als transcrits 1 i 3 descrits.

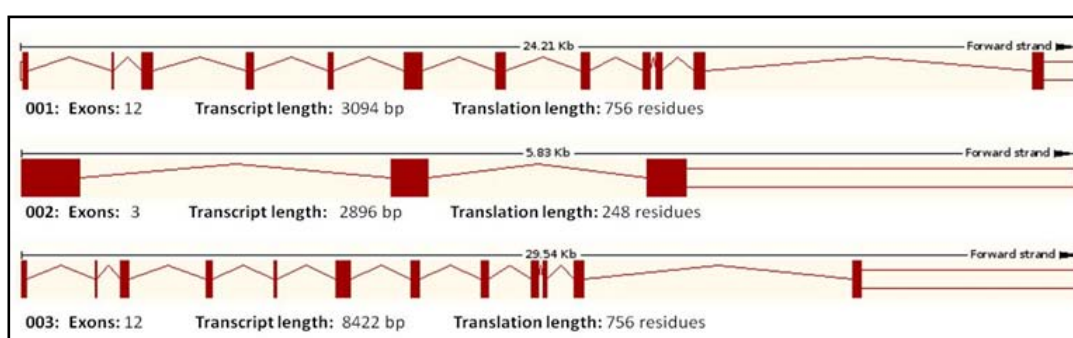


Fig. R.32. Imatge de *l'Ensembl* de les tres isoformes de *DIEXF*.

Es va analitzar l'expressió de *DIEXF* en les diferents línies de CCR (HCT116-DKO, HCT116, HT29, LoVo i Sw480), la línia provinent de cèl·lules mare humanes (hESC4) i els fibroblasts (HFF) (Figura R.33).

Podem veure que hi ha una correspondència entre l'expressió de la proteïna *DIEXF* i l'estat de metilació de l'element *Alu Aj2c1*. Així doncs, a les línies on l'element *Alu* està totalment metilat (hESC4 i HFF) no veiem expressió de la proteïna, en canvi les línies més desmetilades (HCT116-DKO, HCT116, HT29 i LoVo) expressen la proteïna. En el cas de la línia Sw480, que està parcialment metilada, veiem una reducció de l'expressió proteica. La mateixa correlació existeix a nivell d'ARNm. En el Western Blot observem tres bandes que varien d'intensitat depenent de la mostra estudiada, d'aquestes 3 només una es troba present en totes les mostres desmetilades, en canvi la banda superior només apareix en HCT116 i Sw480 i la banda inferior en HCT116-DKO i LoVo. Aquestes bandes podrien correspondre a diferents isoformes o modificacions postraduccionals de la mateixa proteïna, tot i que no podem descartar que siguin inespecífiques.

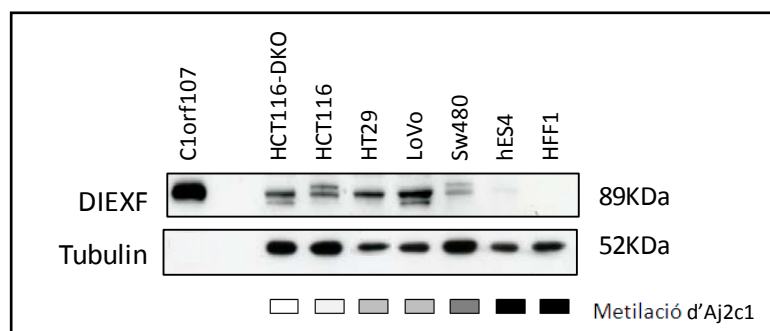


Fig. R.33. Western Blot de les línies de CCR (DKO, HCT116, HT29, LoVo i Sw480), de les cèl·lules mare humanes (hESC4) i la línia de fibroblast (HFF). La proteïna d'interès es detecta amb l'anticòs policlonal Anti-DIEXF (Abcam). Com a control de càrrega utilitzem l'anticòs monoclonal Anti-Tubulin i com a control positiu de l'anticòs Anti-DIEXF utilitzem una proteïna recombinant (DIEXF, també dita C1orf107). Els quadrats inferiors mostren l'estat de metilació global de l'Alu Aj2c1 de les mostres, essent negre 100% metilat i blanc 0% metilat.

En un darrer estudi per veure l'especificitat de l'anticòs i com actua cada transcrit a l'hora de traduir-se a proteïna és va silenciar el gen. Es van provar 4 siRNAs diferents situats al llarg de la seqüència del gen. Els oligos es van triar segons la posició dins del gen i seguint les recomanacions del programa *Block-IT* (Invitrogen). Es van dissenyar dos siRNA situats en la regió ORF (*Open Reading Frame*) i dos en l'extrem 3' en la regió UTR3' (*Untranslated Region*). El disseny dels siRNA es troba explicat en l'apartat 2.5. de la secció de material i mètodes) (Taula R.12).

TAULA R.12. siRNAs DE SILENCIAMENT I LA SEVA POSICIÓ DINS DEL GEN

siRNA	Posició	Nucleòtid d'inici
siRNA1	UTR3' (específic del transcrit 3)	5057
siRNA2	UTR3' (específic del transcrit 3)	7118
siRNA3	ORF (comú a les 3 isoformes)	840
siRNA4	ORF (comú a les 3 isoformes)	2299
siRNAuniv	siRNA comercial	

Cadascun dels siRNAs es va provar en tres temps diferents (24, 48 i 72 hores) així com a diferents concentracions (10  $\mu$ M -20 pmols-, 20  $\mu$ M -40 pmols- i 50  $\mu$ M -100 pmols-), sent la més eficient a 20  $\mu$ M -50 pmols-. Com a controls es va utilitzar la línia cel·lular sense transfectar, la línia cel·lular transfectada amb lipofectamina i la línia cel·lular transfectada amb un siRNA control universal. Els detalls de la tècnica es poden trobar en l'apartat 2.5. de la secció material i mètodes.

Primerament es van utilitzar els siRNA3 i siRNA4 (ambdós corresponents a la regió ORF comú a les tres isoformes) i es va analitzar l'expressió mitjançant RT-qPCR en la línia HCT116. Cal recordar que l'amplicó 8-9 cobreix tots els transcrits, el 12-UTR003 tots els

transcrits excepte el 2, i el UTR003 el transcrit 3. Com s'observa en la figura R.34 els dos siRNAs silencien l'expressió de totes les isoformes de *DIEXF* i aquest silenciament perdura al llarg del temps d'estudi. No obstant, el silenciament del transcrit 3 és menor (Figura R.34).

## HCT116

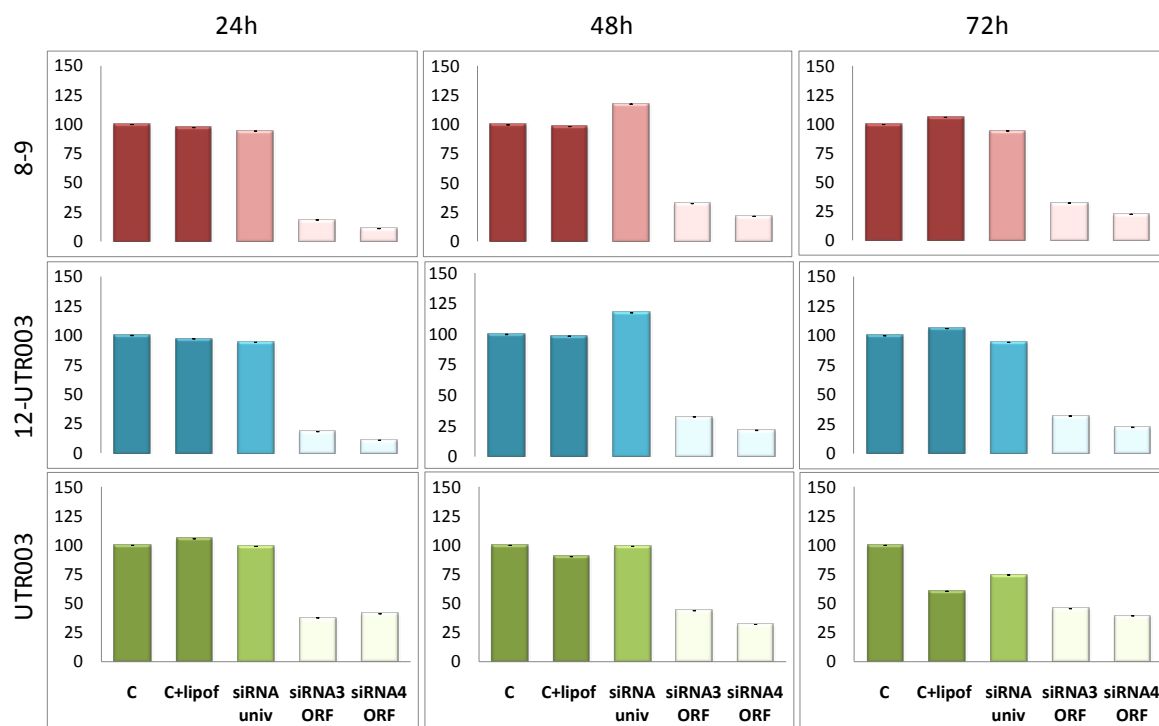


Fig. R.34. Silenciament de *DIEXF* en la línia HCT116 utilitzant siRNA3 i siRNA4 a una concentració de 20 µM. Efecte del silenciament a 24, 48 i 72 hores. L'eix de les Y representa l'expressió relativa, considerant la mostra HCT116 control com al 100% d'expressió. Es miren les regions de gen: 8-9, 12-UTR003 i UTR003 per veure l'efecte per RT-qPCR.

Per Western Blot també confirmem que els diferents siRNAs disminueixen els nivells de proteïna a les 48 hores (tot i que no es mostren les dades, els resultats a 24 i 72 hores són iguals) (Figura R.35). Pel que fa la doble banda veiem que el silenciament amb els siRNAs fa que la banda inferior desapareixi pràcticament en la seva totalitat mentre que la banda superior és manté, el que suggeriria que la segona banda és la corresponent a la proteïna d'interès.

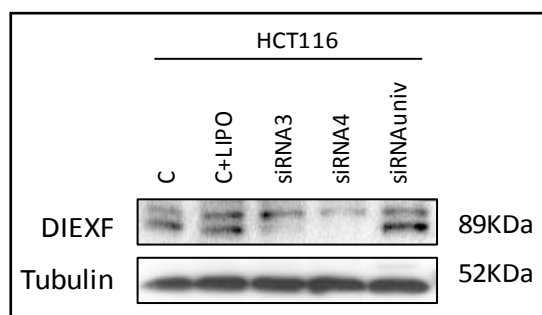


Fig. R.35. Western Blot de la línia HCT116 silenciada amb siRNA3 i siRNA4 a 48 hores. La proteïna d'interès es detecta amb l'anticòs policlonal Anti-*DIEXF* (Abcam). Com a control del silenciament utilitzem la línia control sense tractament, amb lipofectamina i amb el siRNA universal. Com a control de càrrega utilitzem l'anticòs monoclonal Anti-Tubulin.

L'estudi es va repetir en la línia Sw480. En aquesta línia només es va provar el siRNA3 (situat en la regió ORF comú a les 3 isoformes) confirmant de nou l'eficiència del silenciament tant a nivell d'expressió del gen com a nivell de proteïna a les 24, 48 i 72 hores (Figura R.36 i R.37).

### Sw480

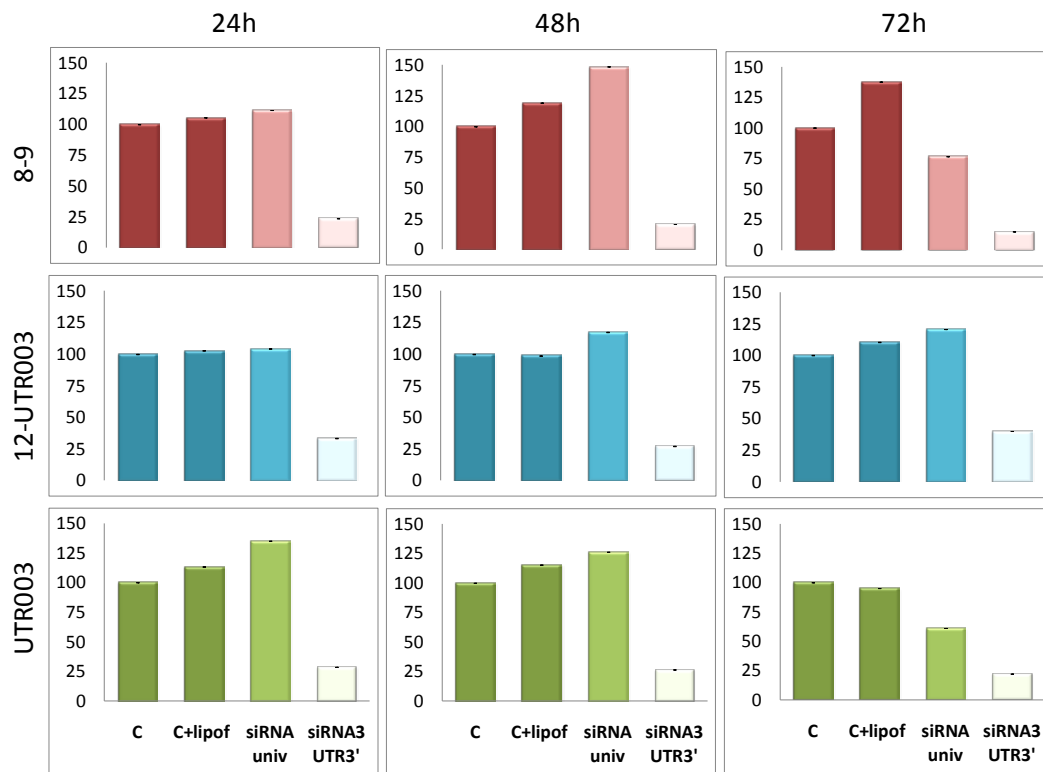


Fig. R.36. Silenciament de DIXF en la línia Sw480 utilitzant siRNA3 a una concentració de 20 µM. Efecte del silenciament a 24, 48 i 72 hores. L'eix de les Y representa l'expressió relativa, considerant la mostra HCT116 control com al 100% d'expressió. Es miren les regions de gen: 8-9, 12-UTR003 i UTR003 per veure l'efecte per RT-qPCR.

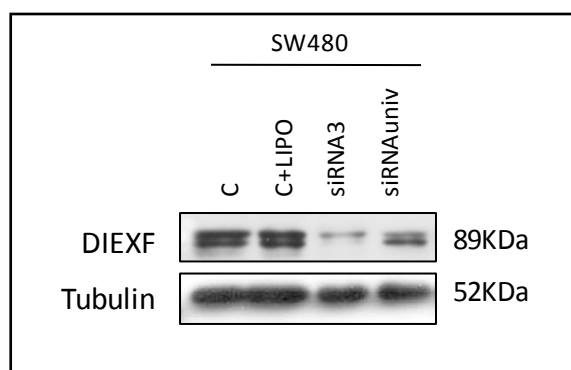


Fig. R.37. Western Blot de la línia Sw480 silenciada amb siRNA3 a 48 hores. La proteïna d'interès es detecta amb l'anticòs policlonal Anti-DIXF (Abcam). Com a control del silenciament utilitzem la línia control sense tractament, amb lipofectamina i amb el siRNA universal. Com a control de càrrega utilitzem l'anticòs monoclonal Anti-Tubulin.

També es va estudiar l'expressió per veure l'efecte de siRNA1 i siRNA2. Recordem que aquests oligos estaven situats en l'UTR3' de la isoforma de major mida, el transcrit 3. Per tant, tot i silenciar un dels transcrits, els altres es continuen expressant, i per això utilitzant encebadors que cobreixen tots els transcrits (8-9) així com el que cobreix el transcrit 1 i 3 (12-UTR003) veiem una lleugera baixada però encara detectem expressió a nivell d'ARNm. Si observem els resultats de la RT-qPCR específica pel transcrit 3 (UTR003) la baixada d'expressió és major (Figura R.38).

HCT116

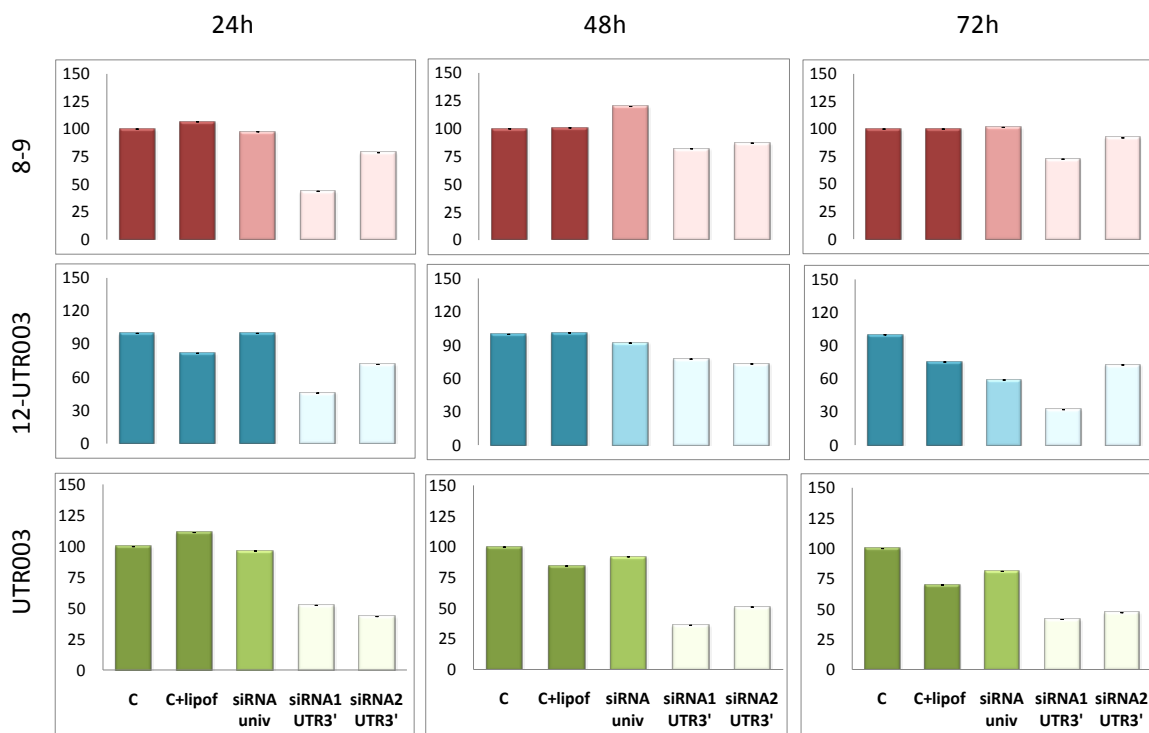


Fig. R.38. Silenciament de DIEXF en la línia HCT116 utilitzant siRNA1 i siRNA2 a una concentració de 20 µM. Efecte del silenciament a 24, 48 i 72 hores. L'eix de les Y representa l'expressió relativa, considerant la mostra HCT116 control com al 100% d'expressió. Es miren les regions de gen: 8-9 i UTR003 per veure l'efecte per RT-qPCR.

A nivell de proteïna veiem la pèrdua de DIEXF al silenciar amb siRNA1, fet que torna a posar de manifest que el transcrit 3 és el que es tradueix a proteïna (Figura R.39).

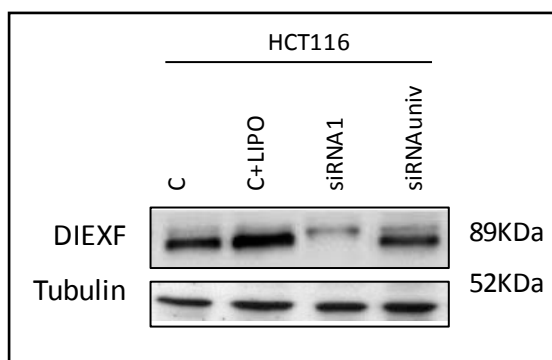
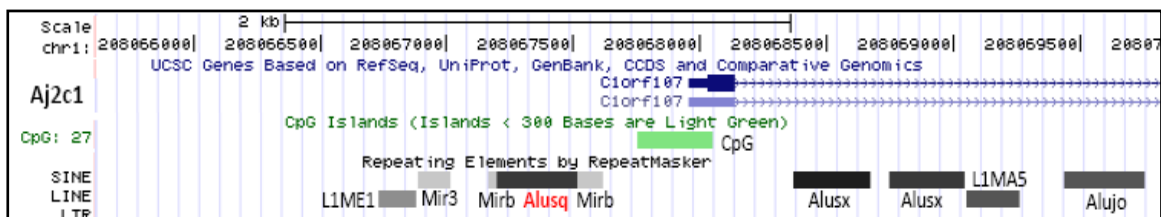


Fig. R.39. Western Blot de la línia HCT116 silenciada amb siRNA1 a 48 hores. La proteïna d'interès es detecta amb l'anticòs policlonal Anti-DIEXF (Abcam). Com a control del silenciament utilitzem la línia control sense tractament, amb lipofectamina i amb el siRNA universal. Com a control de càrrega utilitzem l'anticòs monoclonal Anti-Tubulin

En resum, veiem que podem establir una associació entre l'estat epigenètic de la regió i l'expressió del gen *DIEXF*, sent l'element Alu Aj2c1 el que denota canvis en l'estat de metilació i, per tant, relacionant-se amb la regulació del gen. *DIEXF* té diversos transcrits descrits i hem trobat, mitjançant la tècnica RACE, nous transcrits, el 4, 5 i 6.

Segons els resultats obtinguts tot indica que el transcrit 3 seria l'únic transcrit codificant i és precisament l'expressió d'aquest la que es correlaciona amb l'estat epigenètic de l'element Alu d'estudi Aj2c1 (Taula R.13).

TAULA R.13. RELACIÓ ENTRE L'ESTAT EPIGENÈTIC DE L'ELEMENT ALU AJ2C1 I L'EXPRESSIÓ DE *DIEXF*



		estat metilació <sup>a</sup>			marca d'histona H3K4me3 <sup>b</sup>			EXPRESSIÓ <i>DIEXF</i> <sup>c</sup>				
								TRANSCRITS ARN		PROTEÏNA		
		5'	Aj2c1	3'	5'	Aj2c1	3'	1/4/5/6	3			
Cèl·lules mare embrionàries		[Metilació]			∅	∅	∅	↑	[Verd]	[Vermell]	[Vermell]	
	Cèl·lules pluripotents induïdes		[Metilació]			ND			[Verd]	[Vermell]	ND	
fibroblasts		[Metilació]			∅	∅	∅	↑	[Verd]	[Vermell]	[Vermell]	
	queratinocits		[Metilació]			ND			[Verd]	[Vermell]	ND	
CCR	Sw480		[Metilació]			∅	∅	∅	↑	[Verd]	[Verd]	[Verd]
	LoVo		[Metilació]			∅	∅	∅	↑	[Verd]	[Verd]	[Verd]
	HCT116		[Metilació]			∅	↑	↑	↑	[Verd]	[Verd]	[Verd]
	DKO		[Metilació]			∅	↑	↑	↑	[Verd]	[Verd]	[Verd]

a) Estat de metilació de la regió. La regió 5' i 3' de l'element Aj2c1 i la regió flanquejant a Aj2c1 tant en l'extrem 5' com 3'. Desmetilat indicat en blanc, metilat en negre i en escala de grisos parcialment desmetilat.

b) Marca d'histona H3K4me3 de la regió. La regió 5' i 3' de l'element Aj2c1 i la regió flanquejant a Aj2c1 tant en l'extrem 5' com 3'. ∅ indica que no hi ha marca i ↑ indica presència de marca activa, ND indica que no s'ha analitzat

c) Expressió *C1orf107*. L'expressió a nivell de ARN dels diferents transcrits i l'expressió a nivell de proteïna. En vermell no expressió i en verd expressió, ND indica que no s'ha analitzat.

### 3.1.5. PAPER DE L'ELEMENT AJ2C1 EN CÀNCER

Per veure com es troba la nostra regió en diferents càncers es van consultar les dades del TCGA (*The Cancer Genome Atlas*) (<http://cancergenome.nih.gov/>). TCGA ha recollit i caracteritzat més de 20 tipus tumorals. Entre els anàlisis realitzats han mirat l'estat específic de metilació de dinucleòtids CpG mitjançant les plataformes *Illumina Infinium Human Methylation 450K Bead Array platform* en diverses parelles Normal-Tumor.

En un primer estudi de la regió d'interès es va analitzar el *track CpG Methylation by Methyl 450K Bead Arrays from ENCODE/HAIB* del *Genome Browser* (Figura R.40). Aquesta opció ens permet veure l'estat específic de metilació dels dinucleòtids CpG interrogats per la plataforma *Illumina Infinium Human Methylation 450 Bead Array platform* en diverses línies cel·lulars.

En la figura R.40 veiem les diverses sondes situades al llarg de la regió d'estudi i en groc es troben marcades aquelles en les que ens hem centrat: sonda cg3753785 (agafa la CpG 2 del nostre amplicó en l'extrem 5' de l'Alu Aj2c1), la sonda cg05968010 (agafa la CpG 16 del nostre amplicó en l'extrem 3' de l'Alu Aj2c1) i la sonda cg05505912 (agafa la CpG 3 de l'illa CpG).



Fig. R.40. Mapa del *Genome Browser* amb la localització de les sondes d'estudi segons el *track CpG Methylation by Methyl 450K Bead Arrays from ENCODE/HAIB*. Marcades en groc es troben les sondes en les que hem aprofundit. A sota estat de metilació de la regió de les diferents línies cel·lulars estudiades i analitzades per seqüenciació de bisulfit.

Si mirem diverses línies que surten en el *track* del Genome Browser (hESC, CaCo2 i HCT116) (Figura R.40) veiem que l'estat de metilació es correspon amb el que nosaltres havíem vist. Així, en hESC veiem que l'element Alu Aj2c1 està metilat i l'illa CpG desmetilada; en CaCo2 l'extrem 5' de l'Alu està metilat (l'extrem 3' no l'han pogut analitzar) i l'illa CpG està desmetilada; i en HCT116 l'extrem 5' d' Aj2c1 està metilat i el 3' parcialment desmetilat i l'illa CpG desmetilada.

Aquestes mateixes sondes es van consultar en les dades del TCGA, concretament en 10 càncers, i es van identificar perfils de comportament diferents segons el tipus de teixit d'origen i tumor (Figura R.41).

En teixit normal de còlon veiem que l'extrem 5' de l'Alu està metilat mentre que l'extrem 3' i l'illa CpG estan desmetilats. En canvi, en els adenocarcinomes de còlon l'extrem 5' de l' Aj2c1 es manté metilat en uns casos mentre que en altres s'hipometila i la resta de la regió roman desmetilada, com el en teixit normal. Aquests resultats es corresponen als que havíem observat nosaltres en les 10 parelles N-T.

En mama, pròstata i pulmó veiem que l'extrem 5' l' Aj2c1 està metilat i el 3' presenta alguns casos metilats i altres desmetilats, l'illa CpG està desmetilada. En canvi, els carcinomes corresponents, hi ha una població de tumors que l'Alu es manté l'extrem 5' metilat mentre que en una altra població dels tumors s'hipometilen totalment, i en tots els casos i l'illa CpG roman desmetilada.

Uns altres tipus de càncer són els carcinomes d'úter, glioblastomes i carcinomes de ronyó. En el teixit normal l'Alu està metilada i l'illa CpG desmetilada, i en els tumors veiem que l'extrem 5' de l'Alu es manté metilat mentre que l'extrem 3' s'hipometila en una proporció important de tumors, i l'illa CpG roman desmetilada.

El cas del glioma cerebral de baix grau, sembla mostrar un comportament completament diferent ja que hi ha una hipermetilació tant de l' Aj2c1 com de l'illa CpG. Aquest perfil cal mirar-lo amb precaució donat que només es tenen 2 teixits normals.

Altres parelles N-T consultades mostren un patró heterogeni, és el cas del carcinoma de tiroides així com varis adenocarcinomes (pròstata, estómac, pulmó, pàncrees i recte).



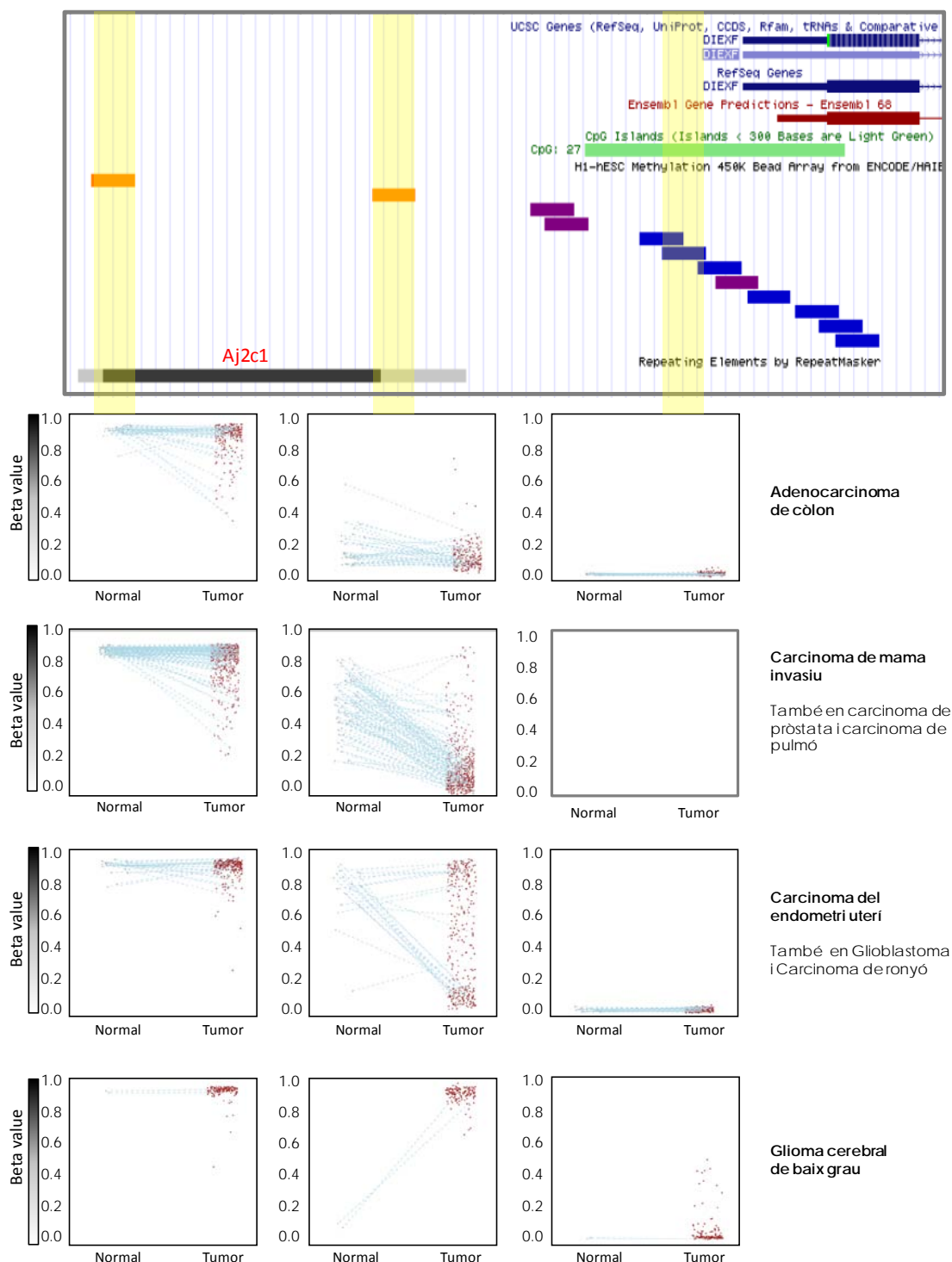


Fig. R.41. Estat de metilació de les 3 regions en diferents tipus de càncer segons les dades del TCGA. Cada gràfic representa l'estat de metilació de diverses mostres normal tumor segons la sonda analitzada (marcat en groc). L'eix de les X respresenta les mostres normals (a l'esquerra de la figura) i tumorals (a la dreta de la figura) i l'eix de les Y el beta value de l'estat de metilació, essent 1 totalment metil·lat i 0 totalment desmetil·lat.

En la figura R.42. presentem un breu esquema dels perfils de comportament segons el tipus de càncer. Dins dels 4 grups de la classificació veiem dos tipus ben definits segons

l'estat de metilació. El grup 1, 2 i 3 formaria part del primer tipus, majoritari en els diferents càncers estudiats: l'element Alu s'hipometila en tumors i l'illa CpG roman desmetilada.

El grup 4, formaria part del segon tipus: hipermetilació en tumor tant de l'element Alu com de l'illa CpG.

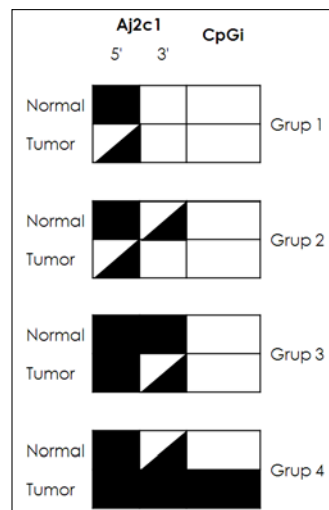


Fig. R.42. Esquema dels diferents grups N-T segons l'estat de metilació de les sondes consultades en el TCGA. Els quadrats en negre representen regions amb CpG metilades i els quadrats blancs regions amb CpG desmetilades.

Per entendre millor la regulació de l'expressió del gen *DIEXF* hem consultat diferents bases de dades d'expressió, entre elles l'*Oncomine* (<https://www.oncomine.org/>). *Oncomine* conté 65 bases de dades d'expressió gènica, cobrint prop de 48 milions de gens en més de 4700 experiments de *microarrays*.

També hem mirat diverses sondes que cobreixen el nostre gen, tant d'*Affymetrix* (*HG-U133 Plus 2.0*) com d'*Agilent* (*Whole human genome 44k*) (Figura R.43).

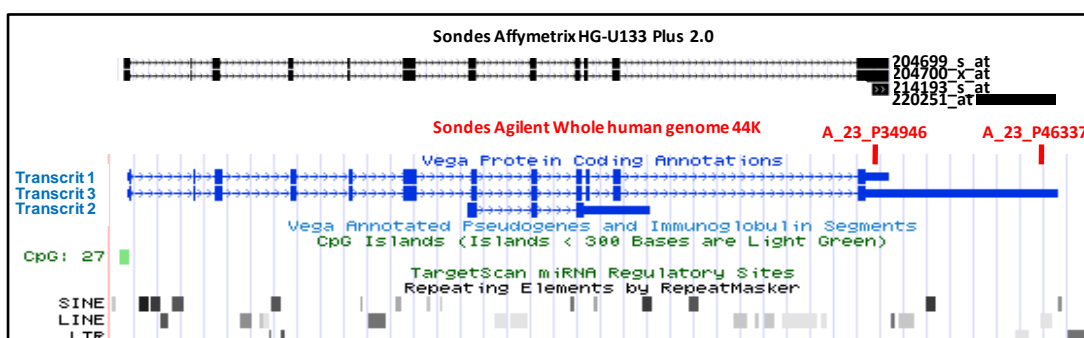


Fig. R.43. Posició de les sondes específiques per *DIEXF*, tant en l'array d'*Affymetrix* *HG-U133 Plus 2.0* com d'*Agilent* *Whole human genome 44k*.

Veiem una sobreexpressió del gen *DIEXF* en aquells càncers que correspondrien al grup 1, 2 i 3, sent major en el grup 1. Del grup 4 no hem trobat cap cas que ens confirmi una infraexpressió de *DIEXF*.

Normalment aquesta sobreexpressió es mostra tant en la sonda específica pels transcrits 1, 3, 4, 5 i 6 com per les específiques del transcrit 3. Tot i que en certs casos la sobreexpressió és específica del transcrit 3. En la figura R.44 mostrem els resultats de càncer colorectal obtinguts de l'*Oncomine*, amb mostres del TCGA utilitzant les sondes d'*Agilent*.

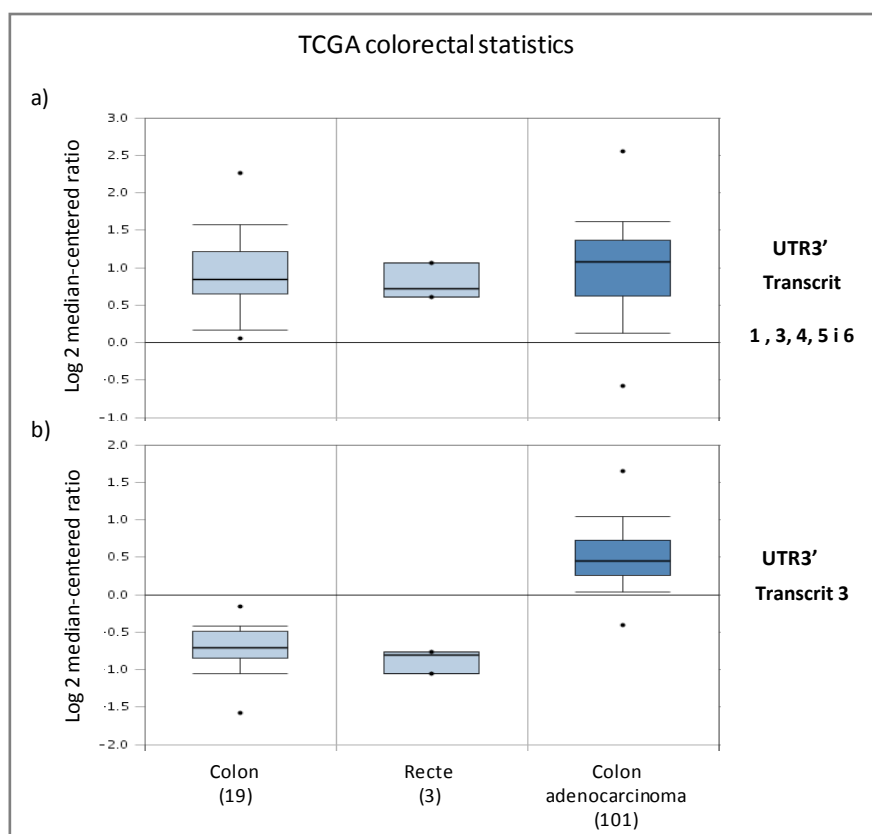


Fig. R.44. Estadística de l'expressió d'ARNm de càncer colorectal obtinguts de l'*Oncomine*, amb 237 mostres del TCGA utilitzant les sondes d'*Agilent* Whole human 44K. Es mostren els resultats de 19 casos de còlon, 3 de recte i 101 d'adenocarcinoma de còlon. a) Resultats amb la sonda A\_23\_P34946, que hibrida amb l'UTR3' del transcrits 1, 4, 5, 6 i 3, b) Resultats amb la sonda A\_23\_P46337, que hibrida amb l'UTR3' específic del transcrit 3.

La informació global trobada en totes aquestes bases de dades tant a nivell d'expressió com de metilació en mostres normals i tumorals ens dona suport a la, la relació entre l'estat de metilació de l'element Alu Aj2c1 i l'expressió del gen *DIEXF*.

### 3.2. ALU ZALU3

L'element ZALU3 es va seleccionar per estudiar-lo en detall en base al seu estat desmetilat en les cèl·lules mare embrionàries així com en les cèl·lules pluripotents induïdes però no en la majoria de les cèl·lules diferenciades analitzades, suggerint la seva possible vinculació en la diferenciació. ZALU3 es tracta d'un Alu Sc, situat a -1064pb de l'inici de transcripció del gen *GLDC* (*glycine dehydrogenase*), un gen relacionat amb el metabolisme de la glicina (Zhang et al., 2012).

Paral·lelament, en el laboratori es va dur a terme un estudi d'expressió gènica per mitjà d'arrays d'expressió (*SurePrint G3 Human Gene Expression 8x60k*). Aquests resultats ens varen mostrar que el gen *GLDC* proper a l'Alu d'estudi només s'expressava en les cèl·lules que tenien l'Alu desmetilada.

#### 3.2.1. ESTUDI EPIGENÈTIC DE L'ELEMENT ZALU3 I LA REGIÓ FLANQUEJANT

Primerament vam voler validar els resultats obtinguts per NSUMA i ampliar l'estudi a tota la regió de l'element ZALU3. Es van analitzar cèl·lules mare embrionàries (hESC4 i hESC2), cèl·lules pluripotents induïdes (KIPS4F8), fibroblasts (F, HFF) queratinòcits (K18 i K37), i una parella normal-tumor de còlon (377N i 377T) i diferents línies cel·lulars de CCR (CaCo2, HCT116-DKO, HCT116, LoVo, Sw480, HT29). En l'annex II figura AII.5 trobem representat l'estat de metilació de cada dinucleòtid CpG dels elements Alu en base als anàlisis de seqüenciació després del tractament amb bisulfit sòdic.

Pel que fa a les cèl·lules mare embrionàries i pluripotents induïdes veiem que tota la regió, incloent la ZALU3, està desmetilada. En canvi en les cèl·lules diferenciades normals veiem tota la regió desmetilada fins arribar a l'element Alu ZALU3, que es troba metilat (Figura R.45).

En les parelles de còlon normal-tumor, veiem que la ZALU3 està molt metilada en ambdós casos i en tumors l'estat de metilació augmenta en la regió, metilant-se també l'AluJb i part de la CpGi (només es mostra una parella, però l'estudi s'ha fet en 10 parelles Normal-Tumor) (Annex II figura AII.5).

Finalment, en les línies cel·lulars tumorals veiem que quan l'Alu ZALU3 està parcialment o totalment metilada, també ho està la resta de la regió incloent l'illa CpG, mentre que quan l'Alu està desmetilada també ho està la regió.

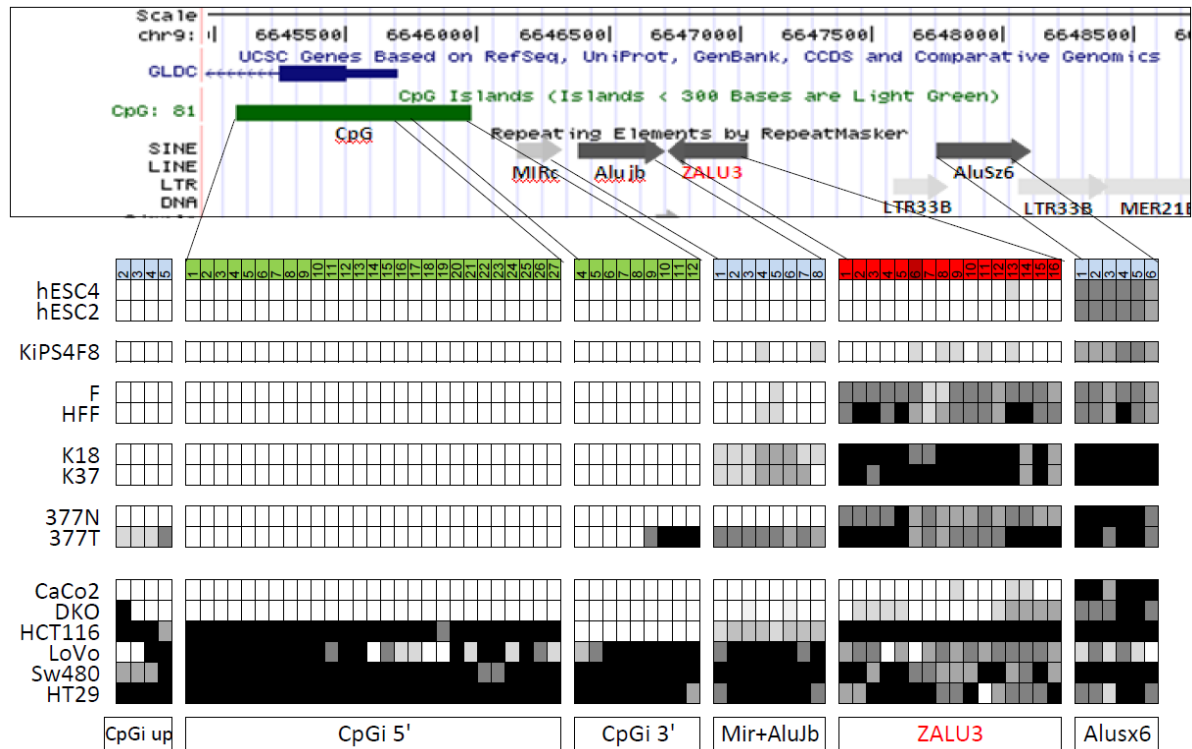


Fig. R.45. Estat de metilació de la regió flanquejant a l'element Alu ZALU3 seleccionat (en vermell) cèl·lules mare embrionàries (hESC4 i hESC2), cèl·lules pluripotents induïdes (KiPS4F8), fibroblasts (F, HFF), queratinòcits (K18 i K37), una parella normal-tumor (377N i 377T) i en diferents línies de CCR. Regió obtinguda del Genome Browser versió Feb. 2009 (GRCh37/hg19). Cada requadre indica una CpG. En vermell trobem l'Alu ZALU3, en verd l'illa CpG i en blau diferents elements repetitius. En blanc de 0-25% de metilació, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre 75-100% de metilació.

A continuació es van analitzar les marques d'histona en diferents línies de CCR (Sw480 HCT116, DKO i CaCo2) que mostraven un patró de metilació divers (Figura R.46). Es van mirar diverses marques d'histona i mostrem, com en els altres casos, les marques actives de Ach3K9, H3K4me3, i l'H3 total. La localització dels encebadors de la regió es troben en l'annex III AIII.2.

Analitzant les marques actives veiem que aquestes només són presents quan tota la regió es troba desmetilada. HCT116 presenta nivells baixos de marca activa en aquells fragments de la regió que es troben parcialment desmetilats. En DKO la presència de marca activa augmenta respecte a HCT116, corresponent-se a la desmetilació que pateix la regió. En el cas de CaCo2, totalment desmetilat en tota la regió, veiem una forta presència de marca activa.

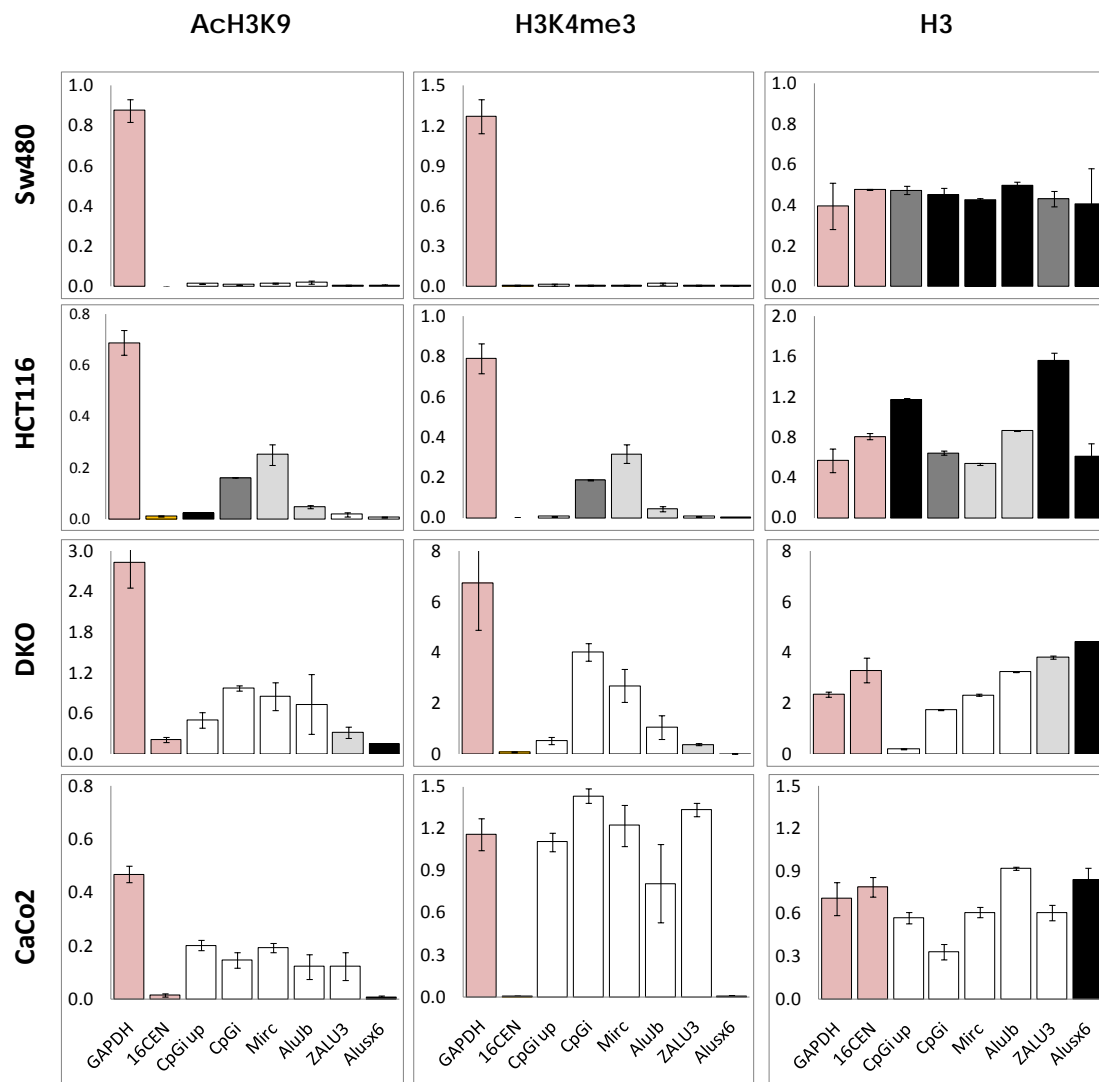


Fig. R.46. Anàlisi per CHIP de les regions flanquejants a l'element Alu ZALU3 en les línies Sw480, HCT116, DKO i CaCo2. Es representen les marques d'activació Ach3K9, H3K4me3 i l'H3 total. L'eix de les Y representa el ràtio de la marca respecte l'H3, a excepció d'H3 que és el valor respecte l'input. Els amplicons representats en les columnes mostren el seu estat de metilació sent blanc desmetilat, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre >75% metilat. Per altra banda en vermell es representen els controls, sent GAPDH control positiu de marques actives i 16CEN control negatiu.

Finalment vam analitzar els fibroblasts i les cèl·lules mare embrionàries (figura R.47). Veiem que les hESC2, totalment desmetilades en la regió, presenten molta marca activa així com una mica de marca inactiva mostrant la bivalència que solen presentar les cèl·lules mare embrionàries. En canvi, els HFF que estaven desmetilats en gairebé tota la regió a excepció de ZALU3 i Alusx6, no presenten marques actives en cap fragment de la regió i en canvi sí presenten marques inactives. Aquest fet resulta interessant i caldrà veure com afecta l'estat de la cromatina a l'expressió del gen proper a la regió, *GLDC*.

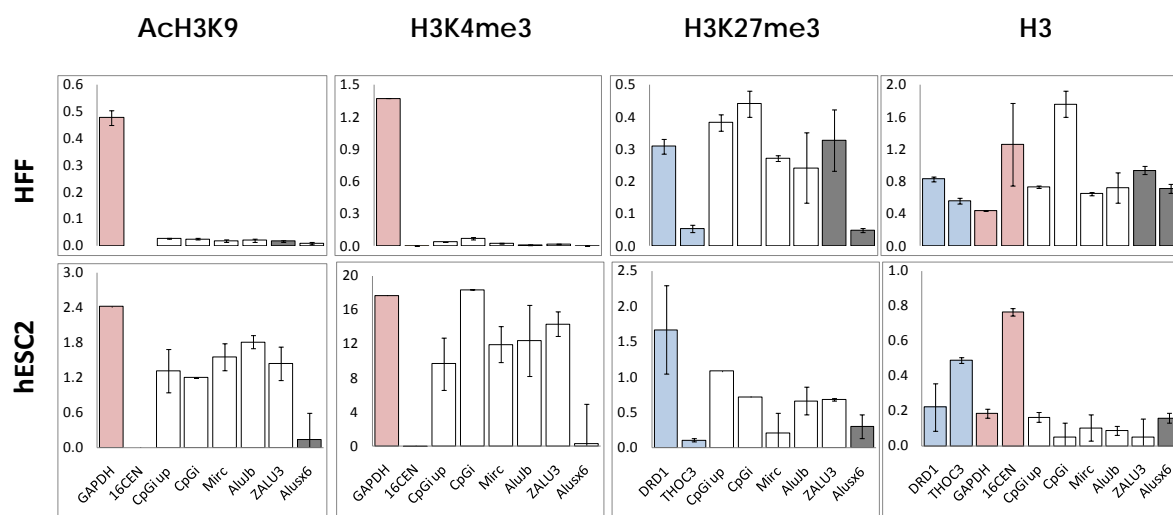


Fig. R.47. Anàlisi per CHIP de les regions flanquejants a l'element Alu ZALU3 en les línies HFF (fibroblasts) i hESC2 (cèl·lules mare embrionàries). Es representen les marques d'activació H3K4me3, Ach3K9, la marca d'inactivació H3K27me3 i l'H3 total. L'eix de les Y representa el ràtio de la marca respecte l'H3, a excepció d'H3 que és el valor absolut. Els amplicons representats en les columnes mostren el seu estat de metilació sent blanc desmetilat, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre >75% metilat. Per altra banda en vermell es representen els controls, sent GAPDH control positiu de marques actives i 16CEN control negatiu. I en blau DRD1 com a control positiu de H3K27me3 i THOC3 control negatiu.

### 3.2.2. ESTUDI DE L'EXPRESSIÓ DE GLDC A NIVELL D'ARNm

El gen *GLDC* conté 3767pb i 25 exons i, a diferència del *DIEXF*, només té un transcrit segons UCSC, RefSeq i l'Ensembl/Havana amb la qual cosa es facilita el seu anàlisi (Figura R.48).

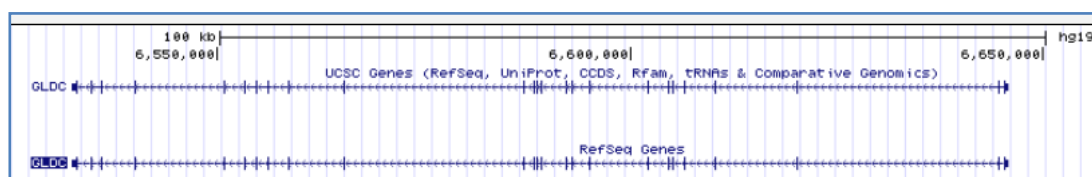


Fig. R.48. Gen *GLDC*. Genome Browser versió Feb. 2009 (GRCh37/hg19).

Com ja s'ha dit, mitjançant *arrays* d'expressió es va veure que *GLDC* s'expressa en cèl·lules on ZALU3 està desmetilada. Vam voler validar els resultats per RT-qPCR de les mostres utilitzades en l'estudi i ampliar-lo amb les mostres d'interès.

Es van analitzar diferents línies de CCR (Sw480, LoVo, HCT116, DKO, CaCo2), una parella Normal-Tumor, fibroblasts (HFF), limfòcits (GM12878), leucòcits, hepatòcits, queratinòcits (K1), cèl·lules pluripotents induïdes derivades de K1 (KiPS4F1 i KiPS4F8) i cèl·lules mare embrionàries (hESC2, hESC4) (Figura R.49).

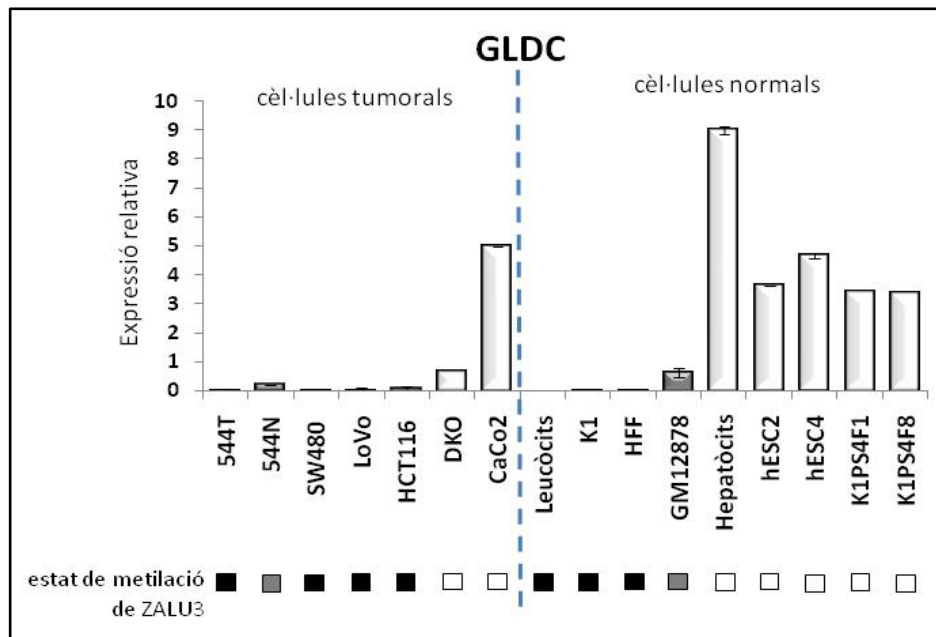


Fig. R.49. Nivells d'expressió de *GLDC* en d'una parella Normal-Tumor (544N-544T), en diferents línies de CCR (Sw480, LoVo, HCT116, DKO, CaCo2), leucòcits, queratinòcits (K1), fibroblasts (HFF), limfòcits (GM12878), hepatòcits, cèl·lules mare embrionàries (hESC2, hESC4) i cèl·lules pluripotents induïdes derivades de K1(K1PS4F1 i K1PS4F8). L'eix de les Y representa l'expressió relativa. La normalització està feta amb la mitjana dels estàndards. L'estat de metil·lació de ZALU3 es representa en escala de grisos, sent el blanc totalment desmetilat i el negre totalment metilat.

Observant els resultats veiem que les línies que expressen el gen *GLDC* són CaCo2, hepatòcits, K1PS4F1, K1PS4F8, hESC2 i hESC4, i en menor grau DKO. Això confirma que les línies cel·lulars que tenen tota la regió desmetilada incloent ZALU3 expressen el gen i aquelles línies que, tot i tenir la regió desmetilada, tenen l'Alu ZALU3 metilada, no expressen el gen, aquest seria el cas de HFF.

Quan fem el tractament amb drogues epigenètiques tant en HCT116 com en Sw480 no aconseguim en cap cas desmetil·lar completament la regió i no observem canvis en l'expressió (dades no mostrades).

### 3.2.3. ESTUDI DE L'EXPRESSIÓ DE *GLDC* A NIVELL DE PROTEÏNA

La proteïna que codifica el gen *GLDC* és de 1020 aa segons l'Ensembl (Figura R.50).

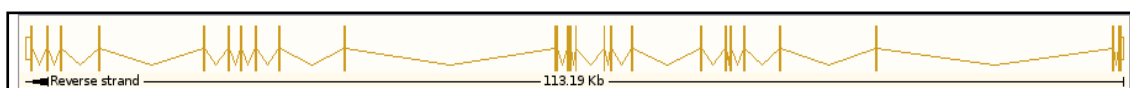


Fig. R.50. Imatge de l'Ensembl del transcrit de *GLDC*.



L'expressió de la proteïna es va analitzar per Western Blot en diferents línies de CCR (Sw480, HCT116, DKO, LoVo, HT29 i CaCo2), així com les cèl·lules d'hepatòcits, cèl·lules mare embrionàries (hESC2), un línia de limfòcits (GM12878) i els fibroblasts (HFF) (Figura R.51).

L'anticòs que es va fer servir va ser l'Abcam, que reconeix la proteïna GLDC entre els aminoàcids 492 i 693.

L'expressió a nivell de proteïna es correspon amb l'obtinguda en l'ARN, així doncs les úniques línies que expressen GLDC són els hepatòcits, les hESC, CaCo2 i, en menor grau, les DKO i, que són les que tenen tota la regió promotora de *GLDC* desmetilada incloent ZALU3.

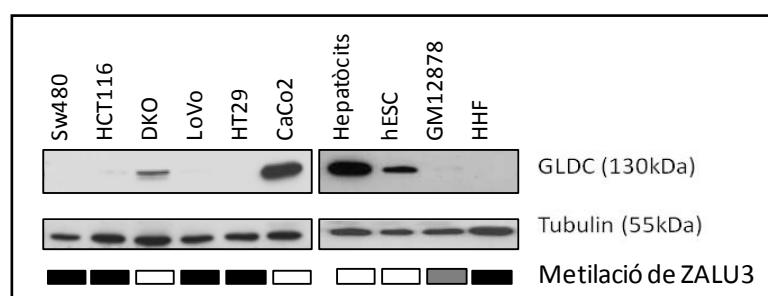


Fig. R.51. Expressió de *GLDC* en les diferents línies cel·lulars per Western Blot. La proteïna d'interès es detecta amb l'anticòs policlonal Anti-*GLDC* (Abcam). Com a control de càrrega utilitzem l'anticòs monoclonal Anti-Tubulin. L'estat de metilació de ZALU3 es representa en escala de grisos, sent el blanc totalment desmetilat i el negre totalment metilat.

### 3.2.4. REGULACIÓ DE L'ELEMENT ZALU3 EN LA DIFERENCIACIÓ

Veient la possible implicació del nostre element Alu en la regulació del gen *GLDC* en la diferenciació es va analitzar un estudi de diferenciació a diferents temps de cèl·lules mare embrionàries cap a "embrioid bodies" (EBs) (mostres facilitades per la Dr. Maria José Barrero del CMRB). Es parteix de cèl·lules mare embrionàries, a dia 0, i es cultiven amb el medi de diferenciació mTeSR1 (*BD Matrigel hESC-qualified matrix and STEMCELL Technologies*), recollint a dia 2, 4, 8 i 12.

Primerament es va mirar si el procés de diferenciació va funcionar fent l'anàlisi d'expressió d'un marcador de diferenciació (*GATA6* - *Transcription factor GATA-binding factor 6*-) i un de pluripotència (*OCT4* - *Octamer-binding transcription factor 4*-) (Figura. R.52). Veiem que a mida que augmenta el temps baixen els nivells d'*OCT4* demostrant la pèrdua de pluripotència de les cèl·lules mare embrionàries i en canvi augmenta *GATA6* indicant la seva diferenciació.

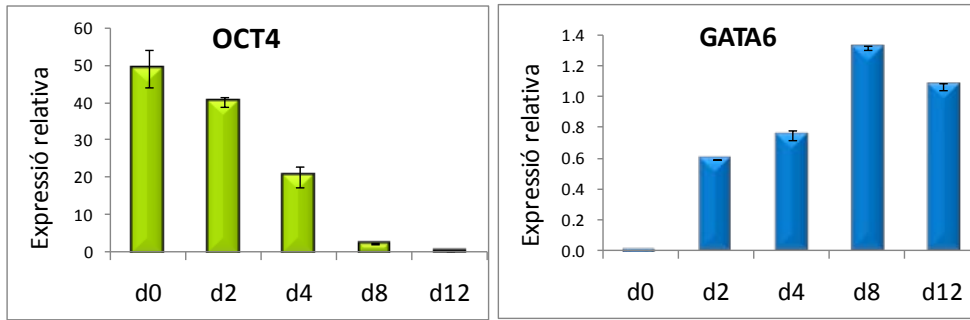


Fig. R.52. OCT4 i GATA6 utilitzats com a controls del correcte funcionament de l'estudi de diferenciació a diferents temps de cèl·lules embrionàries cap a EBs. El nivell d'expressió dels gens està normalitzat amb la mitja de 5 estàndards.

Un cop confirmat el correcte procés de diferenciació es va mirar l'estat de metilació de la regió (Figura R.53). Veiem que, a mesura que es diferencien les cèl·lules mare embrionàries, l'element ZALU3 es metila. Aquest procés té lloc ja a dia 2 de diferenciació.

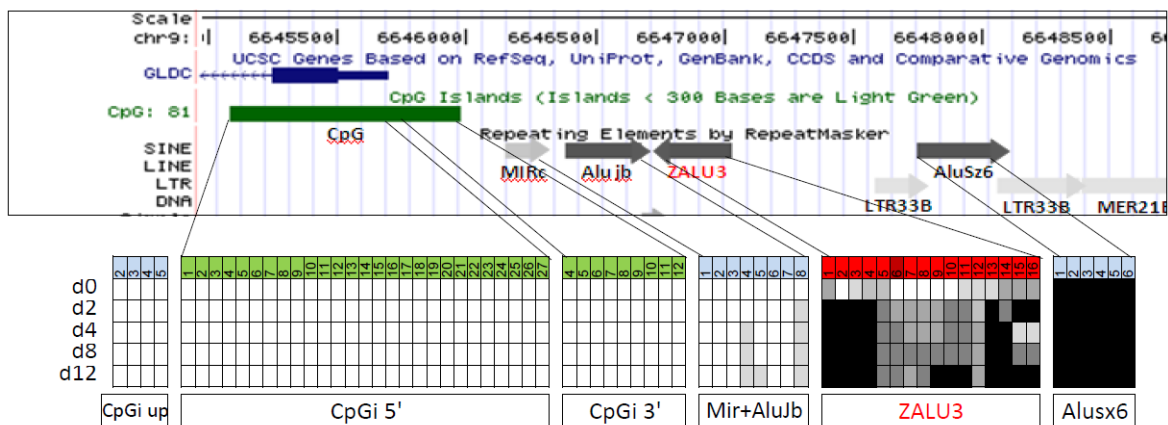


Fig. R.53. Estat de metilació de la regió flanquejant a l'element Alu ZALU3 seleccionat (en vermell) en diferents temps del procés de diferenciació. Regió obtinguda del Genome Browser versió Feb. 2009 (GRCh37/hg19). Cada requadre indica una CpG. En vermell trobem l'Alu ZALU3, en verd l'illa CpG i en blau diferents elements repetitius. En blanc de 0-25% de metilació, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre 75-100% de metilació.

Continuant amb l'estudi vam analitzar les marques d'histona per veure si aquest canvi de metilació en un estadi tan primerenc, també es manifestava a nivell de la cromatina (Figura R.54). Ja a dia 2 veiem una pèrdua de marca activa (tant de AcH3K9 com H3K4me3) i una compactació del domini de cromatina activa a mesura que es diferencien. Pel que fa a la marca inactiva H3K27me3, observem com aquesta augmenta i es propaga al llarg de la regió a mesura que les hESC es diferencien, sent molt major a dia 12.

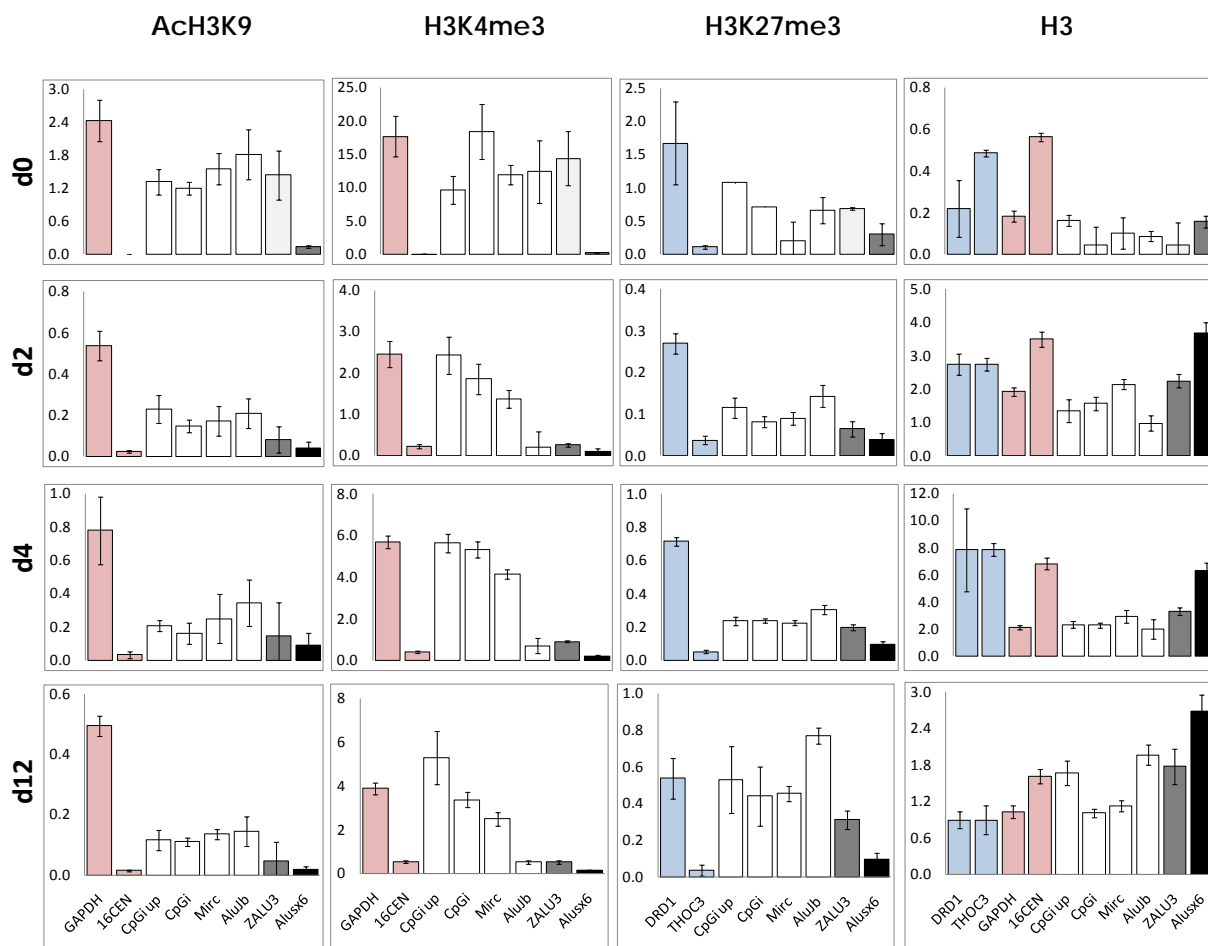


Fig. R.54. Anàlisi per ChIP de les regions flanquejants a l'element ZALU3 en els diferents temps de diferenciació. Es representen les marques d'activació H3K4me3, AcH3K9, la marca la inactivació H3K27me3 i l'H3 total. L'eix de les Y representa el ratió de la marca respecte a l'H3, a excepció d'H3 que és el valor absolut. Els amplicons representats en les columnes mostren el seu estat de metilació sent blanc desmetilat, gris clar 25-50% de metilació, gris fosc 50-75% de metilació i negre >75% metilat. Per altra banda en vermell es representen els controls, sent GAPDH control positiu de marques actives i 16CEN control negatiu. I en blau DRD1 com a control positiu de H3K27me3 i THOC3 control negatiu.

Si mirem com afecta aquest canvi en l'expressió del gen *GLDC* veiem que, ja a dia 2, hi ha una pèrdua tant a nivell d'ARN com de proteïna (Figura R.55).

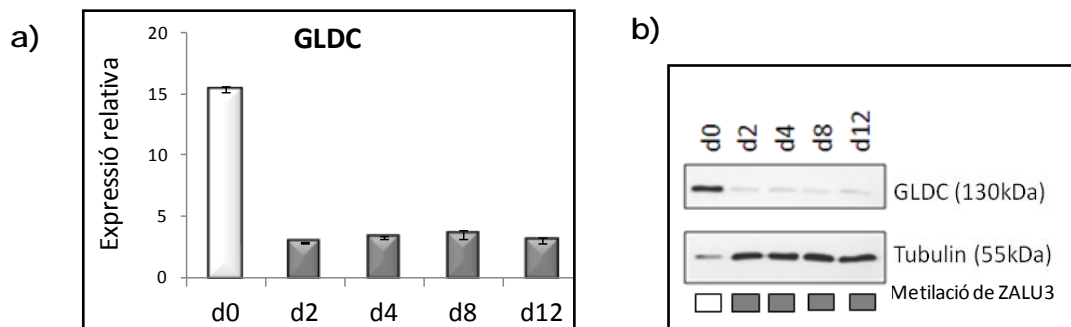


Fig. R.55. Expressió de *GLDC* dels diferents temps de diferenciació. a) Expressió de l'ARNm. L'eix de les Y representa l'expressió relativa. La normalització està feta amb la mitjana de 5 estàndards. L'estat de metilació de *ZALU3* es representa en escala de grisos, essent el blanc totalment desmetilat i el negre totalment metilat. b) Western Blot de l'estudi de diferenciació a diferents temps. La proteïna d'interès es detecta amb l'anticòs policlonal Anti-*GLDC* (Abcam). Com a control de càrrega utilitzem l'anticòs monoclonal Anti-Tubulin.

Finalment i veient que *CaCo2* expressava molt *GLDC* vam fer un estudi induint la diferenciació cel·lular en *CaCo2*. Morfològicament, *CaCo2* mostra un cultiu molt heterogeni i conté subpoblacions de cèl·lules majoritàriament indiferenciades però amb diversa capacitat de diferenciació. Al deixar en cultiu en confluència les *CaCo2* es diferencien i polaritzen exhibint moltes característiques i funcions de les cèl·lules diferenciades de l'epiteli de la mucosa intestinal. En la figura R.56 veiem que, al diferenciar *CaCo2*, l'expressió de *GLDC* es redueix tot i que no fins al punt d'eliminar totalment l'expressió. L'element *ZALU3* no s'arriba a metilar, segurament degut a que la població és molt heterogènia i caldria analitzar diferents clons per veure en quin percentatge de cèl·lules es manté desmetilat i en quin es metila.

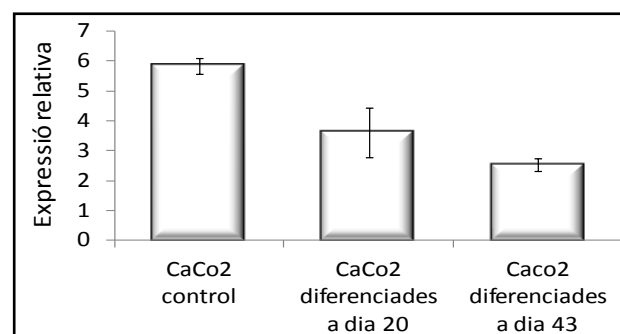
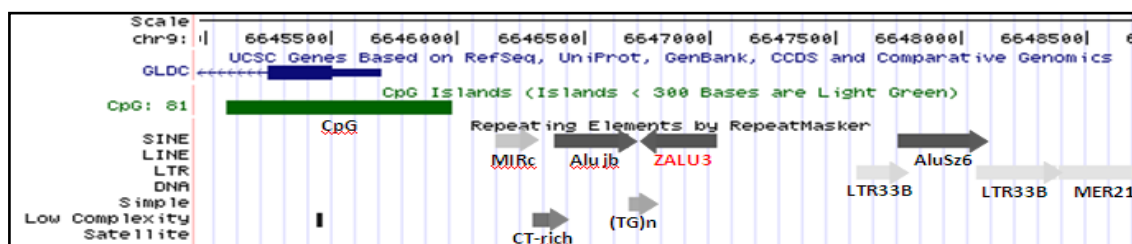


Fig. R.56. Expressió de *GLDC* a diferents temps després d'induir la diferenciació en la línia *CaCo2*. L'eix de les Y representa l'expressió relativa. La normalització està feta amb la mitjana de 5 estàndards. L'estat de metilació de *ZALU3* es representa en escala de grisos, essent el blanc totalment desmetilat i el negre totalment metilat.

Tal i com resumim en la taula R.14 i, en conclusió, veiem que l'estat epigenètic de la regió regula l'expressió del gen *GLDC*, sent l'element *Alu ZALU3* el que denota canvis associats més clars. Aquestes dades suggereixen que *ZALU3* podria tenir un paper determinant en la regulació del gen *GLDC* durant la diferenciació.

TAULA R.14. RELACIÓ ENTRE L'ESTAT EPIGENÈTIC DE L'ELEMENT *ALU ZALU3* I L'EXPRESSIÓ DE *GLDC*



		estat metilació <sup>a</sup>			marca d'histona H3K4me3 <sup>b</sup>			marca d'histona H3K27me3 <sup>b</sup>			EXPRESSIÓ <i>GLDC</i> <sup>c</sup>	
		5'	ZALU3	3'	5'	ZALU3	3'	5'	ZALU3	3'	RNA	PROTEÏNA
Cèl·lules mare embrionàries					↑↑	↑↑	↑↑	↑	↑	↑		
	Cèl·lules pluripotents induïdes					ND			ND			ND
Cèl·lules mare diferenciades a dia 2					↑↑	∅	∅	↑	↑	↑		
	Cèl·lules mare diferenciades a dia 12				↑↑	∅	∅	↑↑	↑↑	↑↑		
	fibroblasts				∅	∅	∅	↑↑	↑↑	↑↑		
queratinocits					ND			ND			ND	
CCR	LoVo					ND			ND			
	HT29					ND			ND			
	Sw480				∅	∅	∅	∅	∅	∅		
	HCT116				↑	∅	∅	∅	∅	∅		
	DKO				↑	∅	∅	∅	∅	∅		
	CaCo2				↑↑	↑↑	↑↑	∅	∅	∅		

a) Estat de metilació de la regió. L'element *ZALU3* i la regió flanquejant a *ZALU3* tant en l'extrem 5' com 3'. Desmetilat indicat en blanc, metilat en negre i en escala de grisos parcialment desmetilat.

b) Marca d'histona *H3K4me3* de la regió. L'element *ZALU3* i la regió flanquejant a *ZALU3* tant en l'extrem 5' com 3'. ∅ indica que no hi ha marca, ↑ indica presència moderada de marca activa o repressiva, ND indica que no s'ha analitzat i ↑↑ indica presència de molta marca activa o repressiva,

c) Expressió *GLDC*. L'expressió a nivell de ARN dels diferents transcrits i l'expressió a nivell de proteïna. En vermell no expressió i en verd expressió, ND indica que no s'ha analitzat.

**DISCUSSIÓ**

## 1. ELS ELEMENTS ALU DESMETILATS CONSTITUEIXEN UNA SUBPOBLACIÓ ESPECÍFICA DINS L'ALUOMA

Els elements Alu són la família de retrotransposons més abundant del genoma humà i cobreixen el 33% del total de CpGs, la majoria de les quals es troben altament metilades en teixits somàtics. En aquest treball ens hem centrat en aquella fracció d'elements Alu desmetilats en cèl·lules normals i/o tumorals per determinar la seva importància biològica i la possible implicació en la regulació transcripcional.

S'han desenvolupat nombroses tècniques per l'anàlisi de la metilació de l'ADN. Aquestes tècniques es poden classificar en tres categories: les basades en enzims de restricció, les basades en anticossos específics per ADN metilat i les basades en el tractament amb bisulfit sòdic (Laird, 2010). Totes elles es poden combinar amb diferents metodologies enfocades a anàlisi específics com la PCR o a anàlisi a escala genòmica com els *microarrays* o la seqüenciació de nova generació. Com és evident cada metodologia té els seus pros i contres. Les dades generades a partir d'enzims de restricció sensibles a metilació troben la limitació dels llocs de reconeixement dels enzims emprats. Els mètodes basats en anticossos són poc resolutius i poc sensibles en l'estudi de seqüències repetitives, especialment si els que ens interessin són els elements no metilats. Finalment, els mètodes que utilitzen la conversió de l'ADN tractat amb bisulfit sòdic, tot i considerar-se el *gold estàndar* per la seva alta resolució i sensibilitat, són poc eficients per estudis a escala genòmica. Per tant, totes les tècniques presenten cert biaix i avantatges o inconvenients depenent del disseny de l'estudi, a més de les limitacions associades a l'estudi de seqüències repetitives.

Una de les principals línies de recerca del nostre laboratori en els darrers anys és la d'identificar seqüències repetitives desmetilades, per això es va optar per l'estudi de l'estat de metilació dels elements repetitius Alu amb metodologies basades en enzims de restricció, concretament es va desenvolupar la tècnica AUMA (*Amplification of Unmethylated Alus*). Recordem que les seqüències generades per l'AUMA es basen en el lloc CpG dins de la seqüència consens de l'Alu, AACCCGGG (AA + diana SmaI). Aquesta seqüència consens es troba en uns 200.000 elements Alu del total d'Alus del genoma humà (aproximadament un milió) (Rodríguez et al., 2008b). La tècnica AUMA ens va permetre la identificació de 94 elements Alu parcial o totalment desmetilats en mucosa de còlon normal i/o en cèl·lules de carcinoma colorectal (Rodríguez et al., 2008b). Aquest abordatge és adient per l'anàlisi comparatiu d'un gran nombre de mostres, però els resultats no permeten conèixer directament el nivell absolut de metilació del elements representats.

Mitjançant bisulfit sòdic i la posterior seqüenciació d'alguns dels elements Alu identificats no només vam validar els resultats de l'AUMA sinó que vam veure que l'estat de metilació de la CpG dins la diana SmaI era representativa de l'estat global de metilació de l'element Alu estudiat. Aquests resultats concorden amb els que Barrera i col·laboradors (Barrera and Peinado, 2012) van veure usant dades de metilomes d'alta resolució, on demostraven la correlació entre l'estat de metilació de la diana Hpall situada dins una illa CpG i l'estat global de metilació de l'illa CpG. Tanmateix, la metilació diferencial entre mostres trobada per AUMA es va poder confirmar mitjançant la seqüenciació amb bisulfit.

Per determinar si els elements Alu desmetilats constituïen una subpoblació particular i diferenciada dins de l'ALUOMA es van analitzar una sèrie de trets a nivell de seqüència (nombre de CpGs, divergència, mida, etc) i de context genòmic (distància a un gen, localització intragènica, etc.). A l'hora de fer l'estudi bioinformàtic, en primer lloc es van comparar 100 elements *random* representatius del genoma i 100 elements virtuals, potencialment amplificables per la tècnica AUMA (Taula R.1). Vam observar que la tècnica AUMA presentava alguns biaixos ja que els elements Alu potencialment amplificables per la tècnica presentaven una cua poli (A) més curta que la mitja de les Alus del genoma, una baixa divergència, un alt nombre de GC i es trobaven enriquits en membres de la família Y i empobrits en membres de la família J. En conseqüència, per poder definir trets genòmics característics i específics dels elements Alu desmetilats vam considerar que els elements Alu virtuals eren el millor grup control. Els paràmetres analitzats van determinar que els elements Alu desmetilats es caracteritzen per presentar una llargada major de la cua poli (A), estar enriquits en Alus S i empobrits en Alus Y, presentar una baixa divergència respecte la seqüència consens, i localitzar-se majoritàriament prop d'inicis de transcripció associats a illes CpG (Figura R.1).

Tal i com ja hem mencionat en altres apartats, els elements Alu no codifiquen per cap proteïna però tenen un promotor ARN Pol III que permet la seva transcripció i utilitzant la maquinària de retrotransposició dels elements LINE-1 poden mobilitzar-se. S'ha descrit que la llargada de la cua poli (A) podria ser un factor important per permetre la retrotransposició de l'element Alu i, coincidint amb els nostres resultats, normalment s'associa una major allargada de la cua poli (A) a elements Alu actius (Roy-Engel et al., 2002b).

Com ja s'havia descrit en altres treballs (Rodríguez et al., 2008b; Szpakowski et al., 2009), vam veure que hi havia una relació inversa entre l'estat de metilació i l'edat de l'element Alu. Així doncs, al comparar Alus virtuals i *random* vam veure que la família Alu Y (la més jove) tenia més prevalença de la diana Smal mentre que la família J (la més antiga) presentava menys seqüències amb la diana Smal. No obstant, al mirar els elements Alu desmetilats vam veure un empobriment de la família Y. La majoria de les insercions dels elements Alu són antigues i inactives però algunes famílies (com la Alu Ya i la AluYb) romanen encara actives (Jurka et al., 2004; Xing et al., 2007). Així doncs, el fet de trobar menor percentatge d'elements Alu Y desmetilats s'explicaria per la forta pressió epigenètica que pateixen aquests elements per reprimir la seva activitat. Pel que fa a les Alus de la família J, vam observar un percentatge d'elements desmetilats molt baix i semblant a l'observat en els elements Alu virtuals. Els elements Alu J es caracteritzen per acumular un gran nombre de mutacions a l'atzar, la qual cosa ve determinada pel seu origen ancestral. Aquesta acumulació de mutacions podria haver afectat a la diana Smal, fent que la pèrdua d'aquesta impedeixi identificar-les per la tècnica AUMA. Finalment, cal destacar un lleuger enriquiment d'Alus S, les més freqüents i d'edat intermèdia, entre les Alus desmetilades que hem identificat.

Al mirar la divergència, vam veure que amb la tècnica AUMA capturàvem un major nombre d'elements Alu amb baixa divergència i aquesta es mantenia en els elements desmetilats. Així doncs sembla que les Alus desmetilades estan força conservades a nivell de seqüència, suggerint-nos el seu paper actiu dins del genoma (Medstrand et al., 2002). En aquest sentit, el fet que citosines metilades siguin més fàcilment mutables a timines que les citosines desmetilades (Polak and Arndt, 2008) explicaria que els



elements Alu desmetilats no pateixin tantes mutacions i que estiguin més conservats a nivell de seqüència que els elements metilats.

Finalment vam veure que els elements Alu desmetilats es localitzaven en regions promotores associades a illes CpG amb major freqüència que els elements Alu virtuals, suggerint el seu paper regulador de l'expressió gènica. En concordança amb aquests resultats, en humans s'ha estimat que més de 1.000 gens podrien estar regulats per elements Alu associats a illes CpG (Oei et al., 2004). Un gran nombre d'articles apunten que les regions promotores són gairebé les úniques regions del genoma desmetilades (Goll and Bestor, 2005; Jones and Baylin, 2002), incloent per tant els elements Alu situats en aquest context. Ara bé, en els nostres estudis hem observat que no tots els elements Alu propers a un inici de transcripció es troben desmetilats, indicant que l'estat de metilació de l'element Alu no només depèn de la posició. En aquest sentit, s'ha vist que els elements Alu podrien actuar com a dianes o llocs d'unió a factors de transcripció modulant la transcripció gènica (Polak and Domany, 2006) o bé com a aïlladors genòmics (Molto et al., 2009), és a dir, actuant com a barrera entre dominis de cromatina activa i inactiva. Més endavant ja entrarem en detall sobre el paper dels elements Alu en la regulació transcripcional.

Xie i col·laboradors (Xie et al., 2009) van obtenir resultats similars als nostres quan van estudiar el metiloma dels elements Alu en cervell. En el seu cas, la metodologia emprada es basava en la seqüenciació massiva de ADN tractat amb bisulfit sòdic. Després de la digestió amb AluI (un enzim insensible a metilació) lligaven uns adaptadors, tractaven l'ADN amb bisulfit sòdic i amplificaven els fragments amb un encebador específic d'elements Alu; d'aquesta manera, a diferència de nosaltres, amplificaven elements Alu del genoma independentment del seu estat de metilació. Com en el nostre cas, l'anàlisi tenia biaixos, concretament cap a les Alu més joves degut a l'ús d'un encebador capaç d'amplificar elements Alu rics en dinucleòtids CpG. Un altre punt de concordança amb els nostres resultats va ser el fet d'observar que el nivell de metilació es reduïa en les regions properes a inicis de transcripció i que augmentava proporcionalment a mida que es distanciaven del TSS. No obstant, donat l'alt cost de l'abordatge realitzat per Xie i col·laboradors, es van limitar a una sola mostra i amb una cobertura molt limitada. Nosaltres, considerant que els elements Alu desmetilats són els que podrien jugar un paper funcionalment rellevant, vam dirigir específicament l'estudi cap als elements desmetilats, el que va suposar el disseny d'una estratègia menys costosa i amb més cobertura.

Cal dir que en el nostre laboratori, s'ha desenvolupat recentment una altra metodologia basada en enzims de restricció, la tècnica NSUMA (de l'anglès *Next Generation Sequencing of Unmethylated Alus*) que és una millora de la tècnica AUMA. El NSUMA aprofita la plataforma de seqüenciació massiva Illumina de manera que permet estudiar més de 100.000 elements Alu a diferència dels gairebé 5.000 analitzables per la tècnica AUMA. Amb la tècnica NSUMA s'han analitzat tres pacients de càncer de còlon, incloent ADN de sang, les línies colorectals HCT116 i HCT116-DKO, així com una línia cel·lular embrionària mare. Resultats preliminars confirmen el ja vist en aquest treball, que les Alus desmetilades estan empobrides en Alus Y i enriquides en Alus S, que presenten una baixa divergència i una proporció important es localitza prop d'inicis de transcripció. A més, aquests elements Alu tenen un alt contingut de CpGs i un alt percentatge de GC.

## 2. ELS ELEMENTS ALU DESMETILATS PRESENTEN UN MAJOR GRAU DE CONSERVACIÓ EVOLUTIVA

Com ja s'ha comentat el 5% del genoma humà s'ha conservat evolutivament en els últims 200 milions d'anys, incloent la gran majoria dels gens i seqüències reguladores, no obstant la major part de la seqüència genòmica amb gran importància funcional és desconeguda.

En l'apartat 2.1.3 dels resultats analitzàvem 17 elements Alu que havíem trobat desmetilats total o parcialment en mostres de còlon tumoral i/o normal. Donat que l'aparició de les seqüències Alu es situa fa aproximadament 65 milions d'anys, coincidint amb l'origen i expansió dels primats, vam voler estudiar diferents espècies de primats a fi d'establir una correlació entre la conservació a nivell de seqüència i l'estat de metilació de les mostres de sang analitzades. Així doncs, amb aquests elements Alu vam fer un estudi *in silico* utilitzant com a genoma de referència el genoma humà i alineant-lo amb el genoma de ximpanzé, goril·la, orangutan i macaco (Taula R.7).

A mesura que ens allunyàvem en l'escala evolutiva vam veure que l'estat de conservació genètica de les Alus s'anava reduint, sent en macaco on més divergència trobàvem respecte l'humà. Recordem que aquestes espècies pertanyen a la família del *Catarrhinis*, sent el *Homo Sapiens* (humà), el *Pan troglodytes* (ximpanzé), el *Gorilla gorilla* (goril·la) i *Pongo pygmeus* (orangutan) de la subfamília *Hominidae* i el *Macaca Mulatta* (Rhesus) de la subfamília *Cercopithecidae*, les quals van divergir fa 25-30 milions d'anys. La subfamília *Hominidae* està formada per primats bípedes i grans simis, i diversos estudis genètics han vist la gran similitud entre aquestes espècies, sent l'orangutan el que es troba més separat filogenèticament. Per altra banda, la subfamília *Cercopithecidae* està formada pels micos del "vell món" (macacos i babuins) i, tal i com vèiem en els nostres resultats, la més allunyada filogenèticament a l'humà.

Cal mencionar també que el genoma de Goril·la, recentment seqüenciat (Gorilla (May 2011 (gorGor3.1/gorGor3)), encara està en fase d'assemblatge, per la qual cosa el genoma referència emprat s'ha de considerar de baixa qualitat, explicant que els resultats obtinguts en goril·la no mostrin l'homologia que esperàvem i siguin provisionals.

A nivell de família, dels 17 elements Alu estudiats tenim 4 elements de la família Alu Y (la més jove), 13 elements de la família Alu S (la més freqüent i d'edat intermèdia) i cap element de la família J (la família més antiga). Veient els resultats no vam poder establir cap correlació entre l'estat de conservació dels elements Alu i la família a la qual pertanyen. No obstant, cal destacar el fet de no trobar cap dels 4 elements Y, la família més jove, en macaco.

Al mirar l'estat de metilació de les sangs en les diferents espècies vam veure que els elements Alu desmetilats presentaven un major grau de conservació en la seqüència que els elements més metilats. Diferents estudis "filoepigenòmics" han permès dilucidar els canvis epigenètics que s'han anat succeint durant l'evolució (Martin et al., 2011). Aquestes diferències epigenètiques regulen l'activitat de regions no codificants, intervenint en la funció genòmica a través de la regulació transcripcional. És doncs

interessant veure que aquells elements que estan completament desmetilats en sangs humanes, i probablement actius, estan més conservats en les diferents espècies de primats, indicant doncs la seva importància biològica. Per altra banda, els elements més metilats mostren més variació, en part degut a que les citosines metilades tendeixen a mutar a timines molt més ràpid que les citosines desmetilades. Aquesta observació ens fa pensar en el fet, ja conegut, que els elements Alu no són únicament *junk DNA* sinó que tenen un rol important en la regulació de l'expressió gènica i la regulació funcional jugant un paper més que probable en l'evolució dels primats i la divergència d'altres mamífers.

L'estudi comparatiu, mitjançant bisulfit sòdic, de 6 elements Alu entre les diferents espècies de primats ens va permetre observar que l'estat de metilació estava relacionat amb l'estat de conservació, no només genètica, confirmant el que ja havíem vist *in silico*, sinó també epigenètica. Vam veure que l'estat de metilació de les diferents CpGs dels elements Alu no variava entre les diferents espècies estudiades, mantenint-se metilades o desmetilades en totes elles, però observant una major tendència a mutar-se a CpA o a TpG en les CpG metilades que en les CpG desmetilades.

Actualment, en el laboratori, estem pendent d'analitzar bioinformàticament un experiment fet amb la tècnica NSUMA, on es van analitzar 3 mostres de sang humanes, 3 mostres de sang de ximpanzé, 3 de goril·la i 3 d'orangutan. Aquests resultats ens permetran analitzar un gran nombre d'elements Alu desmetilats i comparar-los amb la seqüència humana de referència. D'aquesta manera veurem més clarament l'estat de conservació d'aquests elements Alu desmetilats en les diferents espècies de primats.

### 3. ESTAT EPIGENÈTIC DELS ELEMENTS ALU I EL SEU PAPER COM A BARRERA DE METILACIÓ

Al llarg d'aquesta tesi hem anat analitzant 17 elements Alu identificats per la tècnica AUMA, tant a nivell de metilació com de marques d'histona, en diferents mostres i línies cel·lulars, aportant-nos informació sobre el seu estat epigenètic en càncer, més en concret en càncer de còlon i en la diferenciació cel·lular.

En termes globals vam veure que els elements Alu estudiats mostraven majoritàriament un perfil diferencial entre cèl·lules normals i tumorals, tendint a estar més desmetilats en les cèl·lules tumorals. Per altra banda, a l'analitzar diferents tipus cel·lulars vam veure que alguns elements Alu canviaven el seu estat de metilació segons el teixit on es trobaven indicant una possible implicació en la diferenciació o l'organoespecificitat. Altres elements Alu es mantenien desmetilats en tots els teixits/línies cel·lulars estudiades. Com ja hem mencionat en alguna altra ocasió, l'estat desmetilat dels elements repetitius està altament penalitzat ja que constitueix una probable font d'inestabilitat genòmica. En conseqüència, la preservació d'aquest elements en estat epigenèticament actiu suggereix una funcionalitat dins del genoma, convertint-los doncs en dianes d'interès. Finalment també vam observar un gran nombre d'elements Alu que mantenien un elevat grau de metilació en tots els tipus cel·lulars analitzats (incloent la línia DKO deficient en metilació del ADN), el que és consistent amb la pressió per mantenir la seva inactivació dins el genoma.

Pel que fa a les marques d'histona analitzades vam veure una concordança entre l'estat de metilació i les marques d'activació, en concret d'Ach3K9 i H3K4me3. Els elements Alu desmetilats presentaven Ach3K9 i H3K4me3 i els que es trobaven altament metilats, no. Més concretament, aquestes marques actives s'associen a elements Alu desmetilats propers a inicis de transcripció. En aquest sentit, està àmpliament descrit que les marques de di/trimetilació de la lisina 4 incrementen la transcripció i es troben en regions eucromàtiques, sobre gens actius o activables (Schneider et al., 2004). Per altra banda, l'acetilació es veu implicada en múltiples processos biològics i, en general, s'associa a regions del genoma transcripcionalment actives. L'acetilació de la lisina 9 així com de la lisina 14 són necessàries pel reconeixement de l'inici de transcripció (Bannister and Kouzarides, 2011), però nosaltres, pel que fa a la marca activa Ach3K14, vam veure que es trobava present en tot els elements Alu sense presentar grans diferències entre el seu estat de metilació, així doncs segons els nostres resultats, Ach3K14 s'associaria a elements Alu independentment del seu estat de metilació i de la seva localització genòmica.

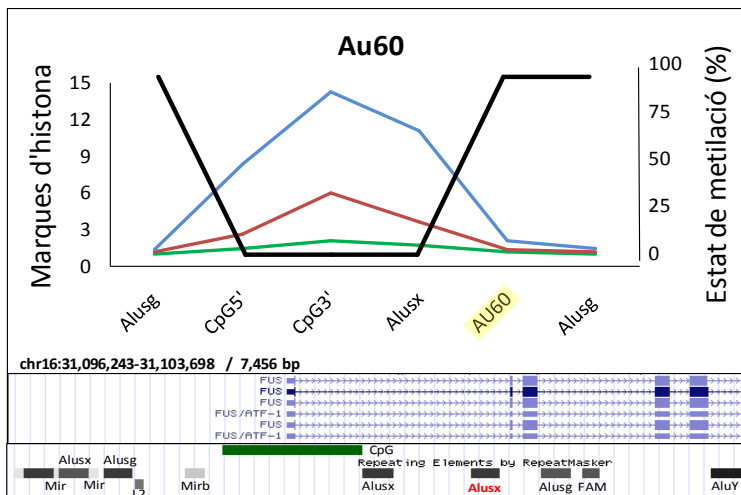
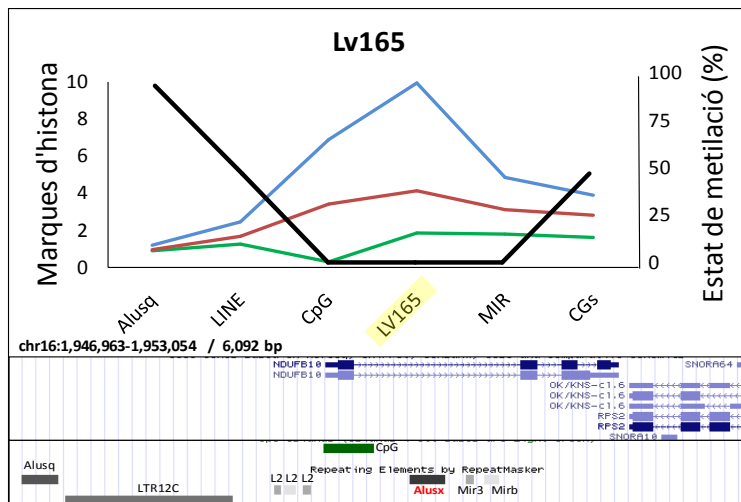
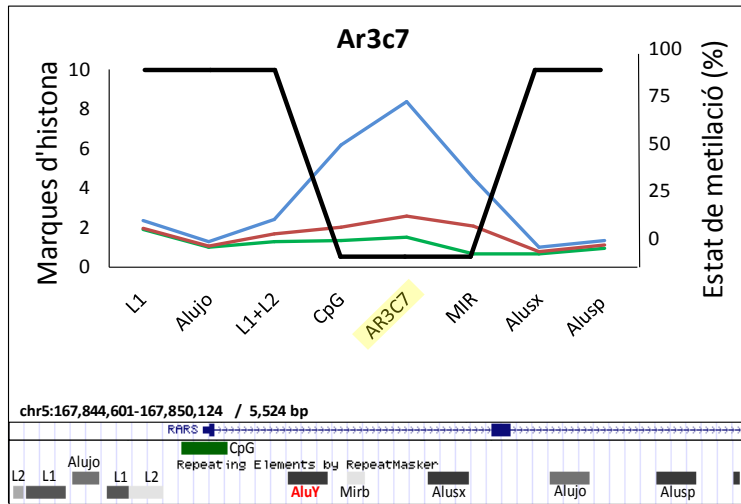
Finalment, es va analitzar l'estat de la cromatina de l'ALUOMA mitjançant l'ús de primers consens per elements Alu (ChIP-qAlu), representatiu dels elements Alu totals del genoma. Donat que ChIP-qAlu es una representació global dels elements Alu, els quals es troben majoritàriament metilats els nivells de H3K4me3 (que està present només en els elements Alu desmetilats) van ser molt baixos. Per contra, l'Ach3K14, associada tant a elements Alu metilats com desmetilats, va donar alts nivells.

L'anàlisi dels elements Alu ens va portar a estudiar la regió flanquejant tant a nivell de metilació com d'estat de la cromatina. L'estudi es va fer en diverses línies, presentant sempre la mateixa correlació entre marques actives i estat de metilació de la regió. En

la figura D.1 es mostren els diversos elements Alu analitzats (Ar3c7, Lv165, Au60, Aj2c1 i Au104) en la línia de CCR HCT116. Aquest gràfic ens mostra la gran correlació entre la desmetilació de la regió i les marques d'histona activa (H3K4me3, AcH3K9).

Els resultats indicaven que els elements Alu estarien actuant com a barrera protectora de la metilació i semblaria que l'amplada del domini de cromatina activa aniria marcada per l'estat de metilació de l'element Alu. Coincidint amb aquesta hipòtesi, s'ha vist que els elements AluY / S, tendeixen a trobar-se en el marge entre les regions genòmiques hipermetilades i hipometilades (Rollins et al., 2006).

L'expressió d'un gen és un fenomen complex i encara no entès en la seva totalitat. La seva expressió està relacionada amb les seqüències veïnes i la seva posició dins del nucli. Aquests efectes de posició permeten explicar com s'organitzen els gens en els cromosomes, en forma d'unitats independents, generant dominis d'expressió. Aquestes unitats sovint contenen elements reguladors que poden estar localitzats a grans distàncies del cos principal del gen. No obstant, es sap ben poc d'aquests elements reguladors distants, i encara que cada cop n'hi han més exemples en la literatura, desconeixem en bona part l'extensió i els mecanismes que governen la regulació genòmica de llarg abast (Smallwood and Ren, 2013). Aquests elements reguladors són seqüències adjacents a un locus, amb la capacitat de restringir o garantir l'expressió d'un gen determinat en un temps i un lloc concret (Krivega and Dean, 2012). S'ha vist que dins d'aquests elements reguladors, que actuen com a *insulators*, trobem regions *scaffold/matrix-attachments*, elements repetitius com SINES o LINES, llocs diana de CTCF i enhancers (Krivega and Dean, 2012; Molto et al., 2009). Així doncs diferents seqüències genòmiques convergeixen en una mateixa funció, la de regular i participar en la correcta organització, estructura i funcionament del genoma dels vertebrats (Bushey et al., 2008). Dins d'aquestes seqüències genòmiques encarregades de regular dominis adjacents trobem els elements Alu. Com hem anat veient, els elements Alu desmetilats i conservats al llarg de l'evolució sembla que juguen un paper important dins els processos biològics, regulant l'expressió gènica, actuant com a potenciadors, activadors de l'expressió gènica, o promotors (Kreahling and Graveley, 2004; Schulz et al., 2006; Slotkin and Martienssen, 2007). S'ha vist que els elements Alu podrien actuar com a dianes o llocs d'unió a factors de transcripció, formant un mecanisme alternatiu de regulació en els processos biològics (Polak and Domany, 2006). Així doncs, segons la seva localització dins el genoma, podrien facilitar, juntament amb altres elements, l'estat de cromatina oberta i permetre la unió de factors de transcripció i la conseqüent transcripció gènica de determinats gens.



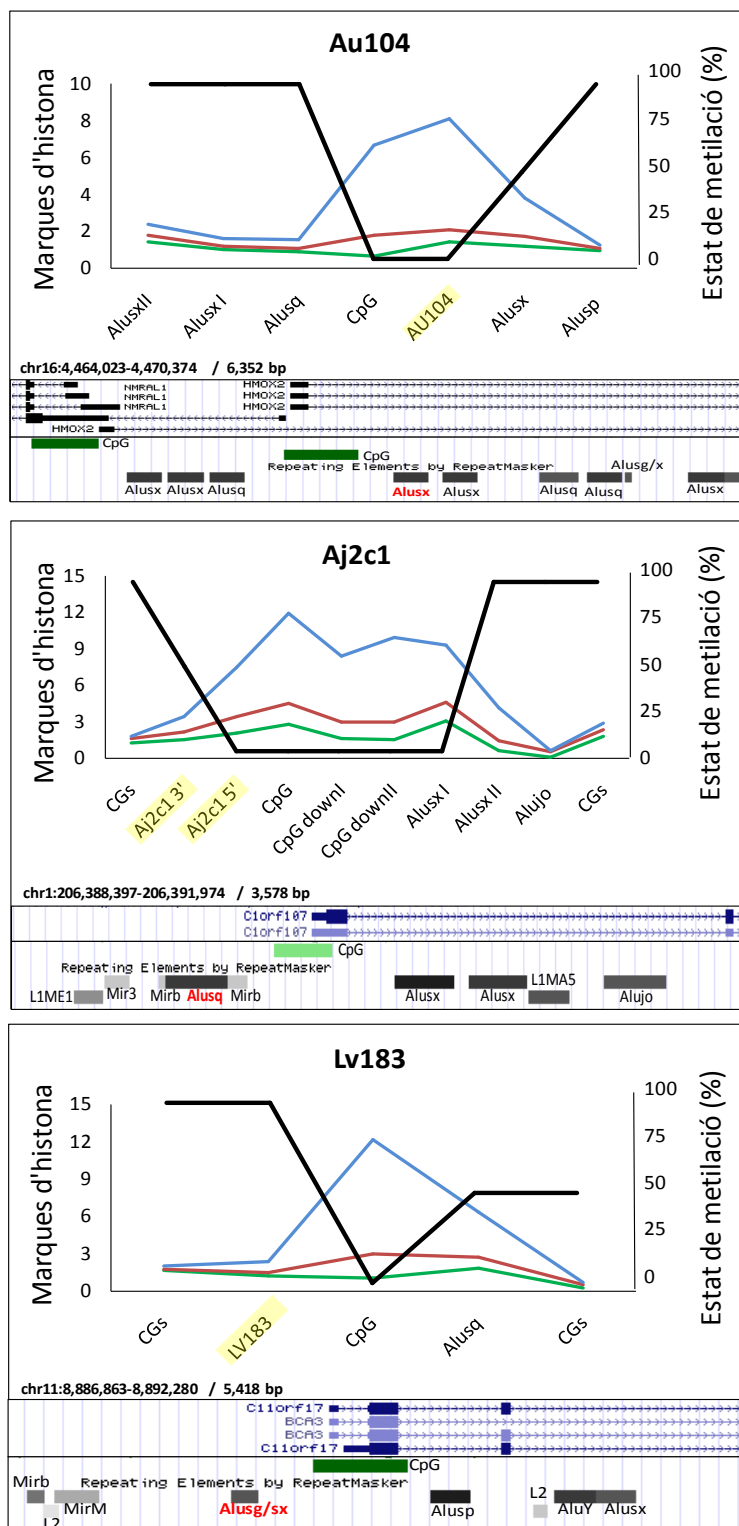


Fig. D.1. Mapa de la regió dels diferents elements Alu analitzats obtingut a partir del Genome Browser versió Mar. 2006 (NCBI36/hg18). Estudi fet amb la línia de CCR HCT116. En la gràfica estan representades les marques d'activació: en blau H3K4me3, en vermell Ach3K9 i en verd Ach3k14. En negre representem l'estat de metilació corresponent a cadascun dels elements de la regió.

#### 4. HIPOMETILACIÓ DELS ELEMENTS ALU EN CÀNCER

La majoria d'elements Alu estan metilats i transcripcionalment silenciats en cèl·lules somàtiques i es desmetilen amb freqüència en càncer. Aquesta desmetilació, però, no només afecta les seqüències Alu sinó que durant la tumorigènesi té lloc una hipometilació global del genoma (Ehrlich, 2002; Weisenberger et al., 2005). En aquest sentit, Hansen i col·laboradors (Hansen et al., 2011) van determinar mitjançant *Whole Genome Bisulphite Sequencing* que més de la meitat del genoma en càncer colorectal es desmetila, incloent seqüències úniques i repetitives.

Les implicacions de la hipometilació dels elements Alu en la tumorigènesi no han estat àmpliament estudiades, encara que hi ha certes evidències. Hi ha un gruix significatiu de treballs que relacionen la hipometilació d'elements transposables amb la inestabilitat genòmica (Daskalos et al., 2009; Kazazian, 2004). Per altra banda, també s'ha vist que els elements Alu podrien actuar com a llocs d'unió a factors de transcripció (Polak and Domany, 2006) o bé com a aïlladors genòmics (insulators) (Molto et al., 2009) promovent o impeding, segons l'estat de metilació, la unió de la maquinària transcripcional.

Són cada cop més els estudis que relacionen la hipometilació dels elements Alu amb càncer. Un d'ells, portat a terme en el nostre laboratori en càncer colorectal utilitzant la tècnica AUMA (Rodríguez et al., 2008b), ha donat lloc al desenvolupament d'aquest treball. Per altra banda, Xiang i col·laboradors (Xiang et al., 2010) van analitzar 48 mostres de carcinomes gàstrics i les seves parelles normals, així com 55 mostres de gastritis i diverses línies cel·lulars utilitzant la tècnica COBRA-DHPLC a fi de detectar l'estat de metilació de les Alus. Xiang i col·laboradors van observar que, mitjançant aquesta tècnica, quantificaven un gran nombre d'elements Alu hipometilats en carcinomes gàstrics. Un altre estudi portat a terme per Daskalos i col·laboradors (Daskalos et al., 2009) en càncer de pulmó mitjançant la piroseqüenciació, va determinar els nivells de metilació de seqüències Alu i LINE-1 de 48 carcinomes i els seus teixits adjacents. En ell observaven la reducció significativa dels nivells de metilació d'ambdós elements i van associar aquesta hipometilació a la inestabilitat genòmica.

Encara no s'ha elucidat ni s'ha trobat una resposta des del punt de vista mecanístic de si la hipometilació dels elements repetitius és causa o conseqüència de la inestabilitat genòmica. No obstant, es pot suggerir que la hipometilació dels elements transposables i la inestabilitat genòmica podria ser conseqüència de la reprogramació genètica i l'epigenètica en tumors, per exemple, la sobreexpressió de LINE-1 en sistemes models han demostrat que indueixen inestabilitat genòmica i delecions (Gilbert et al., 2007; Kazazian, 2004). Per altra banda Suzuki i col·laboradors (Suzuki et al., 2006) van estudiar la relació de les alteracions genètiques i epigenètiques en càncer gastrointestinal observant que la hipometilació s'acumulava de manera progressiva amb l'edat, cosa que d'altres alteracions genètiques no fan. Aquest fet posa de manifest que la hipometilació precedeix a l'aparició d'alteracions genètiques en tumors. Així doncs, l'acumulació progressiva d'alteracions en la metilació de cèl·lules que esdevindran tumorals serà la base d'algunes alteracions genètiques que s'aniran acumulant durant la progressió tumoral.



Tal i com comentaven Rodríguez i col·laboradors (Rodríguez et al., 2008b), les causes de la inestabilitat genètica no estan clares i poden tenir múltiples orígens. Tot i això, la correlació entre grau de desmetilació i alteracions cromosòmiques en càncers colorectals trobades en els seus estudis indiquen que la hipometilació podria tenir un pes important en la generació d'alteracions genètiques en tumors. No obstant, s'ha d'assumir que la hipometilació és un dels mecanismes que participa en la desestabilització del genoma tumoral però no l'únic.

#### 4.1. RESISTÈNCIA A LA DESMETILACIÓ DE CERTS ELEMENTS ALU

Els agents químics utilitzats en la majoria d'estudis experimentals d'epigenètica en càncer a fi de modular l'expressió dels gens han estat la 5-aza-2'-deoxicitidina (5-AzadC), un inhibidor de la metilació de l'ADN, i el TSA, un inhibidor de la desacetilació d'histones. S'ha vist que tant l'5-AzadC com el TSA poden dirigir-se preferentment cap a una metiltransferasa concreta o cap a una compactació de la cromatina i presenten efectes farmacològics no específics sobre l'expressió gènica, per això la inhibició de metiltransferases mitjançant un abordatge genètic sembla ser una metodologia experimental més rigorosa (Gius et al., 2004).

El mecanisme de desmetilació d'aquestes drogues epigenètiques així com de les HCT116-DKO(DNMT1-/-, DNMT3b-/-) és passiu, així doncs en cada ronda de replicació la metilació es redueix en absència de l'activitat de metilació de manteniment, a més s'ha vist que els patrons de desmetilació induïts per aquestes drogues són altament específics, no aleatoris, i reproduïbles. No obstant això, en el nostre estudi, i al analitzar cèl·lules tractades amb l'agent desmetilant 5-AzadC o modificades genèticament, en el cas de HCT116-DKO, vam veure que no totes les Alu metilades en HCT116 podien desmetilar-se totalment, suggerint la presència d'algun factor o necessitat biològica que fa que alguns elements Alu siguin més o menys susceptibles a desmetilar-se.

Els patrons de metilació en cèl·lules de mamífers estan regulats a través d'una complexa interacció d'almenys tres ADN metiltransferases (DNMTs) codificades independentment: DNMT1, DNMT3a i DNMT3b. Prèviament s'ha demostrat que DNMT1 i DNMT3b cooperen per mantenir la metilació de l'ADN i el silenciament de gens (Rhee et al., 2002) i l'alteració genètica d'ambdós inhibeix significativament l'activitat metiltransferasa i redueix la metilació de l'ADN genòmic un 95%. Malgrat això, i tot i el canvi en l'expressió gènica, certs estudis demostren que la inhibició de les DNMTs no provoquen un canvi de metilació global, sent certes regions resistents a desmetilar-se (Gius et al., 2004), coincidint amb el que observem en alguns dels elements Alu estudiats. Aquesta resistència a la desmetilació podria explicar-se per l'existència de certs *loci* resistents i que, al replicar-se, es "remetilen". Aquesta remetilació preferencial específica de *loci* suggereix la participació de factors moduladors de l'eficiència de metilació cel·lular després de la replicació (Hagemann et al., 2011).

Per altra banda, al analitzar les marques d'histona d'HCT116 i HCT116-DKO, vam veure que els elements Alu desmetilats en ambdues línies presentaven marques actives, a diferència dels elements metilats, tal i com caldria esperar. No obstant, en el cas de la línia HCT116-DKO, vam veure que, tot i desmetilar-se la majoria d'elements Alu, les marques d'histona activa només incrementaven en aquells elements situats a prop

d'inicis de transcripció, suggerint el paper regulador de les Alus situades prop de l'extrem 5' d'un gen i indicant que la inhibició de les DNMTs no és suficient per regular epigenèticament aquests elements.

## 4.2. PAPER DE L'ELEMENT AJ2C1 EN LA REGULACIÓ DE *DIEXF* I LA SEVA IMPLICACIÓ EN CÀNCER

En l'apartat 3.1 dels resultats estudiàvem l'element Aj2c1, que es va obtenir a partir de la tècnica AUMA, trobant-se hipometilat en tumors de càncer colorectal respecte la mucosa normal, la qual cosa suggeria la seva possible implicació en la tumorigènesi. A més, aquest element presentava una barrera interna metilació/demetilació que variava entre els diferents tipus cel·lulars estudiats, la qual cosa indicava una certa organoespecificitat del patró de metilació. La tria d'aquest element Alu Aj2c1 per l'estudi de la seva possible implicació en càncer va venir no només pel seu comportament i perfil epigenètic sinó també pel gen que tenia situat a -440pb, *C1orf107* o *DIEXF* (*digestive-organ expansion factor*), ja que s'ha descrit la seva associació amb la via de p53 i possible participació en el desenvolupament de l'aparell digestiu.

### 4.2.1. *DIEXF* I EL SEU PAPER EN CÀNCER

En Zebrafish s'havia vist que *DEF* (homòleg de *C1orf107/DIEXF* en humana) codificava per un factor pan-endoderm específic. Chen i col·laboradors (Chen et al., 2005) van veure que la pèrdua d'expressió de *DEF* provocava hipoplàsia en òrgans digestius i un augment de l'expressió de  $\Delta 113p53$ .  $\Delta 113p53$  és una isoforma de p53 produïda per un promotor alternatiu en l'intró 4 del gen p53 en humans (Murray-Zmijewski et al., 2006). L'augment d'expressió de  $\Delta 113p53$ , específica d'òrgans digestius anòmals, indueix l'expressió de gens associats a p53 activant l'aturada del cicle cel·lular però no l'apoptosi, i donant lloc a un fenotip hipoplàstic dels òrgans digestius.

Mentre que la inducció de l'expressió de p53 i/o les seves isoformes resulta crucial per suprimir el creixement cel·lular anòmal, la isoforma  $\Delta 113p53$  està estretament regulada pel factor *DEF*, específic d'òrgan/teixit, especialment durant l'organogènesi, prevenint la inhibició adversa del creixement en òrgans i teixits (Chen et al., 2009).

Segons la versió *Genome Browser*, *UCSC Genome Browser Mar. 2006 (NCBI36/hg18)*, i consultant l'Ensembl/Havana, el gen *DIEXF* té diversos transcrits descrits en humans, el 1, 2 i 3. La manca de correlació entre els nostres estudis d'expressió per RT-qPCR i Northern Blot ens van portar a pensar en la possibilitat de l'existència de nous transcrits, i a identificar, mitjançant la tècnica RACE, tres nous transcrits, el 4, 5 i 6. Per altra banda, del transcrit 2 no vam trobar cap evidència de la seva existència en les nostres mostres ni amb la tècnica de Northern Blot ni amb la tècnica RACE.

Excepte el transcrit 2, que com hem dit, no vam detectar en les nostres mostres, la resta de transcrits comparteixen els mateixos exons codificants i es diferencien bàsicament en les regions UTRs dels extrems 3' i 5'.

L'extrem 3' de la majoria dels pre-ARNm de gens codificants de proteïnes i lncRNAs (de l'anglès *Long non-coding RNA*) és tallat i poliadenilat, procés necessari per l'exportació nuclear, l'estabilitat dels ARNm madurs i l'eficiència de traducció. Descobriments recents han posat de manifest que una gran proporció de gens conté més d'un lloc de poliadenilació, indicant que la poliadenilació alternativa (APA) és un fenomen molt generalitzat. L'APA dona lloc a ARNm amb finals 3' alternatius, contribuint a la complexitat del transcriptoma mitjançant la generació d'isoformes que difereixen ja sigui en la seva seqüència codificant o en els seus UTR3' (Elkon et al., 2013). La longitud de l'UTR3' mostra una variació considerable en els ARN missatgers dels mamífers, oscil·lant de 60 fins a 4000 nucleòtids. Aquest fet posa de manifest que el genoma humà ha evolucionat per incrementar els mecanismes de control postranscripcional en l'expressió dels gens. S'han identificat més de 2000 seqüències curtes de nucleòtids en l'extrem UTR3' on s'uneixen diferents elements reguladors de l'ARNm, participant en el transport, degradació o estabilització i traducció de l'ARNm.

En el gen *DIEXF* vàrem observar l'existència de 5 llocs de poliadenilació que corresponien a les isoformes ja descrites (transcrit 1 i 3) i a les noves isoformes identificades per RACE (transcrit 4, 5 i 6), i que donaven lloc a UTR3' diferents, sent el del transcrit 3 el més llarg, concretament uns 5500 pb més respecte els altres transcrits. Els UTR3' serveixen com a principals connectors de factors que controlen l'estabilitat, la localització i l'eficiència de traducció (com per exemple microARNs (miARN) i proteïnes d'unió a ARN) (Elkon et al., 2013). Tot i que la majoria d'articles en aquest camp solen parlar d'un escurçament de l'UTR3' associat a una major expressió a nivell de proteïna donat que isoformes més curtes sovint presenten més estabilitat i perden les dianes a microARN (mediadors de la repressió) (Lin et al., 2012; Mayr and Bartel, 2009), segons els nostres resultats, sembla ser que el transcrit 3 del gen *DIEXF*, que té el UTR3' més llarg, seria l'únic transcrit codificant, tal i com demostren els resultats obtinguts en comparar els nivells d'ARNm de cada transcrit i els nivells de proteïna. A més, també es va silenciar l'expressió de *DIEXF* mitjançant siRNAs i vam veure que els siRNAs específics del UTR3' llarg del transcrit 3 eren suficients per reduir els nivells de proteïna. Cal recordar que la cua poli (A) facilita el transport de l'ARN missatger al citoplasma i el fa més resistent a la degradació de les exonucleases cel·lulars endògenes. Donada la importància de la llargada de la cua poli (A) potser la resta de transcrits generats tenen una cua poli (A) massa curta i pateixen una degradació enzimàtica o són emmagatzemats en un estat inactiu (Elkon et al., 2013).

Els diferents transcrits de *DIEXF* no només difereixen en l'extrem 3' sinó també en el 5', presentant diferents TSS (de l'anglès *transcription start site*), donant lloc a diferents UTR5'. Segons el que hi ha publicat, el TSS del transcrit 1 comença 24 pb més *upstream* que el TSS del transcrit 3. A l'analitzar per RACE l'extrem UTR5' vàiem una gran variació de seqüències segurament causada per la presència de múltiples TSS corresponents als diferents transcrits identificats. És interessant destacar que el TSS majoritari que apareixia en les Sw480 no es detectava en les hESC, suggerint que es tractava del TSS del transcrit 3, present en Sw480 però no en hESC.

Cada vegada és major el nombre de treballs que indiquen que un mateix gen pot tenir diferents TSS. Kawaji i col.laboradors (Kawaji et al., 2006) van veure que l'inici de transcripció pot tenir lloc en múltiples posicions dins el *core promoter*, de manera que molts promotors no tindrien un únic TSS sinó una sèrie d'inicis de transcripció situats uns molt propers als altres (a certs pb), aquest concepte dista de la definició dels

anomenats “promotors alternatius” en els que els *core promoters* solen estar separats per una certa distància genòmica (a certes kb). Mitjançant l’anàlisi CAGE (de l’anglès *Cap Analysis of Gene Expression*) en 22 teixits van identificar 4 tipus diferents de promotors múltiples segons la distribució dels TSS: un únic pic dominant, una ampla distribució general, una distribució bimodal o multimodal i una distribució àmplia amb un pic majoritari (Figura D.2).

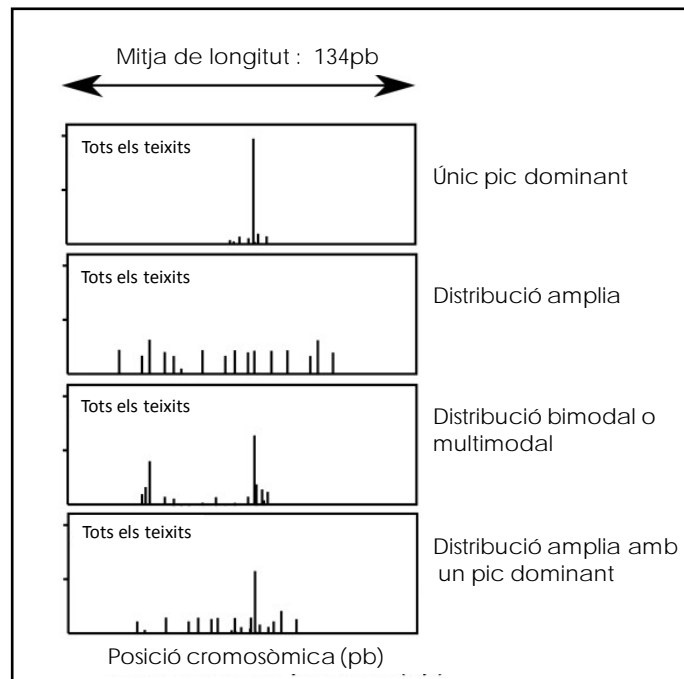


Fig. D.2. Tipus de promotors múltiples establerts segons la distribució de TSS. Les agrupacions es classifiquen en quatre classes basades en el recompte CAGE de tots els teixits. Els recomptes es mostren per histogrames, on l'eix x indica les coordenades genòmiques o localització cromosòmica, i l'eix y indica els recomptes totals dels tags CAGE. Figura adaptada de l'article de Kawaji i col·laboradors (Kawaji et al., 2006)).

Aquest fenomen dona lloc a un increment de la complexitat del transcriptoma i permet una especificitat de teixit en l'ús dels diferents TSS. En la figura D.3. es mostren diversos casos de l'ús dels TSS depenent del teixit: en alguns casos s'utilitzaria el mateix TSS independentment del teixit (Figura D.3.a) i en altres casos hi hauria un ús del TSS específic de teixit (Figura D.3.b. i c). Segons aquesta hipòtesi seria possible que una cèl·lula, segons el context en el que es trobés, modifiqués el seu lloc d'inici de transcripció, la qual cosa podria explicar el que observem en el gen *DIEXF*. Així doncs segons els resultats obtinguts amb el 5'-RACE, i d'acord amb aquest model, semblaria que *DIEXF* es comportaria com la figura D.3.c, on SW480 seria el teixit A ja que presentaria tant el TSS corresponent al transcrit 1 com el TSS del transcrit 3 (específic de teixit i 24 pb downstream de l'altre) i, en canvi, hESC seria el teixit B, presentant un TSS, el del transcrit 1, comú en ambdues mostres.

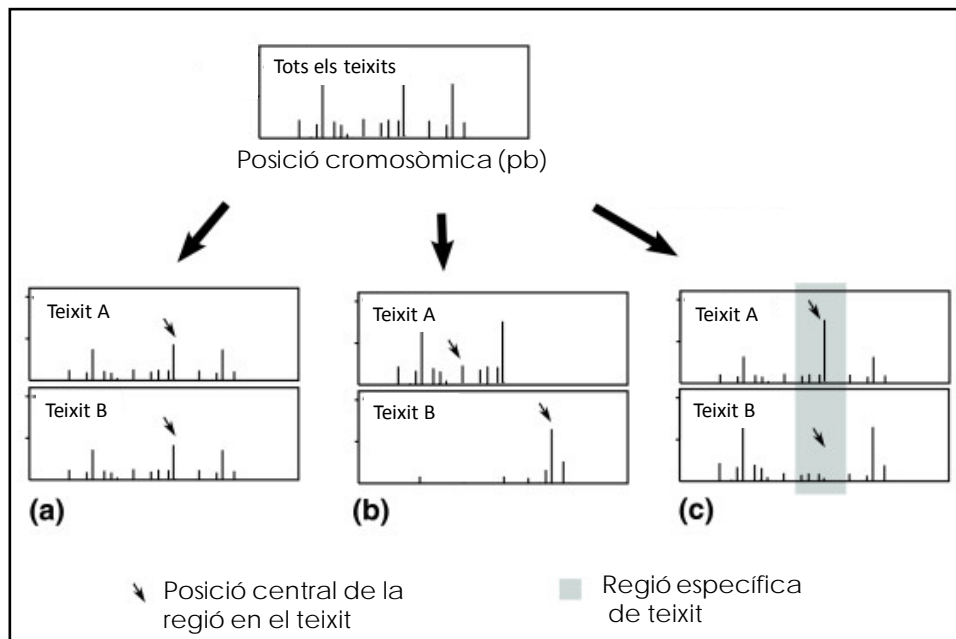


Fig. D.3. Possibles TSS segons el teixit. (a) El TSS s'utilitza d'igual manera en el teixit A i B, (b) El teixit A tendeix a utilitzar regions més upstream que el teixit B i (c) Algunes subregions mostren diferent ús dels TSS segons el teixit. Figura adaptada de l'article de Kawaji i col·laboradors (Kawaji et al., 2006).

L'ús de múltiples TSS podria estar regulat epigenèticament, ja sigui per la metilació de l'ADN, per modificacions d'histones o per remodelació de la cromatina (Kawaji et al., 2006). Com ja hem comentat, l'Alu Aj2c1 situada prop del TSS del gen *DIEXF* presenta un patró de metilació diferent segons el tipus cel·lular, la qual cosa podria regular l'ús dels diferents TSS. Així doncs, l'element Aj2c1 desplaçaria el TSS segons el seu estat de metilació permetent o no la transcripció segons el tipus cel·lular o el context biològic en el que es trobés la cèl·lula.

#### 4.2.2. AJ2C1 ACTUA COM A BARRERA DE METILACIÓ, REGULANT LA CONFORMACIÓ DE LA CROMATINA I AFECTANT L'EXPRESSIÓ DE *DIEXF*

L'element Aj2c1 es va escollir per aprofundir en el paper dels elements Alu durant la tumorigènesi ja que es va trobar hipometilat en tumors de càncer colorectal amb una elevada recurrència. Com ja hem comentat, Aj2c1 és un Alu Sq, situat en l'extrem 5' del gen *DIEXF*, concretament a uns -450 pb dels múltiples TSS identificats. Atès la proximitat a aquest gen important en l'organogènesi de l'aparell digestiu, ens vam proposar investigar el paper de l'Alu en la regulació de l'expressió de *DIEXF*.

L'element Aj2c1 pertany a la família Alu S i està insertada dins d'una seqüència Mirb. Aj2c1 està col·locat en posició invertida. S'ha observat que, normalment, la posició de l'element dins el genoma no afecta al seu rol dins d'aquest. No obstant, els elements Alu invertits s'ha vist que mostren patrons d'inscripció diferents als elements directes, suggerint diferents mecanismes d'integració. També s'ha observat que els elements Alu invertits poden representar una font d'instabilitat cromosòmica (Cook et al., 2011). En l'apartat de resultats mostràvem diversos estudis epigenètics i d'expressió per tal d'esbrinar el paper de l'element Aj2c1 en la regulació del gen *DIEXF*. En primer lloc es va analitzar l'estat de metilació de l'Alu no només en còlon normal i tumoral sinó també en altres tipus cel·lulars, el que va permetre veure que el patró de metilació era específic de tipus cel·lular. A més, també es van analitzar marques d'histona i pel que fa a les marques actives vàiem que a mesura que l'element Alu es desmetila aquestes es feien presents. Al analitzar la regió vam veure com Aj2c1 marcava una barrera de metilació que variava segons el tipus cel·lular, determinant l'obertura de la cromatina (Figura R.18). Així doncs, observàvem que quan més metilada estava l'Alu Aj2c1, més petit era el domini de cromatina activa (seria el cas de hESC2 i HFF) i a mesura que la regió s'anava desmetilant augmenta el domini de cromatina activa (HCT116 i DKO).

Al correlacionar l'estat epigenètic de l'Alu i l'expressió de *DIEXF*, tot apunta a que la regulació del gen, concretament del transcrit codificant 3, ve determinada per la metilació de l'Alu. Vàiem doncs que tots els transcrits no codificants s'expressaven en totes les mostres analitzades, independentment de l'estat de metilació de l'Alu, mentre que només hi havia expressió del transcrit 3 codificant, i per tant de la proteïna, quan l'element Alu Aj2c1 estava totalment desmetilat o bé quan la regió 3' de l'Alu es trobava altament desmetilada (Taula R.13). En Sw480 i LoVo, tot i tenir part de l'Alu desmetilada, aquesta no presentava marques actives, però el gen s'expressava. Aquest escenari suggereix que la desmetilació del element Alu seria el pas que desencadenaria un desplaçament del TSS, permetent la unió de factors de transcripció i la conseqüent transcripció gènica. Com hem vist, els resultats de RACE, ens van permetre veure l'existència de múltiples TSS de *DIEXF* i que el transcrit 3 en concret presentava un TSS específic i diferent de la resta. Segons aquests resultats semblaria que l'estat de metilació de l'element Aj2c1 controlaria l'inici de transcripció del transcrit 3.

Com ja hem comentat, la complexitat dels mecanismes de regulació gènica no es coneixen completament i malgrat que els nostres resultats no acaben d'aclarir el mecanisme d'acció de l'element Aj2c1 i per tant calen molts més estudis, podríem postular que Aj2c1 actuaria com a font de llocs d'unió a factors de transcripció o altres complexos proteics que determinessin l'ús o no del TSS del transcrit 3, facilitant a més

l'estat de la cromatina oberta. En la seqüència d'Aj2c1 s'han predit múltiples llocs d'unió a factors de transcripció, entre ells trobem Sp1 (involucrat en desenvolupament) i SRF (implicat en cicle cel·lular, apoptosi, diferenciació i creixement). En aquesta tesi no hem tractat aquest punt però seria interessant trobar què s'uneix o no a Aj2c1 segons el seu estat de metilació així com fer diversos estudis per determinar el mecanisme que regula l'ús dels múltiples inicis de transcripció.

Per altra banda, i segons els resultats observats, l'expressió de *DIEXF* aniria associada, de manera específica a un context tumorigènic. Com ja hem comentat el gen *DIEXF* està associat a la via de p53 i a l'aparell digestiu. En l'apartat de resultats confirmàvem el rol de Aj2c1, així com de *DIEXF*, en la tumorigènesi al consultar, en diverses parelles normal-tumor, l'estat de metilació de l'element Alu en les dades del TCGA (*The Cancer Genome Atlas*) i l'expressió del gen *DIEXF* en la bases de dades de l'*Oncomine*.

Segons l'estat de metilació es van poder classificar en 4 perfils de comportament (Figura R.42). Dins dels 4 grups de la classificació vèiem dos tipus ben definits segons l'estat de metilació. El grup 1, 2 i 3 formaria part del primer tipus, majoritari en els diferents càncers estudiats i on l'element Alu s'hipometila en tumors i l'illa CpG roman desmetilada, en aquests grups vèiem una sobreexpressió del gen *DIEXF*. Dins d'aquests grups trobem d'adenocarcinoma de còlon (grup1), el carcinoma de mama, pròstata i pulmó (grup 2), el carcinomes d'uter, ronyó i glioblastomes (grup 3).

El grup 4 (on trobem els glioma cerebral de baix grau) formaria part del segon tipus: hipermetilació en tumor tant de l'element Alu com de l'illa CpG. Del grup 4, no obstant no hem trobat cap cas que ens confirmi una infraexpressió de *DIEXF*.

Normalment aquesta sobreexpressió es veia tant en la sonda específica pels transcrits 1, 3, 4, 5 i 6 com per les específiques del transcrit 3. Tot i que en certs casos la sobreexpressió era específica del transcrit 3. Altres tipus de tumors de diferents teixits mostren un patró heterogeni fent que no encaixessin en cap dels grups establerts, és el cas del carcinoma de tiroides així com els adenocarcinomes de pròstata, estómac, pulmó, pàncrees i recte.

La informació trobada en totes aquestes bases de dades tant a nivell d'expressió com de metilació en mostres normals i tumorals ens ajuda a confirmar, de nou, la relació entre l'estat de metilació de l'element Alu Aj2c1 i l'expressió del gen *DIEXF*.

## 5. ELS ELEMENTS ALU DURANT LA DIFERENCIACIÓ

La diferenciació cel·lular implica l'activació i repressió de milers de gens d'una cèl·lula mare a fi de donar una cèl·lula diferenciada. Les cèl·lules mare embrionàries pluripotents posseeixen la capacitat de diferenciar-se en múltiples llinatges cel·lulars en el cos. En aquesta reprogramació juga un paper important l'epigenètica, tant a nivell de metilació de l'ADN com a nivell de cromatina. En les cèl·lules mare, l'epigenoma és molt mal·leable, receptiu i versàtil, a diferència de les cèl·lules somàtiques (Meshorer and Misteli, 2006; Mohn and Schubeler, 2009). Trobem molts estudis en aquest camp, però la majoria d'ells s'han centrat en seqüències úniques.

En mamífers s'ha vist que un subconjunt dels promotors que contenen illes CpG pateixen metilacions *de novo* dependent del llinatge cel·lular, causades per factors específics de seqüència i modificacions epigenètiques. No obstant això, encara no s'han aclarit els mecanismes de la metilació dinàmica de l'ADN. Lienert i col·laboradors van observar que els promotors proximals mostraven autonomia pròpia per definir els seu estat de metilació en cèl·lules pluripotents i capacitat de reprogramar aquests estats de metilació durant la diferenciació cel·lular (Lienert et al., 2011).

En l'apartat de resultats vam determinar l'estat de metilació d'un grup d'elements Alu en cèl·lules mare, cèl·lules pluripotents induïdes i cèl·lules diferenciades (queratinòcits, fibroblasts, leucòcits i cèl·lules de còlon) (Taula R.6). Trobàvem certs elements que mantenien el seu patró desmetilat en tots els tipus cel·lulars així com una forta presència de marques d'histona actives al analitzar tant cèl·lules mare embrionàries com fibroblasts (Figura R.13). Aquest fet, com ja s'ha comentat anteriorment, ens fa pensar en la seva importància biològica. D'altres estaven metilats en cèl·lules mare embrionàries i es mantenien metilats en les cèl·lules diferenciades analitzades, i al analitzar la cromatina vam observar que no presentaven marques actives. Com ja hem anat veient la metilació de l'ADN és una marca repressora i el fet que aquests elements es mantinguin silenciatos en els diferents contextos biològics indicaria la forta pressió epigenètica a la que estan sotmesos.

Per altra banda, vam observar certs elements metilats en cèl·lules mare embrionàries i cèl·lules pluripotents induïdes però que es desmetilaven parcialment o totalment en algunes cèl·lules diferenciades. Resulta interessant que aquests elements Alu, tot i estar metilats en cèl·lules mare embrionàries presentaven marques d'histona actives que es mantenien en els fibroblasts, on es desmetilaven. Com s'ha vist al llarg del treball, les marques actives AcK9H3 i H3K4me3 estan associades específicament a Alus desmetilades, però el fet de trobar Alus amb marca activa en cèl·lules mare embrionàries tot i estar metilades indicaria una "preparació" de la cromatina, fent-la més accessible i donant-li versatilitat a l'hora de diferenciar-se, similar al que ocorre amb els dominis bivalents (Bernstein et al., 2006). Aquest fenomen només l'hem observat en uns pocs elements Alu i caldria realitzar altres anàlisis, com per exemple ChIP-Seq, a fi de determinar si es tracta d'un mecanisme més generalitzat en cèl·lules mare embrionàries. En aquest cas, la bivalència vindria donada per la metilació dels elements Alu i les marques AcK9H3 i/o H3K4me3, de manera, que durant la diferenciació, en aquelles cèl·lules on l'Alu tingués un paper funcional es desmetilaria i



es mantindrien les marques actives mentre que en les cèl·lules on l'Alu hagués d'estar reprimida, es perdrien les marques actives i es mantindria la metilació.

Un altre element Alu que vam identificar diferencialment metilat va ser l'element ZALU3, desmetilat en cèl·lules mare embrionàries i pluripotents induïdes i metilat en la majoria de les cèl·lules diferenciades estudiades. ZALU3 presentava marca d'histona activa (H3K4me3) i inactiva (H3K27me3) en cèl·lules mare embrionàries mostrant la bivalència característica d'aquestes cèl·lules mare descrita en múltiples treballs (Bernstein et al., 2006; Mohn and Schubeler, 2009). Al diferenciar-se i metilar-se, vèiem com només es mantenia la marca inactiva H3K27me3.

## 5.1. PAPER DE L'ELEMENT ZALU3 EN LA REGULACIÓ DE GLDC I LA SEVA IMPLICACIÓ EN EL PROCÉS DE DIFERENCIACIÓ

L'element ZALU3 es va obtenir més recentment, a partir de la tècnica NSUMA. En l'apartat 3.2 dels resultats estudiàvem l'element ZALU3 i el gen pròxim *GLDC*. Vam considerar ZALU3 interessant ja que es trobava altament desmetilat en cèl·lules mare embrionàries i metilat en la majoria de cèl·lules diferenciades analitzades en el NSUMA. ZALU3 es tracta d'un Alu Sc, situat a -1.064 pb del TSS del gen *GLDC* (*glycine dehydrogenase*), un gen relacionat amb el metabolisme de la glicina (Kanno et al., 2007; Zhang et al., 2012).

### 5.1.1. *GLDC* I EL SEU PAPER EN EL METABOLISME DE LA GLICINA

Abans d'entrar en detall en la regulació de l'expressió de *GLDC* és important entendre el paper de *GLDC* dins del metabolisme de la glicina (Figura D.4).

La glicina és un aminoàcid nutricionalment "no essencial" i la seva principal funció és com a precursor de proteïnes i proporcionar la subunitat C<sub>2</sub>N a les purines. També participa en la biosíntesi del col·lagen, del grup hemo de l'hemoglobina i d'altres porfirines i és un neurotransmissor inhibitori del sistema nerviós central (Petrat et al., 2012) (Petrat et al.; Wang et al., 2013).

La glicina es pot sintetitzar a través de diferents vies tot i que la principal és a partir de la serina seguint una important reacció reversible catalitzada per la serina hidroximetiltransferasa (SHMT). Pel que fa a la degradació, existeixen tres vies, essent la més important la catalitzada pel sistema d'escissió de la glicina o GCS (de l'anglès *Glycine Cleavage System*) que degrada la glicina a CO<sub>2</sub>, amoníac, NADH i un grup metilè, acceptat pel tetrahidrofolat (Wang et al., 2013). GCS catalitza una reacció reversible, però és important mencionar que degut a les concentracions intracel·lulars de substrats i productes i a la seva cinètica enzimàtica, s'afavoreix la degradació de la glicina. A més a més, la reacció de síntesi de glicina mitjançant GCS és un procés altament reductor, de manera que en condicions fisiològiques aeròbiques no es dona (Kikuchi et al., 2008).

GCS és un complex enzimàtic mitocondrial format per 4 proteïnes (P, H, T i L). *GLDC* o *glycine dehydrogenase (decarboxylating)* codifica la proteïna P (*pyridoxal phosphate*-

*dependent glycine decarboxylase*) d'aquest complex enzimàtic, la qual posseeix una activitat glicina descarboxilasa dependent de fosfat piridoxal i s'expressa en cervell, ronyó, fetge i placenta. La presència o l'absència de la proteïna P és clau per determinar els teixits actius o inactius en la degradació de la glicina, però no es coneix la importància fisiològica dels diferents nivells d'activitat GCS en els diferents teixits. S'ha vist que el complex GCS s'expressa durant l'embriogènesi (Tibbetts and Appling, 2010). Es creu que la regulació de l'expressió dels gens que codifiquen pels diferents components de GCS pot ser crucial per al metabolisme de la glicina. En aquest sentit, els nostres resultats indiquen que l'expressió de *GLDC* estaria regulada per l'estat de metil·lació de *ZALU3*.

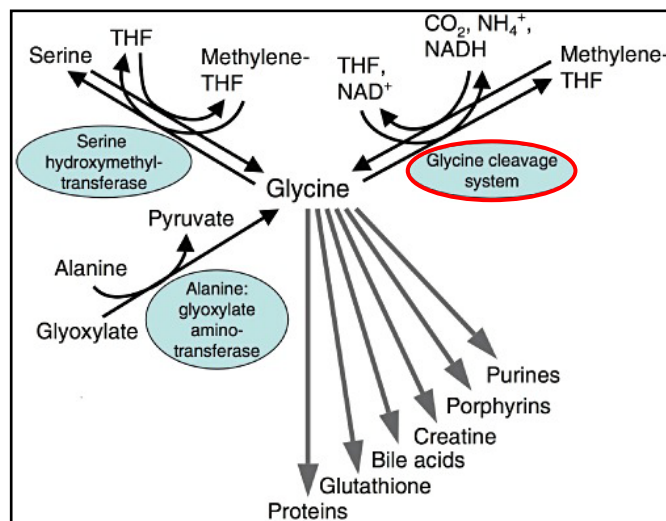


Fig. D.4. Metabolisme de la Glicina. En vermell marquem el GCS (Glycine Cleavage System) del que forma part la proteïna P, codificada per *GLDC*. Figura adaptada de (Petrat et al., 2012).

### 5.1.2. ALTERACIONS EN *GLDC*

S'ha vist que la deficiència en *GLDC* és causa d'hiperglicèmia no cetòsica (NKH, de l'anglès *Nonketotic Hyperglycemia*), també coneguda com encefalopatia per glicina. Aquesta malaltia autosòmica recessiva es caracteritza per una gran acumulació de glicina en els fluids corporals i per símptomes neurològics severos (Kanno et al., 2007). Altres malalties associades a una mutació en *GLDC* són: retard mental, diabetis i anomalies congènites.

El 80% de les alteracions que pateix *GLDC* són mutacions o delecions en l'extrem 5' del gen (Wongsurawat et al., 2012). La regió 5' de *GLDC* és rica en R-loops i en elements Alu. S'han detectat alguns casos de NKH associats a delecions en *GLDC* causades per recombinacions dutes a terme per elements Alu. Un total de 120 còpies d'elements Alu es troben en el gen *GLDC*, el qual té una llargada de 113 kb. Això implica una elevada densitat d'elements Alu en aquesta regió. Aquest no és el primer cas de malalties associades a recombinacions/delecions dutes a terme per Alus (Kanno et al., 2007). No

obstant, ZALU3 no es troba dins del gen, sinó a més de 1.000 pb de l'inici de transcripció i el paper que jugaria seria a nivell de regulació transcripcional tal i com discutirem més endavant.

### 5.1.3. GLDC I EL SEU PAPER EN CÀNCER

En un treball recent es proposa GLDC com un enzim oncogènic (Zhang et al., 2012). En l'estudi portat a terme per Zhang i col·laboradors (Zhang et al., 2012) en càncer de pulmó van observar un subgrup de cèl·lules canceroses que sobreexpressaven GLDC. Aquest subgrup es va identificar com a TIC (de l'anglès *Tumor Initiating Cells*), es tracta de cèl·lules que expressen marcadors de cèl·lules mare i són les responsables del manteniment i la progressió tumoral (Benayoun et al., 2012). En aquest estudi van generar un *knock-out* de GLDC en cèl·lules de càncer de pulmó (A549) i també van emprar les cèl·lules d'adenocarcinoma de còlon CaCo2; al suprimir GLDC veien una reducció de la proliferació cel·lular així com una reducció en el potencial tumoral d'aquestes cèl·lules. Així doncs la sobreexpressió de GLDC s'associaria a cèl·lules mare canceroses, al menys en càncer de pulmó.

El concepte de les cèl·lules mare canceroses o CSCs (de l'anglès *Cancer Stem Cells*) ha adquirit molta importància en la recerca oncològica dels últims anys. Les CSC es defineixen com una petita subpoblació de cèl·lules tumorals que tenen la capacitat d'iniciar un tumor i de conduir la proliferació neoplàstica. Les cèl·lules CSC tenen capacitat de produir un nombre il·limitat de còpies de si mateixes mitjançant l'autorenovació i potencial per donar lloc a qualsevol cèl·lula dins d'un tumor mitjançant el procés de diferenciació. Així doncs les CSC presenten propietats similars a les de les cèl·lules mare, com la capacitat d'autorenovar-se i diferenciar-se (Suva et al., 2013).

Coincidint amb els resultats de Zhang i col·laboradors, en els nostres estudis vam observar que les cèl·lules d'adenocarcinoma de còlon humà, CaCo2 expressaven GLDC. Morfològicament, els cultius d'aquesta línia són molt heterogenis i contenen subpoblacions de cèl·lules amb diversa capacitat de diferenciació. Així doncs, en condicions estàndards de cultiu, la població CaCo2 conté un gran nombre de cèl·lules morfològicament indiferenciades, en canvi sota condicions d'estrès, com la deprivació de glucosa o sèrum, aquestes cèl·lules es diferencien i polaritzen, exhibint moltes característiques i funcions de les cèl·lules diferenciades de l'epiteli de la mucosa intestinal. Aquestes cèl·lules han estat emprades en múltiples treballs per realitzar estudis *in vitro* relacionats amb la funció i la diferenciació de les cèl·lules intestinals (Yang et al., 2013). En el nostre laboratori i sota les condicions adequades de cultiu *in vitro* es van diferenciar i polaritzar aquestes cèl·lules i vam poder observar com es reduïa l'expressió de GLDC (Figura R.56).

Per altra banda Mohit Jain i col·laboradors (Jain et al., 2012) van observar que el patró de consum de glicina així com els enzims relacionats amb la seva biosíntesi estan estretament relacionats amb la velocitat de divisió de les cèl·lules canceroses. Les cèl·lules canceroses que es divideixen ràpidament consumeixen altíssimes quantitats de glicina, mentre que les que presenten velocitats molt baixes exporten aquest aminoàcid a l'exterior de la cèl·lula. Al desproveir de glicina a les cèl·lules canceroses,

eliminant l'aminoàcid del medi de cultiu o bloquejant els enzims implicats en el seu metabolisme), les cèl·lules amb una ràpida divisió reduïen el seu ritme de proliferació, mentre que aquelles que proliferaven poc a poc no modificaven el seu patró de divisió. Aquests estudis també s'han confirmat *in vivo*, on una major expressió d'aquests enzims implicats en el metabolisme de la glicina s'associen a un pitjor pronòstic en càncer de mama (Jain et al., 2012). Així doncs sembla que la glicina té un paper destacat en la tumorigènesi i depenent del tipus de càncer s'alteren determinats gens afectant a les vies de síntesi i degradació de la glicina.

#### 5.1.4. ZALU3 REGULA L'EXPRESSIÓ DE *GLDC* I ACTUA COM A AÏLLADOR GENÒMIC

Tal i com hem comentat, l'element ZALU3 es va escollir per aprofundir en el paper dels elements Alu durant la diferenciació. Es tracta d'un Alu Sc (família més freqüent i d'edat intermèdia), situat en l'extrem 5' del gen, a -1.064pb del TSS del gen *GLDC* (*glycine dehydrogenase*).

En l'apartat de resultats mostràvem els diversos estudis epigenètics i d'expressió per tal d'elucidar el paper de l'element ZALU3 en la regulació del gen *GLDC*. El patró d'expressió venia aparentment determinat per la metilació de ZALU3. *GLDC* només s'expressa quan l'element ZALU3 es troba totalment desmetilat, és el cas de les cèl·lules mare embrionàries, les cèl·lules pluripotents induïdes, els hepatòcits i la línia de CCR CaCo2 (Taula R.14).

Al analitzar diferents temps de diferenciació de les cèl·lules mare cap a *embrioid bodies* (EBs), vàiem que la metilació de ZALU3 ocorria ja a dia 2, moment en el qual *GLDC* deixava d'expressar-se. Pel que fa a la cromatina, les cèl·lules mare embrionàries, les quals expressaven *GLDC* i tenien ZALU3 desmetilat, presentaven no només marques d'histona actives sinó també la marca inactiva H3K27me3, mostrant una bivalència característica de les cèl·lules mare. Aquesta bivalència no només es dona en l'Alu sinó en tota la regió promotora. Com ja hem comentat, diversos estudis *a escala genòmica* de la metilació de l'ADN i les modificacions d'histones revelen estats dinàmics d'aquestes marques epigenètiques en els promotors i les illes CpG (Bernstein et al., 2006; Mohn and Schubeler, 2009). Al llarg del temps de diferenciació observàvem com la marca activa H3K4me3 començava a baixar a dia 2, sobretot en l'element ZALU3 i vàiem com el domini de cromatina activa s'anava fent més estret al llarg de la diferenciació. Pel que fa a la marca inactiva H3K27me3, vàiem que es mantenia al llarg de la diferenciació i que a dia 12 aquesta s'estenia al llarg de la regió (Figura R.54). Per tant, tot indica que la metilació de l'Alu seria el primer o un dels primers esdeveniments de la disminució de l'expressió del gen *GLDC*.

Pel que fa a les cèl·lules diferenciades no vam observar bivalència. Així doncs, quan *GLDC* no s'expressava i l'Alu estava metilada vàiem presència de marca H3K27me3, com en el cas d'HFF. En canvi quan *GLDC* s'expressava i l'Alu estava desmetilada observàvem marca activa, com en el cas de la línia d'hepatòcits (en aquest cas basant-nos les dades del ENCODE ja que no es va fer un anàlisi per CHIP d'aquestes línies).

De la mateixa manera, en les cèl·lules tumorals vam observar molta presència de marca activa quan *GLDC* s'expressava i l'Alu estava desmetilada (CaCo2). No

obstant, en aquells casos on *GLDC* no s'expressava, com en HCT116 i Sw480, la regió estava altament metilada i no presentava marques actives o nivells molt baixos, ni tampoc H3K27me3. En el cas de DKO, on observàvem una lleu expressió de *GLDC*, vèiem que la presència de marca activa augmentava respecte a HCT116, però sense assolir els nivells presents en CaCo2, corresponent-se a la desmetilació que patia la regió. Com ja hem comentat, DKO es tracta d'una línia deficient en DNMTs (DNMT1 -/-, DNMT3b -/-), fet que provoca una desmetilació global del seu genoma, suggerint que la desmetilació seria la que facilitaria l'obertura de la cromatina, fent-la més accessible.

Així doncs, l'estat epigenètic de la regió regularia l'expressió del gen *GLDC*, sent l'element Alu ZALU3 el que marcaria una barrera en l'estat de metilació, jugant un paper com a aïllador genòmic, i reclutant possiblement factors de transcripció i/o altres proteïnes que participarien en la regulació del gen. Com en el cas de *Aj2c1*, en la regió de ZALU3 s'han predit múltiples llocs d'unió a factors de transcripció, entre ells trobem Sp1 (involucrat en desenvolupament) i SRF (implicat en cicle cel·lular, apoptosi, diferenciació i creixement).

En aquest cas sembla que la metilació de ZALU3 jugaria un paper dins la diferenciació regulant l'expressió de *GLDC* en cèl·lules mare embrionàries i en òrgans com el cervell, el ronyó, el fetge i la placenta, on jugaria un paper en la síntesi i degradació de glicina.

## 6. IMPLICACIÓ DELS ELEMENTS ALU EN LA REGULACIÓ GÈNICA

Molts estudis a escala genòmica han ignorat elements Alu a causa de les dificultats tècniques en l'anàlisi d'aquests elements repetitius. No obstant, l'abundància d'elements Alu en genoma humà demostra que han tingut un gran impacte en la mutagènesi causada per la seva inserció i en l'evolució del genoma dels primats. La seva distribució en tot el genoma ha agreujat l'impacte, substituint seqüències primàries per esdeveniments de recombinació homòloga no-al·lèlics en tot el genoma. Diversos estudis de seqüenciació han mostrat que aquestes formes d'inestabilitat no només han donat lloc a importants canvis evolutius en els genomes, sinó segueixen causant diversitat dins la mateixa espècie i contribueixen en certes malalties. La ubiqüitat d'elements Alu en el genoma, i la seva presència en zones riques en gens, també els ha portat a participar, tal i com hem vist, sobre l'expressió i regulació gènica.

Com ja s'ha dit, els elements Alu són transcrits per l'ARN polimerasa III, però no codifiquen cap proteïna funcional sinó que utilitzen la transcriptasa inversa dels elements LINE-1, per fer còpies d'ADNc i retrotransposar-se de nou en el genoma de l'hoste (Dewannieux et al., 2003). Encara que les insercions Alu poden ser mutagèniques (Deininger and Batzer, 2002), la gran majoria d'Alus integrats no tenen cap influència aparent sobre el genoma. No hi ha cap funció òbvia pels elements Alu, però la seva presència en el genoma s'ha implicat en diversos processos biològics incloent la recombinació ectòpica, la creació de nous exons, i la donació de nous elements reguladors (Deininger and Batzer, 2002; Krehling and Graveley, 2004).

Cada vegada que un element Alu s'insereix en un gen, o prop d'ell, té el potencial d'influir en l'expressió d'aquest per diversos mecanismes, no obstant, és molt probable que la majoria d'aquestes influències estiguin sota una selecció negativa i per tant reprimides. Però tal i com hem vist, en algunes ocasions la inserció d'un element Alu i la seva evolució conjunta amb un gen específic el portarà a ser regulador d'aquest gen (Deininger, 2011; Polak and Arndt, 2008).

Com ja hem comentat els elements Alu són rics en residus CpG, àmpliament subjectes a la metilació i per tant responsables d'aproximadament el 25% de la totalitat de la metilació al genoma. Per altra banda el fet que les citosines metilades siguin més fàcilment mutables que les citosines desmetilades (Polak and Arndt, 2008) explica que els elements Alu desmetilats no pateixin tantes mutacions i que estiguin més conservats a nivell de seqüència que els elements metilats (Xing et al., 2004). A més hem vist que la metilació d'alguns elements Alu varia entre diferents teixits i disminueix en molts tumors. És probable doncs que la desmetilació d'un element Alu augmenti l'expressió del locus on es troba. D'acord amb el que hem anat observant, i el que d'altres estudis també han postulat, els elements Alu podrien ser una font de noves illes CpG, formant una barrera de metilació que influiria en la regulació dels gens propers (com ocorre amb ZALU3, proper al gen *GLDC*), o bé, segons el seu estat epigenètic, canviaria la conformació de la cromatina i desplaçaria l'inici de transcripció del gen proper permetent o no la seva expressió (com és el cas de *Aj2c1*, proper al gen *DIEXF*) (Figura D.5).

Els elements Alu també destaquen per albergar un gran nombre de llocs d'unió a factors de transcripció, i alguns d'aquests llocs d'unió són específics de certes

subfamílies Alu. Per tant, tenen una enorme capacitat per influir en l'expressió dels gens propers (Deininger, 2011; Polak and Domany, 2006).

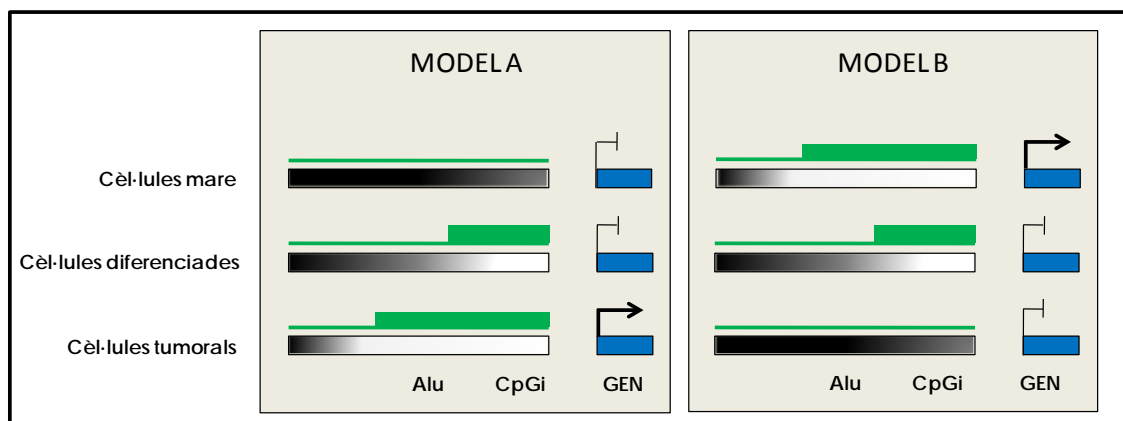


Fig. D.5. Model resum del paper dels elements Alu en la regulació transcripcional segons els nostres resultats. El verd indica els nivells de marques d'histona actives i la barra inferior mostra l'estat de metilació, essent negre metilat i blanc desmetilat. *Aj2c1* (*DIEXF*) estaria representat pel model A i *ZALU3* (*GLDC*) pel model B.

A part dels mecanismes de regulació transcripcional en els que ens hem centrat al llarg d'aquesta tesi, s'ha vist que els elements Alu també intervenen en processos post-transcripcionals. Dins d'aquests processos els elements Alu situats dins de gens participen proporcionant llocs de poliadenilació, llocs d'*splicing* alternatiu i llocs per l'edició de l'ARN (Stower, 2013).

Els elements Alu, com ja hem comentat, tenen dues cues d'adenines en la seva seqüència consens que poden mutar fàcilment i donar lloc a la cua de poliadenilació consens AATAAA. L'*splicing* alternatiu generat per elements Alu es coneix com 'exonització' dels elements Alu (Stower, 2013). Aquest fenomen està molt estès, i sens dubte afecta a centenars, si no milers, dels gens humans. En general, l'ús de seqüències Alu per generar un *splicing* alternatiu sembla ser la causa només d'una disminució d'expressió. No obstant, els *splicings* alternatius que sobreviuen evolutivament es tornen dominants afectant a la funció del gen on s'ha produït l'esdeveniment. Una àmplia gamma de malalties són causades per aquest mecanisme, com per exemple el síndrome d'Alport, el síndrome de Leigh, la degeneració coriorretinal, mucopolisacaridosi VII, l'angioedema hereditari causat per la mutació del gen *C1-INH*, l' $\alpha$ -talassèmia, causada per la mutació del gen  $\alpha$ -globina, síndrome Ehlers-Danlos causat per la mutació del gen *PLOD* i, com hem vist anteriorment, la hiperglicèmia no cetòsica causada per la mutació del gen *GLDC*. També hi ha casos de distròfia muscular de Duchenne, causats per la mutació del gen *DMD* (Stower, 2013).

Finalment també s'ha vist que, del prop d'un milió de repeticions Alu en el genoma, el 0,5% resideix en les regions UTR3' dels gens (Smalheiser and Torvik, 2006). De nou els elements Alu tindrien una implicació en la regulació gènica ja que podrien actuar com a possibles llocs d'unió de microRNAs. Els microRNAs són miRNAs altament conservats i relacionats amb la regulació de l'expressió gènica unint-se a les regions UTR3' dels gens, inhibint així la seva traducció a ARNm (Smalheiser and Torvik, 2006).

**CONCLUSIONS**



1. L'anàlisi sistemàtic de la metilació en les seqüències repetitives Alu ha permès identificar un subgrup d'elements amb perfils epigenètics distintius i propis d'estats funcionals actius.
2. Els elements Alu desmetilats en teixit normal presenten un alt grau de conservació epigenètica.
3. Els elements Alu presenten perfils epigenètics específics d'estats fisiològics (segons el tipus de cel·lular) o patològics (com el càncer).
4. S'han identificat i caracteritzat dues regions promotores corresponents als gens *DIEXF* i *GLDC* que presenten elements Alu (Aj2c1 i ZALU3, respectivament) amb un patró epigenètic dinàmic durant el càncer i el desenvolupament.
5. Els elements Alu Aj2c1 i ZALU3 presenten canvis epigenètics que delimiten la regió reguladora dels gens adjacents, traduint-se en perfils transcripcionals i proteics associats a l'estat epigenètic de l'Alu.

## **BIBLIOGRAFIA**

- Aguilera, O., Fernandez, A. F., Munoz, A., and Fraga, M. F. (2010). Epigenetics and environment: a complex relationship. *J Appl Physiol* 109, 243-251.
- Ahmed, F. E. (2004). Effect of diet, life style, and other environmental/chemopreventive factors on colorectal cancer development, and assessment of the risks. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev* 22, 91-147.
- Ahnen, D. J. (1991). Genetics of colon cancer. *West J Med* 154, 700-705.
- Allis, C. D., Berger, S. L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhhattar, R., *et al.* (2007). New nomenclature for chromatin-modifying enzymes. *Cell* 131, 633-636.
- Andrews, N. C., and Faller, D. V. (1991). A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res* 19, 2499.
- Aprile, G., Mazzer, M., Moroso, S., and Puglisi, F. (2009). Pharmacology and therapeutic efficacy of capecitabine: focus on breast and colorectal cancer. *Anticancer Drugs* 20, 217-229.
- Atkinson, S., and Armstrong, L. (2008). Epigenetics in embryonic stem cells: regulation of pluripotency and differentiation. *Cell Tissue Res* 331, 23-29.
- Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkschlager, M., and Fisher, A. G. (2006). Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* 8, 532-538.
- Bannister, A. J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res* 21, 381-395.
- Barak, M., Levanon, E. Y., Eisenberg, E., Paz, N., Rechavi, G., Church, G. M., and Mehr, R. (2009). Evidence for large diversity in the human transcriptome created by Alu RNA editing. *Nucleic Acids Res* 37, 6905-6915.
- Barrera, V., and Peinado, M. A. (2012). Evaluation of single CpG sites as proxies of CpG island methylation states at the genome scale. *Nucleic Acids Res* 40, 11490-11498.
- Barrero, M. J., Boue, S., and Izpisua Belmonte, J. C. (2010). Epigenetic mechanisms that regulate cell identity. *Cell Stem Cell* 7, 565-570.
- Bassett, A., Cooper, S., Wu, C., and Travers, A. (2009). The folding and unfolding of eukaryotic chromatin. *Curr Opin Genet Dev* 19, 159-165.
- Batzler, M. A., and Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nat Rev Genet* 3, 370-379.
- Bayani, J., and Squire, J. A. (2004). Traditional banding of chromosomes for cytogenetic analysis. *Curr Protoc Cell Biol Chapter 22*, Unit 22 23.
- Baylin, S. B., and Herman, J. G. (2000). DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet* 16, 168-174.
- Baylin, S. B., and Schuebel, K. E. (2007). Genomic biology: the epigenomic era opens. *Nature* 448, 548-549.
- Becker, P. B., and Horz, W. (2002). ATP-dependent nucleosome remodeling. *Annu Rev Biochem* 71, 247-273.
- Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R., and Deininger, P. (2010). Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res* 38, 3909-3922.
- Benayoun, L., Gingis-Velitski, S., Voloshin, T., Segal, E., Segev, R., Munster, M., Brill, R., Satchi-Fainaro, R., Scherer, S. J., and Shaked, Y. (2012). Tumor-initiating cells of various tumor types exhibit differential angiogenic properties and react differently to antiangiogenic drugs. *Stem Cells* 30, 1831-1841.

- Bennett, E. A., Keller, H., Mills, R. E., Schmidt, S., Moran, J. V., Weichenrieder, O., and Devine, S. E. (2008). Active Alu retrotransposons in the human genome. *Genome Res* 18, 1875-1883.
- Berdasco, M., and Esteller, M. (2010). Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev Cell* 19, 698-711.
- Berdasco, M., and Esteller, M. (2011). DNA methylation in stem cell renewal and multipotency. *Stem Cell Res Ther* 2, 42.
- Berger, S. L. (2007). The complex language of chromatin regulation during transcription. *Nature* 447, 407-412.
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315-326.
- Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., *et al.* (2010). The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 28, 1045-1048.
- Bertram, J. S. (2000). The molecular biology of cancer. *Mol Aspects Med* 21, 167-223.
- Bhutani, N., Burns, D. M., and Blau, H. M. (2011). DNA demethylation dynamics. *Cell* 146, 866-872.
- Bibikova, M., Chudin, E., Wu, B., Zhou, L., Garcia, E. W., Liu, Y., Shin, S., Plaia, T. W., Auerbach, J. M., Arking, D. E., *et al.* (2006). Human embryonic stem cells have a unique epigenetic signature. *Genome Res* 16, 1075-1083.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev* 16, 6-21.
- Bird, A. (2007). Perceptions of epigenetics. *Nature* 447, 396-398.
- Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. *Science* 330, 612-616.
- Borras, J. M., Sanchez-Hernandez, A., Navarro, M., Martinez, M., Mendez, E., Ponton, J. L., Espinas, J. A., and Germa, J. R. (2001). Compliance, satisfaction, and quality of life of patients with colorectal cancer receiving home chemotherapy or outpatient treatment: a randomised controlled trial. *Bmj* 322, 826.
- Brosius, J. (1999). RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* 238, 115-134.
- Bushey, A. M., Dorman, E. R., and Corces, V. G. (2008). Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol Cell* 32, 1-9.
- Butcher, L. M., and Beck, S. (2008). Future impact of integrated high-throughput methylome analyses on human health and disease. *J Genet Genomics* 35, 391-401.
- Carter, A. B., Salem, A. H., Hedges, D. J., Keegan, C. N., Kimball, B., Walker, J. A., Watkins, W. S., Jorde, L. B., and Batzer, M. A. (2004). Genome-wide analysis of the human Alu Yb-lineage. *Hum Genomics* 1, 167-178.
- Clark, S. J. (2007). Action at a distance: epigenetic silencing of large chromosomal regions in carcinogenesis. *Hum Mol Genet* 16 Spec No 1, R88-95.
- Clark, S. J., Statham, A., Stirzaker, C., Molloy, P. L., and Frommer, M. (2006). DNA methylation: bisulphite modification and analysis. *Nat Protoc* 1, 2353-2364.
- Compton, C. C., and Greene, F. L. (2004). The staging of colorectal cancer: 2004 and beyond. *CA Cancer J Clin* 54, 295-308.
- Cook, G. W., Konkel, M. K., Major, J. D., 3rd, Walker, J. A., Han, K., and Batzer, M. A. (2011). Alu pair exclusions in the human genome. *Mob DNA* 2, 10.

- Cordaux, R., and Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nat Rev Genet* 10, 691-703.
- Costello, J. F., Fruhwald, M. C., Smiraglia, D. J., Rush, L. J., Robertson, G. P., Gao, X., Wright, F. A., Feramisco, J. D., Peltomaki, P., Lang, J. C., et al. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* 24, 132-138.
- Cui, H., Onyango, P., Brandenburg, S., Wu, Y., Hsieh, C. L., and Feinberg, A. P. (2002). Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2. *Cancer Res* 62, 6442-6446.
- Chen, C., Gentles, A. J., Jurka, J., and Karlin, S. (2002). Genes, pseudogenes, and Alu sequence organization across human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* 99, 2930-2935.
- Chen, J., Ng, S. M., Chang, C., Zhang, Z., Bourdon, J. C., Lane, D. P., and Peng, J. (2009). p53 isoform delta113p53 is a p53 target gene that antagonizes p53 apoptotic activity via BclxL activation in zebrafish. *Genes Dev* 23, 278-290.
- Chen, J., Ruan, H., Ng, S. M., Gao, C., Soo, H. M., Wu, W., Zhang, Z., Wen, Z., Lane, D. P., and Peng, J. (2005). Loss of function of def selectively up-regulates Delta113p53 expression to arrest expansion growth of digestive organs in zebrafish. *Genes Dev* 19, 2900-2911.
- Chen, L. L., DeCerbo, J. N., and Carmichael, G. G. (2008). Alu element-mediated gene silencing. *Embo J* 27, 1694-1705.
- Cho, N. Y., Kim, B. H., Choi, M., Yoo, E. J., Moon, K. C., Cho, Y. M., Kim, D., and Kang, G. H. (2007). Hypermethylation of CpG island loci and hypomethylation of LINE-1 and Alu repeats in prostate adenocarcinoma and their relationship to clinicopathological features. *J Pathol* 211, 269-277.
- Choi, I. S., Estecio, M. R., Nagano, Y., Kim do, H., White, J. A., Yao, J. C., Issa, J. P., and Rashid, A. (2007). Hypomethylation of LINE-1 and Alu in well-differentiated neuroendocrine tumors (pancreatic endocrine tumors and carcinoid tumors). *Mod Pathol* 20, 802-810.
- Danaei, G., Vander Hoorn, S., Lopez, A. D., Murray, C. J., and Ezzati, M. (2005). Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet* 366, 1784-1793.
- Daskalos, A., Nikolaidis, G., Xinarianos, G., Savvari, P., Cassidy, A., Zakopoulou, R., Kotsinas, A., Gorgoulis, V., Field, J. K., and Liloglou, T. (2009). Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer* 124, 81-87.
- Debrix, I., Avenin, D., Andre, T., Abbas, F., Gligorov, J., Milleron, B., Lotz, J. P., and Becker, A. (2008). [Anticancer chemotherapy in the elderly: a review of the literature]. *Bull Cancer* 95 *FMC Onco*, F37-43.
- Deininger, P. (2011). Alu elements: know the SINEs. *Genome Biol* 12, 236.
- Deininger, P. L., and Batzer, M. A. (2002). Mammalian retroelements. *Genome Res* 12, 1455-1465.
- Denda, A., Rao, P. M., Rajalakshmi, S., and Sarma, D. S. (1985). 5-azacytidine potentiates initiation induced by carcinogens in rat liver. *Carcinogenesis* 6, 145-146.
- Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35, 41-48.
- Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., et al. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 41, 1350-1353.
- Ducasse, M., and Brown, M. A. (2006). Epigenetic aberrations and cancer. *Mol Cancer* 5, 60.
- Duggan, B. D., Felix, J. C., Muderspach, L. I., Tourgeman, D., Zheng, J., and Shibata, D. (1994). Microsatellite instability in sporadic endometrial carcinoma. *J Natl Cancer Inst* 86, 1216-1221.

- Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* 300, 455.
- Edwards, J. R., O'Donnell, A. H., Rollins, R. A., Peckham, H. E., Lee, C., Milekic, M. H., Chanrion, B., Fu, Y., Su, T., Hibshoosh, H., *et al.* (2010). Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Res* 20, 972-980.
- Ehrlich, M. (2002). DNA methylation in cancer: too much, but also too little. *Oncogene* 21, 5400-5413.
- Elkon, R., Ugalde, A. P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* 14, 496-506.
- Estecio, M. R., Gallegos, J., Vallot, C., Castoro, R. J., Chung, W., Maegawa, S., Oki, Y., Kondo, Y., Jelinek, J., Shen, L., *et al.* (2010). Genome architecture marked by retrotransposons modulates predisposition to DNA methylation in cancer. *Genome Res* 20, 1369-1382.
- Esteller, M. (2002). CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* 21, 5427-5440.
- Esteller, M. (2005). Aberrant DNA methylation as a cancer-inducing mechanism. *Annu Rev Pharmacol Toxicol* 45, 629-656.
- Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8, 286-298.
- Esteller, M. (2008). Epigenetics in cancer. *N Engl J Med* 358, 1148-1159.
- Fearon, E. R., and Jones, P. A. (1992). Progressing toward a molecular description of colorectal cancer development. *Faseb J* 6, 2783-2790.
- Feinberg, A. P. (1988). Alterations in DNA methylation in colorectal polyps and cancer. *Prog Clin Biol Res* 279, 309-317.
- Feinberg, A. P. (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature* 447, 433-440.
- Feinberg, A. P. (2010a). Epigenomics reveals a functional genome anatomy and a new approach to common disease. *Nat Biotechnol* 28, 1049-1052.
- Feinberg, A. P. (2010b). Genome-scale approaches to the epigenetics of common human disease. *Virchows Arch* 456, 13-21.
- Feinberg, A. P., and Tycko, B. (2004). The history of cancer epigenetics. *Nat Rev Cancer* 4, 143-153.
- Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. *Nature* 421, 448-453.
- Feltus, F. A., Lee, E. K., Costello, J. F., Plass, C., and Vertino, P. M. (2006). DNA motifs associated with aberrant CpG island methylation. *Genomics* 87, 572-579.
- Filippova, G. N. (2008). Genetics and epigenetics of the multifunctional protein CTCF. *Curr Top Dev Biol* 80, 337-360.
- Filkowski, J. N., Ilnytsky, Y., Tamminga, J., Koturbash, I., Golubov, A., Bagnyukova, T., Pogribny, I. P., and Kovalchuk, O. (2010). Hypomethylation and genome instability in the germline of exposed parents and their progeny is associated with altered miRNA expression. *Carcinogenesis* 31, 1110-1115.
- Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., Heine-Suner, D., Cigudosa, J. C., Urioste, M., Benitez, J., *et al.* (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* 102, 10604-10609.
- Fraser, P., and Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. *Nature* 447, 413-417.

- Friedman, S. (2006). Cancer in Crohn's disease. *Gastroenterol Clin North Am* 35, 621-639.
- Frigola, J., Song, J., Storzaker, C., Hinshelwood, R. A., Peinado, M. A., and Clark, S. J. (2006). Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nat Genet* 38, 540-549.
- Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. *J Mol Biol* 196, 261-282.
- Gasparini, P., Sozzi, G., and Pierotti, M. A. (2007). The role of chromosomal alterations in human cancer development. *J Cell Biochem* 102, 320-331.
- Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* 7, 703-713.
- Gaudet, F., Hodgson, J. G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J. W., Leonhardt, H., and Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. *Science* 300, 489-492.
- Gibbs, R. A., Rogers, J., Katze, M. G., Bumgarner, R., Weinstock, G. M., Mardis, E. R., Remington, K. A., Strausberg, R. L., Venter, J. C., Wilson, R. K., *et al.* (2007). Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316, 222-234.
- Gilbert, N., Thomson, I., Boyle, S., Allan, J., Ramsahoye, B., and Bickmore, W. A. (2007). DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *J Cell Biol* 177, 401-411.
- Gius, D., Cui, H., Bradbury, C. M., Cook, J., Smart, D. K., Zhao, S., Young, L., Brandenburg, S. A., Hu, Y., Bisht, K. S., *et al.* (2004). Distinct effects on gene expression of chemical and genetic manipulation of the cancer epigenome revealed by a multimodality approach. *Cancer Cell* 6, 361-371.
- Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74, 481-514.
- Goodier, J. L., and Kazazian, H. H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* 135, 23-35.
- Grady, W. M. (2005). Epigenetic events in the colorectum and in colon cancer. *Biochem Soc Trans* 33, 684-688.
- Grewal, S. I., and Jia, S. (2007). Heterochromatin revisited. *Nat Rev Genet* 8, 35-46.
- Grover, D., Mukerji, M., Bhatnagar, P., Kannan, K., and Brahmachari, S. K. (2004). Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition. *Bioinformatics* 20, 813-817.
- Gu, T. J., Yi, X., Zhao, X. W., Zhao, Y., and Yin, J. Q. (2009). Alu-directed transcriptional regulation of some novel miRNAs. *BMC Genomics* 10, 563.
- Hagemann, S., Heil, O., Lyko, F., and Brueckner, B. (2011). Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines. *PLoS One* 6, e17388.
- Hansen, K. D., Timp, W., Bravo, H. C., Sabunciyani, S., Langmead, B., McDonald, O. G., Wen, B., Wu, H., Liu, Y., Diep, D., *et al.* (2011). Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43, 768-775.
- Hasler, J., Samuelsson, T., and Strub, K. (2007). Useful 'junk': Alu RNAs in the human transcriptome. *Cell Mol Life Sci* 64, 1793-1800.
- Hasler, J., and Strub, K. (2006). Alu elements as regulators of gene expression. *Nucleic Acids Res* 34, 5491-5497.

- Hattori, N., Nishino, K., Ko, Y. G., Hattori, N., Ohgane, J., Tanaka, S., and Shiota, K. (2004). Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. *J Biol Chem* 279, 17063-17069.
- Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat Rev Genet* 9, 15-26.
- Herman, J. G., and Baylin, S. B. (2003). Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* 349, 2042-2054.
- Herold, M., Bartkuhn, M., and Renkawitz, R. (2012). CTCF: insights into insulator function during development. *Development* 139, 1045-1057.
- Hewitt, K. J., Shamis, Y., Hayman, R. B., Margvelashvili, M., Dong, S., Carlson, M. W., and Garlick, J. A. (2011). Epigenetic and phenotypic profile of fibroblasts derived from induced pluripotent stem cells. *PLoS One* 6, e17128.
- Hodgetts, R. (2004). Eukaryotic gene regulation by targeted chromatin re-modeling at dispersed, middle-repetitive sequence elements. *Curr Opin Genet Dev* 14, 680-685.
- Hollister, J. D., and Gaut, B. S. (2009). Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* 19, 1419-1428.
- Huang, C. R., Schneider, A. M., Lu, Y., Niranjan, T., Shen, P., Robinson, M. A., Steranka, J. P., Valle, D., Civin, C. I., Wang, T., et al. (2010). Mobile interspersed repeats are major structural variants in the human genome. *Cell* 141, 1171-1182.
- Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 363, 558-561.
- Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41, 178-186.
- Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., Van Meir, E. G., Vertino, P. M., and Devine, S. E. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253-1261.
- Issa, J. P. (2000). The epigenetics of colorectal cancer. *Ann N Y Acad Sci* 910, 140-153; discussion 153-145.
- Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33 Suppl, 245-254.
- Jain, M., Nilsson, R., Sharma, S., Madhusudhan, N., Kitami, T., Souza, A. L., Kafri, R., Kirschner, M. W., Clish, C. B., and Mootha, V. K. (2012). Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science* 336, 1040-1044.
- Jasinska, A., and Krzyzosiak, W. J. (2004). Repetitive sequences that shape the human transcriptome. *FEBS Lett* 567, 136-141.
- Ji, X., Tang, J., Halberg, R., Busam, D., Ferriera, S., Pena, M. M., Venkataramu, C., Yeatman, T. J., and Zhao, S. (2010). Distinguishing between cancer driver and passenger gene alteration candidates via cross-species comparison: a pilot study. *BMC Cancer* 10, 426.
- Jirtle, R. L., and Skinner, M. K. (2007). Environmental epigenomics and disease susceptibility. *Nat Rev Genet* 8, 253-262.
- Jones, P. A., and Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 3, 415-428.
- Jones, P. A., and Baylin, S. B. (2007). The epigenomics of cancer. *Cell* 128, 683-692.



- Jorda, M., Rodriguez, J., Frigola, J., and Peinado, M. A. (2009). Analysis of DNA methylation by amplification of intermethylated sites (AIMS). *Methods Mol Biol* 507, 107-116.
- Jurka, J., Kohany, O., Pavlicek, A., Kapitonov, V. V., and Jurka, M. V. (2004). Duplication, coclustering, and selection of human Alu retrotransposons. *Proc Natl Acad Sci U S A* 101, 1268-1272.
- Kanno, J., Hutchin, T., Kamada, F., Narisawa, A., Aoki, Y., Matsubara, Y., and Kure, S. (2007). Genomic deletion within GLDC is a major cause of non-ketotic hyperglycinaemia. *J Med Genet* 44, e69.
- Karpf, A. R., and Matsui, S. (2005). Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. *Cancer Res* 65, 8635-8639.
- Kawaji, H., Frith, M. C., Katayama, S., Sandelin, A., Kai, C., Kawai, J., Carninci, P., and Hayashizaki, Y. (2006). Dynamic usage of transcription start sites within core promoters. *Genome Biol* 7, R118.
- Kazazian, H. H., Jr. (2004). Mobile elements: drivers of genome evolution. *Science* 303, 1626-1632.
- Keenen, B., and de la Serna, I. L. (2009). Chromatin remodeling in embryonic stem cells: regulating the balance between pluripotency and differentiation. *J Cell Physiol* 219, 1-7.
- Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* 11, 345-355.
- Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. *Cell* 116, 259-272.
- Kidd, J. M., Graves, T., Newman, T. L., Fulton, R., Hayden, H. S., Malig, M., Kallicki, J., Kaul, R., Wilson, R. K., and Eichler, E. E. (2010). A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell* 143, 837-847.
- Kikuchi, G., Motokawa, Y., Yoshida, T., and Hiraga, K. (2008). Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. *Proc Jpn Acad Ser B Phys Biol Sci* 84, 246-263.
- Kinzler, K. W., and Vogelstein, B. (1998). Landscaping the cancer terrain. *Science* 280, 1036-1037.
- Knudson, A. G., Jr. (1985). Hereditary cancer, oncogenes, and antioncogenes. *Cancer Res* 45, 1437-1443.
- Kochanek, S., Renz, D., and Doerfler, W. (1993). DNA methylation in the Alu sequences of diploid and haploid primary human cells. *Embo J* 12, 1141-1151.
- Konkel, M. K., and Batzer, M. A. (2010). A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Semin Cancer Biol* 20, 211-221.
- Kornberg, R. D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98, 285-294.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.
- Kreahling, J., and Graveley, B. R. (2004). The origins and implications of Aluternative splicing. *Trends Genet* 20, 1-4.
- Kriegs, J. O., Churakov, G., Jurka, J., Brosius, J., and Schmitz, J. (2007). Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet* 23, 158-161.
- Krivega, I., and Dean, A. (2012). Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev* 22, 79-85.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

- Laird, P. W. (2010). Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 11, 191-203.
- Lamond, A. I., and Earnshaw, W. C. (1998). Structure and function in the nucleus. *Science* 280, 547-553.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.
- Laurent, L., Wong, E., Li, G., Huynh, T., Tsigos, A., Ong, C. T., Low, H. M., Kin Sung, K. W., Rigoutsos, I., Loring, J., and Wei, C. L. (2010). Dynamic changes in the human methylome during differentiation. *Genome Res* 20, 320-331.
- Lee, J. T., and Bartolomei, M. S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* 152, 1308-1323.
- Leedham, S. J., and Wright, N. A. (2008). Expansion of a mutated clone: from stem cell to tumour. *J Clin Pathol* 61, 164-171.
- Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. *Nature* 396, 643-649.
- Lev-Maor, G., Goren, A., Sela, N., Kim, E., Keren, H., Doron-Faigenboim, A., Leibman-Barak, S., Pupko, T., and Ast, G. (2007). The "alternative" choice of constitutive exons throughout evolution. *PLoS Genet* 3, e203.
- Lev-Maor, G., Ram, O., Kim, E., Sela, N., Goren, A., Levanon, E. Y., and Ast, G. (2008). Intronic Alus influence alternative splicing. *PLoS Genet* 4, e1000204.
- Li, Y., Zhu, J., Tian, G., Li, N., Li, Q., Ye, M., Zheng, H., Yu, J., Wu, H., Sun, J., *et al.* (2010). The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol* 8, e1000533.
- Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schubeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* 43, 1091-1097.
- Lin, Y., Li, Z., Ozsolak, F., Kim, S. W., Arango-Argoty, G., Liu, T. T., Tenenbaum, S. A., Bailey, T., Monaghan, A. P., Milos, P. M., and John, B. (2012). An in-depth map of polyadenylation sites in cancer. *Nucleic Acids Res* 40, 8460-8471.
- Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., *et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315-322.
- Liu, G. E., Alkan, C., Jiang, L., Zhao, S., and Eichler, E. E. (2009). Comparative analysis of Alu repeats in primate genomes. *Genome Res* 19, 876-885.
- Liu, K., Wang, Y. F., Cantemir, C., and Muller, M. T. (2003). Endogenous assays of DNA methyltransferases: Evidence for differential activities of DNMT1, DNMT2, and DNMT3 in mammalian cells in vivo. *Mol Cell Biol* 23, 2709-2719.
- Liu, W. M., and Schmid, C. W. (1993). Proposed roles for DNA methylation in Alu transcriptional repression and mutational inactivation. *Nucleic Acids Res* 21, 1351-1359.
- Loeb, K. R., and Loeb, L. A. (1999). Genetic instability and the mutator phenotype. Studies in ulcerative colitis. *Am J Pathol* 154, 1621-1626.
- Lorch, Y., Zhang, M., and Kornberg, R. D. (1999). Histone octamer transfer by a chromatin-remodeling complex. *Cell* 96, 389-392.
- Lupski, J. R. (2010). Retrotransposition and structural variation in the human genome. *Cell* 141, 1110-1112.

- Macia, A., Munoz-Lopez, M., Cortes, J. L., Hastings, R. K., Morell, S., Lucena-Aguilar, G., Marchal, J. A., Badge, R. M., and Garcia-Perez, J. L. (2011). Epigenetic control of retrotransposon expression in human embryonic stem cells. *Mol Cell Biol* 31, 300-316.
- Maegawa, S., Hinkal, G., Kim, H. S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L. A., and Issa, J. P. (2010). Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res* 20, 332-340.
- Martin, D. I., Singer, M., Dhahbi, J., Mao, G., Zhang, L., Schroth, G. P., Pachter, L., and Boffelli, D. (2011). Phyloepigenomic comparison of great apes reveals a correlation between somatic and germline methylation states. *Genome Res* 21, 2049-2057.
- Martin, S. T., Sato, N., Dhara, S., Chang, R., Hustinx, S. R., Abe, T., Maitra, A., and Goggins, M. (2005). Aberrant methylation of the Human Hedgehog interacting protein (HHIP) gene in pancreatic neoplasms. *Cancer Biol Ther* 4, 728-733.
- Mayor, R., Casadome, L., Azuara, D., Moreno, V., Clark, S. J., Capella, G., and Peinado, M. A. (2009). Long-range epigenetic silencing at 2q14.2 affects most human colorectal cancers and may have application as a non-invasive biomarker of disease. *Br J Cancer* 100, 1534-1539.
- Mayr, C., and Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138, 673-684.
- Medstrand, P., van de Lagemaat, L. N., and Mager, D. L. (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res* 12, 1483-1495.
- Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766-770.
- Melki, J. R., Vincent, P. C., and Clark, S. J. (1999). Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia. *Cancer Res* 59, 3730-3740.
- Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* 7, 540-546.
- Mighell, A. J., Markham, A. F., and Robinson, P. A. (1997). Alu sequences. *FEBS Lett* 417, 1-5.
- Mills, R. E., Bennett, E. A., Iskow, R. C., and Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends Genet* 23, 183-191.
- Mohn, F., and Schubeler, D. (2009). Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends Genet* 25, 129-136.
- Molto, E., Fernandez, A., and Montoliu, L. (2009). Boundaries in vertebrate genomes: different solutions to adequately insulate gene expression domains. *Brief Funct Genomic Proteomic* 8, 283-296.
- Mollet, I. G., Ben-Dov, C., Felicio-Silva, D., Grosso, A. R., Eleuterio, P., Alves, R., Staller, R., Silva, T. S., and Carmo-Fonseca, M. (2010). Unconstrained mining of transcript data reveals increased alternative splicing complexity in the human transcriptome. *Nucleic Acids Res* 38, 4740-4754.
- Muotri, A. R., Marchetto, M. C., Coufal, N. G., and Gage, F. H. (2007). The necessary junk: new functions for transposable elements. *Hum Mol Genet* 16 Spec No. 2, R159-167.
- Muramoto, H., Yagi, S., Hirabayashi, K., Sato, S., Ohgane, J., Tanaka, S., and Shiota, K. (2010). Enrichment of short interspersed transposable elements to embryonic stem cell-specific hypomethylated gene regions. *Genes Cells* 15, 855-865.
- Murray-Zmijewski, F., Lane, D. P., and Bourdon, J. C. (2006). p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. *Cell Death Differ* 13, 962-972.

- Nardella, C., Clohessy, J. G., Alimonti, A., and Pandolfi, P. P. (2011). Pro-senescence therapy for cancer treatment. *Nat Rev Cancer* 11, 503-511.
- Nowak, M. A., Komarova, N. L., Sengupta, A., Jallepalli, P. V., Shih Ie, M., Vogelstein, B., and Lengauer, C. (2002). The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci U S A* 99, 16226-16231.
- Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science* 194, 23-28.
- Oei, S. L., Babich, V. S., Kazakov, V. I., Usmanova, N. M., Kropotov, A. V., and Tomilin, N. V. (2004). Clusters of regulatory signals for RNA polymerase II transcription associated with Alu family repeats and CpG islands in human promoters. *Genomics* 83, 873-882.
- Ohshima, K., Hattori, M., Yada, T., Gojobori, T., Sakaki, Y., and Okada, N. (2003). Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol* 4, R74.
- Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247-257.
- Orgel, L. E., Crick, F. H., and Sapienza, C. (1980). Selfish DNA. *Nature* 288, 645-646.
- Ostertag, E. M., and Kazazian, H. H., Jr. (2001). Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 35, 501-538.
- Pace, J. K., 2nd, and Feschotte, C. (2007). The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res* 17, 422-432.
- Pal, A., Srivastava, T., Sharma, M. K., Mehndiratta, M., Das, P., Sinha, S., and Chattopadhyay, P. (2010). Aberrant methylation and associated transcriptional mobilization of Alu elements contributes to genomic instability in hypoxia. *J Cell Mol Med* 14, 2646-2654.
- Pang, A. W., MacDonald, J. R., Pinto, D., Wei, J., Rafiq, M. A., Conrad, D. F., Park, H., Hurles, M. E., Lee, C., Venter, J. C., *et al.* (2010). Towards a comprehensive structural variation map of an individual human genome. *Genome Biol* 11, R52.
- Park, Y., and Kuroda, M. I. (2001). Epigenetic aspects of X-chromosome dosage compensation. *Science* 293, 1083-1085.
- Paz, M. F., Wei, S., Cigudosa, J. C., Rodriguez-Perales, S., Peinado, M. A., Huang, T. H., and Esteller, M. (2003). Genetic unmasking of epigenetically silenced tumor suppressor genes in colon cancer cells deficient in DNA methyltransferases. *Hum Mol Genet* 12, 2209-2219.
- Peterson, C. L., and Laniel, M. A. (2004). Histones and histone modifications. *Curr Biol* 14, R546-551.
- Petrat, F., Boengler, K., Schulz, R., and de Groot, H. (2012). Glycine, a simple physiological compound protecting by yet puzzling mechanism(s) against ischaemia-reperfusion injury: current knowledge. *Br J Pharmacol* 165, 2059-2072.
- Polak, P., and Arndt, P. F. (2008). Transcription induces strand-specific mutations at the 5' end of human genes. *Genome Res* 18, 1216-1223.
- Polak, P., and Domany, E. (2006). Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* 7, 133.
- Ponicsan, S. L., Kugel, J. F., and Goodrich, J. A. (2010). Genomic gems: SINE RNAs regulate mRNA production. *Curr Opin Genet Dev* 20, 149-155.
- Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat Biotechnol* 28, 1057-1068.
- Poulsen, P., Esteller, M., Vaag, A., and Fraga, M. F. (2007). The epigenetic basis of twin discordance in age-related diseases. *Pediatr Res* 61, 38R-42R.

- Price, A. L., Eskin, E., and Pevzner, P. A. (2004). Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. *Genome Res* 14, 2245-2252.
- Probst, A. V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Biol* 10, 192-206.
- Ptashne, M. (2007). On the use of the word 'epigenetic'. *Curr Biol* 17, R233-236.
- Pufulete, M., Al-Ghnam, R., Rennie, J. A., Appleby, P., Harris, N., Gout, S., Emery, P. W., and Sanders, T. A. (2005). Influence of folate status on genomic DNA methylation in colonic mucosa of subjects without colorectal adenoma or cancer. *Br J Cancer* 92, 838-842.
- Quigley, D., and Balmain, A. (2009). Systems genetics analysis of cancer susceptibility: from mouse models to humans. *Nat Rev Genet* 10, 651-657.
- Raptis, S., and Bapat, B. (2006). Genetic instability in human tumors. *Exs*, 303-320.
- Rhee, I., Bachman, K. E., Park, B. H., Jair, K. W., Yen, R. W., Schuebel, K. E., Cui, H., Feinberg, A. P., Lengauer, C., Kinzler, K. W., et al. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* 416, 552-556.
- Richard, G. F., Kerrest, A., and Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev* 72, 686-727.
- Risques, R. A., Moreno, V., Ribas, M., Marcuello, E., Capella, G., and Peinado, M. A. (2003). Genetic pathways and genome-wide determinants of clinical outcome in colorectal cancer. *Cancer Res* 63, 7206-7214.
- Robertson, K. D., Ait-Si-Ali, S., Yokochi, T., Wade, P. A., Jones, P. L., and Wolffe, A. P. (2000). DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat Genet* 25, 338-342.
- Robertson, K. D., and Wolffe, A. P. (2000). DNA methylation in health and disease. *Nat Rev Genet* 1, 11-19.
- Rodenhiser, D., and Mann, M. (2006). Epigenetics and human disease: translating basic biology into clinical applications. *Cmaj* 174, 341-348.
- Rodriguez, J., Frigola, J., Vendrell, E., Risques, R. A., Fraga, M. F., Morales, C., Moreno, V., Esteller, M., Capella, G., Ribas, M., and Peinado, M. A. (2006). Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. *Cancer Res* 66, 8462-9468.
- Rodriguez, J., Munoz, M., Vives, L., Frangou, C. G., Groudine, M., and Peinado, M. A. (2008a). Bivalent domains enforce transcriptional memory of DNA methylated genes in cancer cells. *Proc Natl Acad Sci U S A* 105, 19809-19814.
- Rodriguez, J., Vives, L., Jorda, M., Morales, C., Munoz, M., Vendrell, E., and Peinado, M. A. (2008b). Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells. *Nucleic Acids Res* 36, 770-784.
- Rollins, R. A., Haghghi, F., Edwards, J. R., Das, R., Zhang, M. Q., Ju, J., and Bestor, T. H. (2006). Large-scale structure of genomic methylation patterns. *Genome Res* 16, 157-163.
- Roy-Engel, A. M., Carroll, M. L., El-Sawy, M., Salem, A. H., Garber, R. K., Nguyen, S. V., Deininger, P. L., and Batzer, M. A. (2002a). Non-traditional Alu evolution and primate genomic diversity. *J Mol Biol* 316, 1033-1040.
- Roy-Engel, A. M., Salem, A. H., Oyeniran, O. O., Deininger, L., Hedges, D. J., Kilroy, G. E., Batzer, M. A., and Deininger, P. L. (2002b). Active Alu element "A-tails": size does matter. *Genome Res* 12, 1333-1344.

- Saito, Y., Suzuki, H., Tsugawa, H., Nakagawa, I., Matsuzaki, J., Kanai, Y., and Hibi, T. (2009). Chromatin remodeling at Alu repeats by epigenetic treatment activates silenced microRNA-512-5p with downregulation of Mcl-1 in human gastric cancer cells. *Oncogene* 28, 2738-2744.
- Satterlee, J. S., Schubeler, D., and Ng, H. H. (2010). Tackling the epigenome: challenges and opportunities for collaboration. *Nat Biotechnol* 28, 1039-1044.
- Schmid, C. W. (1991). Human Alu subfamilies and their methylation revealed by blot hybridization. *Nucleic Acids Res* 19, 5613-5617.
- Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol* 6, 73-77.
- Schulz, W. A., Steinhoff, C., and Florl, A. R. (2006). Methylation of endogenous human retroelements in health and disease. *Curr Top Microbiol Immunol* 310, 211-250.
- Sela, N., Mersch, B., Gal-Mark, N., Lev-Maor, G., Hotz-Wagenblatt, A., and Ast, G. (2007). Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol* 8, R127.
- Sempere, L., and Jover, R. (2008). [New perspectives in predicting response to chemotherapy in colorectal cancer]. *Gastroenterol Hepatol* 31, 580-586.
- Sen, S. K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P. A., Dyer, M., Cordaux, R., Liang, P., and Batzer, M. A. (2006). Human genomic deletions mediated by recombination between Alu elements. *Am J Hum Genet* 79, 41-53.
- Sharma, P. C., Grover, A., and Kahl, G. (2007). Mining microsatellites in eukaryotic genomes. *Trends Biotechnol* 25, 490-498.
- Sharma, S., Kelly, T. K., and Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis* 31, 27-36.
- Shen, Y., Chow, J., Wang, Z., and Fan, G. (2006). Abnormal CpG island methylation occurs during in vitro differentiation of human embryonic stem cells. *Hum Mol Genet* 15, 2623-2635.
- Shi, Y. (2007). Histone lysine demethylases: emerging roles in development, physiology and disease. *Nat Rev Genet* 8, 829-833.
- Shibata, D., Peinado, M. A., Ionov, Y., Malkhosyan, S., and Perucho, M. (1994). Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nat Genet* 6, 273-281.
- Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8, 272-285.
- Smalheiser, N. R., and Torvik, V. I. (2006). Alu elements within human mRNAs are probable microRNA targets. *Trends Genet* 22, 532-536.
- Smallwood, A., and Ren, B. (2013). Genome organization and long-range regulation of gene expression by enhancers. *Curr Opin Cell Biol* 25, 387-394.
- Smith, Z. D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat Rev Genet* 14, 204-220.
- Stankiewicz, P., and Lupski, J. R. (2010). Structural variation in the human genome and its role in disease. *Annu Rev Med* 61, 437-455.
- Stenger, J. E., Lobachev, K. S., Gordenin, D., Darden, T. A., Jurka, J., and Resnick, M. A. (2001). Biased distribution of inverted and direct Alus in the human genome: implications for insertion, exclusion, and genome stability. *Genome Res* 11, 12-27.
- Stephens, E. B., Sahni, M., Leung, K., Raghavan, R., Joag, S. V., and Narayan, O. (1998). Nucleotide substitutions in the long terminal repeat are not required for development of neurovirulence by simian immunodeficiency virus strain mac. *J Gen Virol* 79 ( Pt 5), 1089-1100.

- Stirzaker, C., Song, J. Z., Davidson, B., and Clark, S. J. (2004). Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res* 64, 3871-3877.
- Stower, H. (2013). Alternative splicing: Regulating Alu element 'exonization'. *Nat Rev Genet* 14, 152-153.
- Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. *Nature* 403, 41-45.
- Surani, M. A. (2001). Reprogramming of genome function through epigenetic inheritance. *Nature* 414, 122-128.
- Suva, M. L., Riggi, N., and Bernstein, B. E. (2013). Epigenetic reprogramming in cancer. *Science* 339, 1567-1570.
- Suzuki, K., Suzuki, I., Leodolter, A., Alonso, S., Horiuchi, S., Yamashita, K., and Perucho, M. (2006). Global DNA demethylation in gastrointestinal cancer is age dependent and precedes genomic damage. *Cancer Cell* 9, 199-207.
- Szpakowski, S., Sun, X., Lage, J. M., Dyer, A., Rubinstein, J., Kowalski, D., Sasaki, C., Costa, J., and Lizardi, P. M. (2009). Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements. *Gene* 448, 151-167.
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930-935.
- Takai, D., and Jones, P. A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* 99, 3740-3745.
- Thomas, G. A., and Williams, E. D. (1992). Production of thyroid tumours in mice by demethylating agents. *Carcinogenesis* 13, 1039-1042.
- Tibbetts, A. S., and Appling, D. R. (2010). Compartmentalization of Mammalian folate-mediated one-carbon metabolism. *Annu Rev Nutr* 30, 57-81.
- Tinelli, A., Mezzolla, V., Leo, G., Pisano, M., Storelli, F., Alemanno, G., Malvasi, A., Tommasi, S., Ronzino, G., and Lorusso, V. (2010). Microsatellite instability (MSI) as genomic markers in endometrial cancer: toward scientific evidences. *Mini Rev Med Chem* 10, 1356-1365.
- Tomilin, N. V. (2008). Regulation of mammalian gene expression by retroelements and non-coding tandem repeats. *Bioessays* 30, 338-348.
- Tomlinson, I., and Bodmer, W. (1999). Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nat Med* 5, 11-12.
- Towbin, H., Ozbey, O., and Zingel, O. (2001). An immunoblotting method for high-resolution isoelectric focusing of protein isoforms on immobilized pH gradients. *Electrophoresis* 22, 1887-1893.
- Toyota, M., and Suzuki, H. (2010). Epigenetic drivers of genetic alterations. *Adv Genet* 70, 309-323.
- Usdin, K. (2008). The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome Res* 18, 1011-1019.
- Valinluck, V., and Sowers, L. C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res* 67, 946-950.
- van Steensel, B., and Dekker, J. (2010). Genomics tools for unraveling chromosome architecture. *Nat Biotechnol* 28, 1089-1095.
- Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M., and Bos, J. L. (1988). Genetic alterations during colorectal-tumor development. *N Engl J Med* 319, 525-532.

- Voigt, P., Tee, W. W., and Reinberg, D. (2013). A double take on bivalent promoters. *Genes Dev* 27, 1318-1338.
- Walters, R. D., Kugel, J. F., and Goodrich, J. A. (2009). Invaluable junk: the cellular impact and function of Alu and B2 RNAs. *IUBMB Life* 61, 831-837.
- Wallace, J. A., and Felsenfeld, G. (2007). We gather together: insulators and genome organization. *Curr Opin Genet Dev* 17, 400-407.
- Wang, W., Wu, Z., Dai, Z., Yang, Y., Wang, J., and Wu, G. (2013). Glycine metabolism in animals and humans: implications for nutrition and health. *Amino Acids*.
- Weber, M., and Schubeler, D. (2007). Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr Opin Cell Biol* 19, 273-280.
- Weisenberger, D. J., Campan, M., Long, T. I., Kim, M., Woods, C., Fiala, E., Ehrlich, M., and Laird, P. W. (2005). Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res* 33, 6823-6836.
- Wilhelm, C. S., Kelsey, K. T., Butler, R., Plaza, S., Gagne, L., Zens, M. S., Andrew, A. S., Morris, S., Nelson, H. H., Schned, A. R., et al. (2010). Implications of LINE1 methylation for bladder cancer risk in women. *Clin Cancer Res* 16, 1682-1689.
- Wilusz, J. E., Sunwoo, H., and Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 23, 1494-1504.
- Winawer, S. J. (1999). Natural history of colorectal cancer. *Am J Med* 106, 3S-6S; discussion 50S-51S.
- Witherspoon, D. J., Watkins, W. S., Zhang, Y., Xing, J., Tolpinrud, W. L., Hedges, D. J., Batzer, M. A., and Jorde, L. B. (2009). Alu repeats increase local recombination rates. *BMC Genomics* 10, 530.
- Wong, S. M., and Akerley, B. J. (2005). Environmental and genetic regulation of the phosphorylcholine epitope of Haemophilus influenzae lipooligosaccharide. *Mol Microbiol* 55, 724-738.
- Wongsurawat, T., Jenjaroenpun, P., Kwoh, C. K., and Kuznetsov, V. (2012). Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res* 40, e16.
- Wu, H. C., John, E. M., Ferris, J. S., Keegan, T. H., Chung, W. K., Andrulis, I., Delgado-Cruzata, L., Kappil, M., Gonzalez, K., Santella, R. M., and Terry, M. B. (2011). Global DNA methylation levels in girls with and without a family history of breast cancer. *Epigenetics* 6, 29-33.
- Xiang, S., Liu, Z., Zhang, B., Zhou, J., Zhu, B. D., Ji, J., and Deng, D. (2010). Methylation status of individual CpG sites within Alu elements in the human genome and Alu hypomethylation in gastric carcinomas. *BMC Cancer* 10, 44.
- Xie, H., Wang, M., Bonaldo Mde, F., Smith, C., Rajaram, V., Goldman, S., Tomita, T., and Soares, M. B. (2009). High-throughput sequence-based epigenomic analysis of Alu repeats in human cerebellum. *Nucleic Acids Res* 37, 4331-4340.
- Xing, J., Hedges, D. J., Han, K., Wang, H., Cordaux, R., and Batzer, M. A. (2004). Alu element mutation spectra: molecular clocks and the effect of DNA methylation. *J Mol Biol* 344, 675-682.
- Xing, J., Witherspoon, D. J., Ray, D. A., Batzer, M. A., and Jorde, L. B. (2007). Mobile DNA elements in primate and human evolution. *Am J Phys Anthropol Suppl* 45, 2-19.
- Xing, J., Zhang, Y., Han, K., Salem, A. H., Sen, S. K., Huff, C. D., Zhou, Q., Kirkness, E. F., Levy, S., Batzer, M. A., and Jorde, L. B. (2009). Mobile elements create structural variation: analysis of a complete human genome. *Genome Res* 19, 1516-1526.
- Xu, Y., and Price, B. D. (2011). Chromatin dynamics and the repair of DNA double strand breaks. *Cell Cycle* 10, 261-267.



- Yan, P. S., Rodriguez, F. J., Laux, D. E., Perry, M. R., Standiford, S. B., and Huang, T. H. (2000). Hypermethylation of ribosomal DNA in human breast carcinoma. *Br J Cancer* 82, 514-517.
- Yang, A. S., Estecio, M. R., Doshi, K., Kondo, Y., Tajara, E. H., and Issa, J. P. (2004). A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 32, e38.
- Yang, B., Cao, L., Liu, B., McCaig, C. D., and Pu, J. (2013). The transition from proliferation to differentiation in colorectal cancer is regulated by the calcium activated chloride channel A1. *PLoS One* 8, e60861.
- Yoder, J. A., Walsh, C. P., and Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13, 335-340.
- Zhang, W. C., Shyh-Chang, N., Yang, H., Rai, A., Umashankar, S., Ma, S., Soh, B. S., Sun, L. L., Tai, B. C., Nga, M. E., et al. (2012). Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell* 148, 259-272.
- Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev* 15, 2343-2360.
- Zhumabayeva, B., Chenchik, A., Siebert, P. D., and Herrler, M. (2004). Disease profiling arrays: reverse format cDNA arrays complimentary to microarrays. *Adv Biochem Eng Biotechnol* 86, 191-213.

**ANNEXOS**

## ENCEBADORS D'EXPRESSIONI

La seqüència dels encebadors es mostra en direcció 5'-3'. F (encebador directe, la F prové del terme anglès *Forward*) i R (encebador invers, la R prové del terme anglès *Reverse*). S'indica la temperatura d'amplificació segons l'encebador que estudiem, la mida de l'ADNc amplificat i l'eficiència de la PCR. Es va emprar el *Light Cycler® 480*.

**TAULA AI.1. ENCEBADORS D'EXPRESSIONI EMPRATS COM A CONTROLS D'EFICIENCIA DEL TRACTAMENT AMB DROGUES EPIGENETIQUES**

GENS	encebadors	Temperatura amplificació	mida del fragment (pb)	Eficiència
EN1	F: CAGAGGCCAGGATCGCAT R:TCACCCCAGTCCAGTCACA	62	57	1.80
CPLX2	F:CCTCCAGACCCCCACCCATC R:AGGCTTCCCCGCGGCTTCTCAG	62	173	1.72

**TAULA AI.2. ENCEBADORS EMPRATS COM A ESTANDARDS D'EXPRESSIONI**

ESTANDARDS	encebadors	Temperatura amplificació	mida del fragment (pb)	Eficiència
18S	F:GCGAAAGCATTGCCAAGAA R:CATCACAGACCTGTTATTGC	60	510	1,70
CICLOFILINA A (PP1A)	F:CTCCTTTGAGCTGTTTGAG R:CACCACATGCTTGCCATCC	60	325	1,62
$\beta$ -2 MICROGLOBULINA	F:CCAGCAGAGAATGGAAAGTC R:GATGCTGCTTACATGTCTCG	60	269	1,90
PSMC4	F:TGTTGGCAAAGGCGGTGGCA R:TCTCTGGTGGCGATGGCAT	60	182	1,82
PUM1	F:CGGTCGTCCTGAGGATAAAA R:CGTACGTGAGGCGTGAGTAA	60	121	1,66
MRPL19	F:CAGTTTCTGGGGATTTCAT R:TATTCAGGAAGGGCATCTCG	60	197	1,72

**TAULA AI.3. ENCEBADORS EMPRATS PER L'ESTUDI D'EXPRESSIONI DEL GEN *DIEXF***

REGIÓ (EXONS)	encebadors	Temperatura amplificació	mida del fragment (pb)	Eficiència
3_4	F: GACAGTATTGTAGATGATGC R: CTGAACAGTGACTCGTGT	63	212	1.76
6_7	F: AGACGCCGAAGCCAGAAGTT R: CCTCGAGGAGGCTGATGAA	63	142	2.02
8_9	F: TCATTGGTGGAGAAGGAGAG R: ATGTGAGTCCAGGGGTAGT	63	155	1.94
11_12	F: CCTTCAAGGAGAGAAAACAGT R: GGTGGCTCTCAGCATATTAC	63	151	2.00
12-UTR003	F: GCAGGAAGTGGTATTTGGCA R: GGCTGGGGCAGAATAAGTTT	63	478	1.71
UTR003	F: TCCTGTTCCCTTCTGTGAG R: CCTGGAGCTCTGTTCTTCT	63	169	1.83
UTR002	F: GTCATTTGTCAGAGGAGGAA R: GGGCTTACAATAACTTGCTG	61	185	1.81

**TAULA AI.4. ENCEBADORS EMPRATS PER L'ESTUDI D'EXPRESSIONI DEL GEN *GLDC***

GEN	encebadors	Temperatura amplificació	mida del fragment (pb)	Eficiència
<i>GLDC</i>	F: TCTGGAGATCGTATATTGGC R: GTAAACTCTCCAGCCTCCC	58	144	2.03

**TAULA AI. 5. ENCEBADORS D'EXPRESSIONI EMPRATS PER L'ESTUDI DE DIFERENCIACIÓ**

GENS	encebadors	Temperatura amplificació	mida del fragment (pb)	Eficiència
<i>GATA6</i>	F: TCCCCACAACACAACCTAC R: ACGCCTATGTAGAGCCCATC	62	200	1.89
<i>OCT4</i>	F: AAGCGATCAAGCAGCGACTAT R: GGAAAGGGACCGAGGAGTACA	62	163	1.86

## ENCEBADORS DE METILACIÓ

La seqüència dels encebadors es mostra en direcció 5'-3'. F1(encebador directe extern), R1 (encebador invers extern), F2 (encebador directe intern) i R2 (encebador invers intern).

(a) Temperatura d'hibridació de la PCR. (b) Concentració de la que partim del producte de la PCR externa per realitzar la PCR interna. (c) Mida en pb del producte final. (d) Encebador intern necessari per realitzar la PCR de seqüenciació.

TAULA A1.6. ENCEBADORS EMPRATS PER L'ESTUDI DE L'ESTAT DE METIL·LACIÓ DELS ELEMENTS ALU

ELEMENT ALU	encebadors externs	Tm PCR externa (a)	producte (b)	encebadors interns	Tm PCR interna (a)	mida del fragment (pb) (c)	Encebador intern ( d )
<b>Ar3c7</b>	F1:GGGGTGATTTAATTAGATT R1:TTATAACTACTAAATAAAATATACA	Tm1=52 Tm2=54	dilució 1/10	F2:GGGAGGAGTATATATATATAT R2:ATAAACTATCTCCTACCATC	Tm3=52 Tm4=54	402	F2
<b>Lv165</b>	F1:GGGAAAAAGTTAGTTTAG R1:TAAAAAAGTTCAACCCTCTTA	Tm1=50 Tm2=52	mare	F1:GGGAAAAAGTTAGTTTAG R2:CACCAAAAACATACTATACA	Tm3=54 Tm4=56	370	R2
<b>Lv183</b>	F1:GTAGATTAAGTAGAAAGAGG R1:ACCAACTAAATAAACAAATATAA	Tm1=54 Tm2=56	mare	F2:GGAGGATTTGTTTATTTTTATGATG R2:ACACACACTAACCAAAAAATAAT	Tm3=62 Tm4=64	264	F2
<b>Au104</b>	F1:GGATTTTTTGTAGTTTTGTA R1:ATCACACTCCACCITATCT	Tm1=52 Tm2=54	dilució 1/20	F2:AGGTTTGTGTATGTTAT R2:AAAAATAAACCAACCACAAACCA	Tm3=54 Tm4=56	388	F2
<b>Aj2c1</b>	F1:TAGATTTAATGGGTAAAAGTA R1:CAATAAAGTTTACAATATACTAC	Tm1=50 Tm2=52	mare	F2:TGTGGTTGTATAGTTAAGG R2:ACCAACATTTAAAAATACCTAA	Tm3=54 Tm4=56	495	R2
<b>Av2c2</b>	F1:TTTATAAAAAGTAGTATATGTTGA R1:AAAAATATAAATAAATACCTCAA	Tm1=52 Tm2=54	mare	F2:TAAAGTGGAAATAGAGTTAAT R2:TACTCCAAAAACATAAACAAATAAT	Tm3=54 Tm4=56	812	R2
<b>Ar3c5</b>	F1:TTAGATTAATTTAGTTTTAGTT R1:TTATCTACATACTTCATAACAA	Tm1=52 Tm2=54	mare	F2:TATGGGTTTATTTGGGTTTAT R2:TCTAACCTATAAAAAATTTAACAA	Tm3=54 Tm4=56	429	R2
<b>Au60</b>	F1:AATTTGGGATGTTTATAAGAT R1:CACCTCACCTTTAAATTAATCTTA	Tm1=52 Tm2=54	dilució 1/20	F2:TGGTTAAGGTTTTTGGGATTT R2:ATCACATTAATCTTCAACTACC	Tm3=54	526	R2
<b>At1c3</b>	F1:TTTGGGGTTGAGGTGAGAAATG R1:AATAAAAAGTCTAAAATACATAAAA	Tm1=52 Tm2=54	mare	F2:GGGGTGGTTAAAAAGTAAAGAA R2:TCTATTTTTCTTTTCTTCTTCTC	Tm3=54 Tm4=56	412	F2

ELEMENT ALU	encebadors externs	Tm PCR externa (a)	producte (b)	encebadors interns	Tm PCR interna (a)	mida del fragment (pb) (c)	Encebador intern (d)
<b>Ah1c1</b>	F1:GTAAATAAGGGGAGGTAGA	Tm1=52	dilució 1/10	F2:GATGGGATATTGTTGAAAA	Tm3=54	691	F2
	R1:TTAACITTAACCATCCACAIT	Tm2=54		R2:CAATCTACTTCCAAAAATCCAT	Tm4=56		
<b>Aq3c6</b>	F1:GIGTTTTATGTTGGGG	Tm1=50	dilució 1/20	F2:ATTTGAGTGTGTTGTGT	Tm3=54	215	R2
	R1:CCCAAACITCTCCAAAAACTA	Tm2=52		R2:ATAACAAAAAACACTACACAA	Tm4=56		
<b>Ac4c1</b>	F1:GTTAATAATAAAGTAAAGGTATA	Tm1=52	mare	F2:AAAATTTAAAGTTAAGAGTGG	Tm3=54	318	F2
	R1:AAATAACATACIATAIAATAATA	Tm2=54		R2:CCACTACTATAIATCCCTAAA	Tm4=56		
<b>Ao1c4</b>	F1:GTTTAAAAATTTTTGTTGGGA	Tm1=52	mare	F2:GATAIATGTTTATTATTTAIGTIT	Tm3=52	413	F2
	R1:ACCTATAACCTCTAAAAACCA	Tm2=54		R2:ATTTTTTCTTTTTCTATTTTTTAA	Tm4=54		
<b>Ah2c10</b>	F1:TTGAATTTAAAAATGTTTGAGTA	Tm1=52	mare	F2:TTTGGTTAAAAATGTTTGAGTAAAT	Tm3=56	342	F2
	R1:TTTTAATATAAAATAATAACTTCC	Tm2=54		R2:AACTTCCAAAATAATCAAACTAA	Tm4=58		
<b>Ao2c1</b>	F1:ATATTTAGAGAAATTAATGGTIT	Tm1=52	dilució 1/10	F2:TGGATTTTGGGAAAGTAGATTT	Tm3=54	238	F2
	R1:CCATTTACCATAACCCITATAA	Tm2=54		R2:ACCTTCAAAAACCCCAATCAA	Tm4=56		
<b>Ag2c5</b>	F1:ATATTTATTTAGATTTTAGTTAAG	Tm1=52	mare	F2:GTTTTTGTGTTTTGTTAAAAGTTTT	Tm3=54	424	R2
	R1:ACATAATCCCTACCATAAAACT	Tm2=54		R2:TTACTCAATCAAAATACAACAAAT	Tm4=56		
<b>Ar3c1</b>	F1:TATATGAGTAGTTATGGAAA	Tm1=52	dilució 1/10	F2:AGTTAATTTGAAAATTTAGTGGAT	Tm3=54	377	F2
	R1:ACCTACATACCTAAACTAAT	Tm2=54		R2:AACTTTAATAACTTTTAAATCCTTT	Tm4=56		
<b>ZALU3</b>	GGTTAATTTTTTAAAGTAAGA	Tm1=50	dilució	GATAGTGAATAATGTTTATTT	Tm3=52	336	F2
	AAAAATTTCAAACACTACAATATA	Tm2=52	1/10	TATATATAAAACTACAAAAAATA	Tm4=54		

TAULA A1.7. ENCEBADORS EMPRATS PER L'ESTUDI DE L'ESTAT DE METILACIÓ DE LES REGIONS FLANQUEJANTS ALS ELEMENTS ALU

ELEMENT ALU	Regió flanquejant	encebadors externs	Tm PCR externa (a)	producte (b)	encebadors interns	Tm PCR interna (a)	mida del fragment (pb) (c)	Encebador intern (d)
Alu Jo		F1:GTGAATATTAAGATTTATTT	Tm1=48	mare	F1:GTGAATATTAAGATTTATTT	Tm3=50	244	F2
		R1:ATTATATACTTACAATTTTC	Tm2=50		R2:TTTCRTTTTATTTATTTTAAATA	Tm4=52		
CpGi		F1:TATTTAAGTATAGAGTAGTTTT	Tm1=50	mare	F2:GGTTAGTGTTTGTAAATTTTT	Tm3=52	394	R2
		R1:CTTTAAAAATAAATACTTTC	Tm2=52		R2:CTAAAAAAACCCAAAAATAATCCT	Tm4=54		
Ar3c7		F1:GGGGTGATTTATTTATAGATT	Tm1=52	diluïció 1/10	F2:GGAAGGAGTAGTATATATAT	Tm3=52	402	F2
		R1:TTATAACTACTAATTAATAACA	Tm2=54		R2:ATAAACTATCTCCTACCATC	Tm4=54		
Mir b		F1:TTTATGTTTAGGGGGGAG	Tm1=52	mare	F2:GTAGGAAAGATAGTTTAGTAT	Tm3=54	243	R2
		R1:TCACCTAAAAATTCAAAAACT	Tm2=54		R2:CAACCCAAAAACCCAAATACCAA	Tm4=56		
Alu sx		F1:GATATGAGGTATATTTTATA	Tm1=50	diluïció 1/20	F2:TATGATGTGTGGTTAATAATA	Tm3=52	357	R2
		R1:TTCTTCTCCTCTAAAAAT	Tm2=52		R2:TTAACATATATCCCTTAATAAAA	Tm4=54		
Alu sq		F1:GGTTATTAATAATGTTAAGATTG	Tm1=50	mare	F2:TTTGGGTTTATGGTTTATATTT	Tm3=54	370	F2
		R1:ATCTTCACATTAATAAATTTTT	Tm2=52		R2:CACATTAATAAATCTTTTTTCTTT	Tm4=56		
L2		F1:GTTAATGGTATAGTTAGTTT	Tm1=50	diluïció 1/20	F2:ATGAGGGTTTTGTAGAGAG	Tm3=52	434	F2
		R1:ATACACCAATAAAACTTATA	Tm2=52		R2:ATACACCAATAAAACTTATA	Tm4=54		
CpGi		F1:AAITATAAGTTTATGGTGTA	Tm1=50	mare	F2:ATAAGGATGTGTTTTTGAGT	Tm3=52	302	F2
		R1:ATAAAAAATCCATAAAAAATAAAT	Tm2=52		R2:AAATACAAATAAAAAACCCAAAAAAA	Tm4=54		
Lv165		F1:GGGAAAAATGTTTAGTTTAG	Tm1=50	diluïció 1/20	F1:GGGAAAAATGTTTAGTTTAG	Tm3=54	370	R2
		R1:TAATAAACTTCACCCCTCTTA	Tm2=52		R2:CACCAAAAACATACTATAACA	Tm4=56		
Mir 3		AAGGGATGGTTAGTTAG	Tm1=52	mare	F2:GTAGGTTTGTGAAGGTTTAG	Tm3=54	330	R2
		CCAACCTCCCTCTCTTTAT	Tm2=54		R2:ATTATAAAAAACATACTATCCTATA	Tm4=56		
CG rich region		F1:ATTTGGGTATGTTATTTTGTG	Tm1=50	mare	F2:AGGGATTAGTGTGTTTAGTAGT	Tm3=52	250	R2
		R1:CTCCAAAAAAGTTATAAAAAATA	Tm2=52		R2:AACTAAAAAACCTATCCTAAACC	Tm4=54		



ELEMENT ALU	Regió flanquejant	encebadors externs	Tm PCR externa (a)	producte (b)	encebadors interns	Tm PCR interna (a)	mida del fragment (pb) (c)	Encebador intern (d)
Alu sx		F1:ATGAAATATTTAAATTTTAGG	Tm1=46	mare	F2:GGTAAAGATAATTTAAAGTTA	Tm3=48	427	F2
		R1:CAATACAAAATAACAACAACACT	Tm2=48		R2:CAATACAAAATAACAACAACACT	Tm4=50		
Alu sq		F1:AGTTTTAAATTTATTTTGAAG	Tm1=50	mare	F2:AGTTGTGTATTTGTATTG	Tm3=50	368	F2
		R1:TTTAAATTTTACTAAACACAAA	Tm2=52		R2:CACAAAAAAAATTTACTAAAAATC	Tm4=52		
CpGi		F1:AAAACCATCAAAACCTCCC	Tm1=48	mare	F2:CCCCTAACAAACCCCTCCA	Tm3=54	444	F2
		R1:TTGAGTAGAGTTTGAGTTTG	Tm2=50		R2:TAGAGATTGGGAAAAGTTATA	Tm4=56		
<b>Au104</b>		F1:GGATTTTTGTAGTTTTGTGA	Tm1=52	mare	F2:AGTTTGTGTGTATGTAT	Tm3=54	388	F2
		R1:ATCACACTCCACCTATCT	Tm2=54		R2:AAAAATAAACCAACCCACAAACCA	Tm4=56		
Alu sx		F1:GTATTGTGTTTTGTGATGG	Tm1=52	mare	F2:TTAGATAAGGTGGAGTGTGA	Tm3=54	330	R2
		R1:ACCTAAATATTTATCTAATAAC	Tm2=54		R2:TAACATAAACCTCTAAIAAACA	Tm4=56		
Alu sq		F1:GGATAGTTGGTTTTATAT	Tm1=50	mare	F2:TTTTAGGGAGGAAGATATAT	Tm3=52	250	F2
		R1:TTTCATCTCTTCTTCTCC	Tm2=52		R2:ACTAAATTTCAAATTTTACAATAA	Tm4=54		
<b>LV183</b>		F1:GTAGATAAGTAGAAAAGAGG	Tm1=54	mare	F2:GGAGGTATTTGTTTTATGATG	Tm3=62	264	F2
		R1:ACCAACTAAATAAACACATAA	Tm2=56		R2:ACACACACTAACCAAAAAATAAT	Tm4=64		
CpGi		F1:TGGTTTTTTTGTATTTTGA	Tm1=52	mare	F2:TTTTAGAGTAGTTTGTGAA	Tm3=54	238	R2
		R1:ATCCCTCTCCATAAAATAAC	Tm2=54		R2:AAATAAAAAAACTCCCTAAAAATAAC	Tm4=56		
Alu sp		F1:GGATTTAAGACACAGGTTTG	Tm1=48	dil·lució	F2:ATATATAGATGATAGTTGTAAGA	Tm3=50	400	R2
		R1:AAATAAAAAAAATCCTTCAAAAC	Tm2=50	1/10	R2:ACAAATCTTATTAATAACTATAA	Tm4=52		
Alu sx		F1:TAGGATGGTTTTATTTTTG	Tm1=50	mare	F2:TAAAAGTGTGGGATATAGG	Tm3=52	243	R2
		R1:TTCCCTAAATAATTTCTAAC	Tm2=52		R2:CTACCTCAACATTCCTCTAA	Tm4=54		
CG flich region		F1:GTTTGAAGTAGTTATAAAAC	Tm1=48	mare	F2:GTTTGTGTTTTAAGTGTTTT	Tm3=52	357	F2
		R1:TAAAACAAAACCCAAAACATA	Tm2=50		R2:CTACAACCCCTTTCTCTATA	Tm4=54		

ELEMENT ALU	Regió flanquejant	encebadors externs	Tm PCR externa (a)	producte (b)	encebadors interns	Tm PCR interna (a)	mida del fragment (pb) (c)	Encebador intern (d)
	CG rich region	F1: GTTAGTATATAATATAT	Tm1=48	mare	F2: TGTGTTGATGTTTTTATAT	Tm3=50	291	F2
		R1: CTATACTACCAATATCTAT	Tm2=50		ACAAAATACITAAATCATAAAC	Tm4=52		
<b>Aj2c1</b>		F1: TAGATTTTAAATGGTAAAAGTAA	Tm1=50	mare	F2: TGTGGTGTATAGTTAAGG	Tm3=54	495	R2
		R1: CAATAAACCITTAACAATATACIAC	Tm2=52		R2: ACCAACAATTTAAAAATACCTAAA	Tm4=56		
	CpGi	F1: ATATAAGTATAAAGTTTATGTTT	Tm1=48	mare	F2: GTTTTAGGGTTTTGAGAAATAT	Tm3=52	287	R2
		R1: TATCATAAAAAAAATACTCCTC	Tm2=50		R2: ATTAATAAATAAATACTTAAGTCT	Tm4=54		
<b>Aj2c1</b>	Alu sx	F1: GGTITTTTGTGAGTTTGT	Tm1=50	mare	F2: TGTAAATTTTTGTTAAGAGGTA	Tm3=52	374	F2
		R1: ATATCTACAAAAAATTCCTAAA	Tm2=52		R2: TCTCTAACCCACACAAAAAATC	Tm4=54		
	Alu sx II	F1: ATTTTTGTGTTAGAGA	Tm1=52	mare	F2: TTTAGGAAATTTTTGTAGATAT	Tm3=52	256	F2
		R1: CACTATATCCCACAAAATATA	Tm2=54		R2: CACCCTAACTAACTAACTAAATA	Tm4=54		
	Alu Jo	F1: AATTAAGTGTATAGTAGTT	Tm1=48	mare	F2: TGTGTTGTTATAGTTAGGG	Tm3=50	448	F2
		R1: AAAATCACATAACAAACCC	Tm2=50		R2: AAATCACACAAAACTCTCAT	Tm4=52		
	Alu Sg	F1: TTATATATATAGATGAGTATA	Tm1=48	mare	F2: GTATATATATATAATGTTAGT	Tm3=50	452	R2
		R1: AAAATCAAACCTCCAAATAAA	Tm2=50		R2: AAAAAATCCACCCTTATCA	Tm4=52		
	CpGi 5'	F1: GGGTAAAGTTATGAGGAT	Tm1=48	mare	F2: GGGTGGAGGTTTGTAGTGT	Tm3=50	413	R2
		R1: CCCATAAAAAACACAATAATA	Tm2=50		R2: CCCATAAAAAACACAATAATA	Tm4=52		
	CpGi 3'	F1: GTTTTTAGGGGTTTGAAG	Tm1=52	dilució	F2: GTTAGGTGAGGGTGGAAA	Tm3=52	186	R2
		R1: CACTAAAAAACAATAACTCAC	Tm2=54	1/20	R2: ACAAAAAAATAAACCACAC	Tm4=54		
<b>AU60</b>	Alu sx	F1: GTGTGAGTTTGTGTTTATGIG	Tm1=60	mare	F2: GTGAGTTTGTGTTTATGTTGTT	Tm3=62	619	F2
		R1: CACCCTACCCCTACCTCCC			R2: CAAAAATCAACTCATCCCTCCC			
<b>AU60</b>		F1: AATTGTGGATGTTTATTAAGAT	Tm1=52	dilució	F2: TGGTTAAGGTTTTTGGGATTT	Tm3=54	526	R2
		R1: CACTCACCTTTAAAAATACITTA	Tm2=54	1/20	R2: ATCACATATCTTCAACTACC			
	Alu sg	F1: GGTATGGTTAGAGTAG	Tm1=50	mare	F2: GGTATGGTTAGAGTAG	Tm3=52	450	R2
		R1: CTATAAAAAAATACTCAAAATCCC	Tm2=52		R2: TTTAAAAACTATACACTATTTCC	Tm4=54		

ELEMENT ALU	Regió flanquejant	encebadors externs	Tm PCR externa (a)	producte (b)	encebadors interns	Tm PCR interna (a)	mida del fragment (pb) (c)	Encebador intern (d)
CpG up		F1:GATTATAAAGTTTTGTGG	Tm1=48	dilució	F2:GAGTAAAGTTAGAAATGATT	Tm3=50	233	F2
		R1:TCTTAAACAACCTTACTCT	Tm2=50	1/10	R2:AAACAACCTTACTTCTCCT	Tm4=52		
CpGi 5'		F1:TAGTAGAGTTTAGGGTAG	Tm1=48	dilució	F2:GTTTAAAGGTTTGTAGTATT	Tm3=50	178	F2
		R1:CAAAAAAAAAAACCACAAAAACC	Tm2=50	1/10	R2:CCATACAATCTTATACCAA	Tm4=52		
CpGi 3'		F1:GTTTTTAAAGGTTATTGT	Tm1=50	mare	F2:GTTAATGAAAAGTAGATAGAG	Tm3=54	178	R2
		R1:AACTTAAACTATCCAC	Tm2=52		R2:ACCTAATCTTACCCCAAATA	Tm4=56		
Mir C + Alu Jb		F1:GATGGGATTTAAAAAGAGA	Tm1=50	dilució	F2:GGAAAAGTATTTGAAATAGTAGA	Tm3=56	327	R2
		R1:ACATCTCACTACAACCTAA	Tm2=52	1/10	R2:TCAACCTCCAAAAATAACTAA	Tm4=58		
<b>ZALU3</b>		F1:GGTTAATTTTAAAAAGTAAGA	Tm1=50	dilució	F2:GATAGTGAATAATGTTTATT	Tm3=52	336	F2
		R1:AAAAATTCAAACTACAATAATA	Tm2=52	1/10	R2:TATATATAAAAACTACAAAAAATA	Tm4=54		
Alu sx6		F1:GGTGATGGATGTTATTGTTT	Tm1=54	dilució	F2:TATGTATATGAGAATTTGGGGT	Tm3=56	290	R2
		R1:ATTCTCTCTCTCTCTTTT	Tm2=56	1/10	R2:CTTACTCTATCACCCCAAATTTAAA	Tm4=58		

**TAULA A1.8. ENCEBADORS EMPRATS PER L'ESTUDI DE L'ESTAT DE METILACIÓ DELS ELEMENTS ALU EN PRIMATS**

En groc marquem aquells encebadors que no coincideixen amb els encebadors d'humà. I en minúscula i vermell aquells nucleòtids diferents a la seqüència en humà

ELEMENT ALU	espècie	encebadors externs	Tm PCR externa (a)	producte (b)	encebadors interns	Tm PCR interna (a)	mida del fragment (pb) (c)	Encebador intern ( d )	
<b>Ar3 c7</b>	humà	F1:GGGGTGATTATTAGATT	Tm1=52	dilució	F2:GGAAGGAGTAGTATAITATAT	Tm3=52	402	F2	
		R1:TTATAACTACTAATAATAACA	Tm2=54	1/10	R2:ATAAACIATCTCCACCATC	Tm4=54			
	chimpanzé/gorila	F1:GGGGTGATTATTAGATT	Tm1=52	mare ó	F2:GGAAGGAGTAGTATAITATAT	Tm3=54	402	R2	
		R1:TTATAACTACTAATAATAAaAA	Tm2=54	1/10	R2:ATAAACIATCTCCACCATC	Tm4=56			
orangutan	humà	F1:GGGGTGATTATTAGgT	Tm1=52	mare ó	F2:GGAaAGAGyGGTATAITATAT	Tm3=54	402	R2	
		R1:TAACACTAATAATAAaAACT	Tm2=54	1/10	R2:ACTATCTCCACCATCC	Tm4=56			
	humà	F1:GTGTTTATGTGGGG	Tm1=50	mare	B16.2_F2:ATTGAGTGTGTTTGTGTT	Tm3=54	215	R2	
		R1:CCCCAACTCTCCAAAACTA	Tm2=52		B16.2_R2:ATAACAAAAAACACTACACAA				
<b>Aq3 c6</b>	chimpanzé/gorila	F1:GTGTTTATGTGGGG	Tm1=50	dilució	F2:gtTGAGTGTGTTTGTGTT	Tm3=52	215	F2	
		R1:CCAAACTCTCCAAAACT	Tm2=52	1/10	R2:ATCACAAAAAACACTACACAA	Tm4=54			
	orangutan		X	X		X	X	X	X
		humà	F1:GGATTTTTGTAGTTTTGTA	Tm1=52	mare	F2:AGGTTTGTGTTATGTTAT	Tm3=54	388	F2
	R1:ATCACACTCCACCTTATCT	Tm2=54		R2:AAAAATAAACAAACCAAAACCA	Tm4=56				
<b>AU104</b>	chimpanzé/gorila	F1:GGATTTTTGTAGTTTTGTA	Tm1=52	dilució	F2:AGGTTTGTGTTATGTTAT	Tm3=54	388	F2	
		R1:CCATCAACAAAAACAATAC	Tm2=54	1/10	R2:AAAAACAACCAACCAAAACCA	Tm4=56			
	orangutan	F1:GGATTTTTGTAGTTTTGTA	Tm1=52	dilució	F2:AGGTTTGTGTTATGTTAT	Tm3=54	388	F2	
		R1:ATCACACTCCACCTTATCT	Tm2=54	1/10	R2:AAAAACAACCAACCAAAACCA	Tm4=56			

ELEMENT ALU	espècie	encebadors externs	Tm PCR externa (a)	producte (b)	encebadors interns	Tm PCR interna (a)	mida del fragment (pb) (c)	Encebador intern (d)
LV183	humà	F1:GTAGATTAAGTAGAAAGGG	Tm1=54	dilució	F2:GTATTGTTTTTATGATG	Tm3=54	264	F2
		R1:ACCAACTAAATAAACACATAA	Tm2=56	1/20	R2:ACACACACTAACCAAAAAATAAT	Tm4=56		
	chimpanzé/gorila	F1:GTAGATTAAGTAGAAAGGG	Tm1=52	mare ó	F2:GTATTGTTTTTATGATG	Tm3=52	264	F2 o R2
		R1:cTAACCAAAAAATAATATAAAATa	Tm2=54	1/10	R2:TACCAACTAAATAAACCAACATAA	Tm4=54		
orangutan	F1:GTGAGAAATTTGTtATATA	Tm1=50	mare ó	F2:taTATATTAGTTAGGAGTGGT	Tm3=50	264	R2	
	R1: AATAAACCAAAACAAAAaAAAC	Tm2=52	1/10	R2: ACTTACCTTATCCCCATAA	Tm4=52			
LV165	humà	F1:GGGAAAAAGTTTGTAGTTTGG	Tm1=50	dilució	F1:GGGAAAAAGTTTGTAGTTTGG	Tm3=54	370	R2
		R1:TAATAAACATCACCCCTCTTA	Tm2=52	1/20	R2:CACCAAAAAACATACTATAACA	Tm4=56		
	chimpanzé	F1:GGGAAAAAGTTTGTAGTTTGG	Tm1=52	mare	F1:GGGAAAAAGTTTGTAGTTTGG	Tm3=58	370	R2
		R1:TAATAAACATCACCCCTCTTA	Tm2=54	mare	R2:CACCAAAAAACATACTATAACA	Tm4=56		
orangutan	F1:GGGAAAAAGTTTGTAGTTTGG	Tm1=52	dil1/10	F1:GGGAAAAAGTTTGTAGTTTGG	Tm3=58	370	R2	
	R1:TAATAAACATCACCCCTCTTA	Tm2=54	dil1/10	R2:CACCAAAAAACATACTATAACA	Tm4=56			
AJ2c1	humà	F1:TAGATTTAAATGGGTAAAGTA	Tm1=50	mare	F2:TGTGGTTGTATTAGTTAAGG	Tm3=54	495	R2
		R1:CAATAAATTTACAATATACTAC	Tm2=52	mare	R2:ACCAACATTTAAAAATAACCTAA	Tm4=56		
	chimpanzé	F1:TAGATTTAAATGGGTAAAGTA	Tm1=50	mare	F2:TGTGGTTGTATTAGTTAAGG	Tm3=54	495	R2
		R1:CAATAAATTTACAATATACTAC	Tm2=52	mare	R2:ACCAACATTTAAAAATAACCTAA	Tm4=56		
orangutan	F1:GTAGTGAATTTATAGTATATTA	Tm1=50	mare	F2:AITAAATTTAAAGGGGTGTTTGA	Tm3=54	495	R2	
	R1:CACACACAAAAACAACAAACC	Tm2=52	mare	R2:TATAATTTACATCAATTAATAA	Tm4=56			

**ENCEBADORS PER CHIP**

La seqüència dels encebadors es mostra en direcció 5'-3'. F (encebador directe) i R (encebador invers). S'indica la temperatura d'amplificació segons l'encebador que estudiem, la mida de l'ADNc amplificat i l'eficiència de la PCR. Es va emprar el *Light Cycler® 480*.

**TAULA AI.9. ENCEBADORS DISSENYATS PER L'ESTUDI DE CHIP DELS ELEMENTS ALU**

Element ALU	encebadors	Temperatura amplificació	mida del fragment (pb)	Eficiència
Ar3c1	F:AGTAACTGAAATTTAGTGGA R:GGAAGATAACTTTGATGTCC	52	385	1.84
Ah2c10	F:GCTAAAATGTTGAGTAATCTG R:TTAGTGTGGATAATTGGCTT	60	377	1.62
Av2c2	F:TAAGTGGAAGTACAGAGCC R:TGCTCCAAAACATAGACAGTA	60	316	1.84
Ag2c5	F:TGCCCTTGCCAAAGTTCTA R:TAATAAAGAAATAGACCTGTCA	56	342	1.61
Ao2c1	F:TGGATTCTGGAGAAGTAGAT R:CCTTCAAGGCCCCATTCA	56	244	1.86
Ar3c5	F:GGGTTCACTTGGGTTAC R:TGACCTATGGGGTATTGG	56	424	2.03
Ac4c1	F:TTCAAAGTCATCAAGAGTGG R:ACTGCTGTGTGCCCTGA	56	313	1.51
Ah1c1	F:GATGGGATTATCTGCCTGA R:ATTCTGCTCCAGAATCCAT	60	398	1.84
Ao1c4	F:TCACTTATGTTCTTTCTACATT R:ACGGTGTTTTTTCTTTTTCT	59	326	1.82
Aj2c1	F:TTGCATCAGTTAACGTTTTTCT R:ATATCAACCAACATTTAAGGGT	55	371	1.79
Aj2c1 5'	F:TTGCATCAGTTAACGTTTTTCT R:ATATCAACCAACATTTAAGGGT	55	172	1.96
Aj2c1 3'	F:TTGCATCAGTTAACGTTTTTCT R:ATATCAACCAACATTTAAGGGT	55	199	1.78
Aq3C6	F:GTGGGGCGGTGACAAAGACG R:AGGGGCACTGCACAGAAACGA	63	346	1.97
Lv165	F:GGGAAAATGCCAGTTTCTAG R:GCACCCCAAGTGTTCAAAAG	61	354	2.06
Ar3c7	F:AAGGAGCAGCCACACCATAT R:GTGAACTATCTTCTACCAT	56	400	1.85
Lv183	F:GCACTTGCCTTATTCCCA R:ATAGCGAGAACCCGTGTAG	61	379	1.83
Au104	F:CTGCCTGTGTCATGCCATT R:TAAGGATGGACAGCCACAG	62	385	2.15
Au60	F:CTTTGTTGCAACATGGTCCTA R:CCTTAACTCAAGCACTGATGA	63	346	1.84
ZALU3	F:GCCACAGAACGAGACTCAG R:CGCCTGTAATGACGGATCACGA	64	376	2.04

Element ALU	encebadors	Temperatura amplifació	mida del fragment (pb)	Eficiència
qAlu	F: ATTCTCCTGCCTCAGCCT R: CCGTCTCTACTAAAAATACA	56	87 *	1,96
GAPDH	F: TTGCAACCGGGAAGGAAA R: TAGCCTCCGTCAGCTGACTT	60	120	1,85
16CEN	F: CAGTTTGCCAGATAGTCTCTTT R: GAGACATTTGGAAGGTCAGTGAAT	60	211	1,83
CPLX2	F: CCTCCAGACCCACCCATCC R: AGGCTCCCCGCGGGCTTCTCAG	66	173	2
THOC3	F: TTCTTCTGGGGTGTTCGGTAATC R: CACACCCGCTAGCCCTTTTCAT	64	119	2,04
DRD1	F: GAGGACGCCCCGGTGAGTGC R: GCCGGCCGTCTAGGAGTTGGTT	66	245	2,07

\* Els encebadors de qAlu es van dissenyar segons la seqüència consensu dels elements Alu de diferents famílies. Els amplicons obtinguts no corresponen a un únic element Alu sinó a un nombre representatiu del total d'Alus del genoma i tenen una mida mitja de 87 pb.

**TAULA AI.10. ENCEBADORS DISSENYATS PER L'ESTUDI DE CHIP DE LES REGIONS FLANQUEJANTS ALS ELEMENTS ALU**

La denominació *up* es dona a aquells elements situats a l'extrem 5' de l'element Alu i el terme *down* als elements situats a l'extrem 3'.

Element ALU	Regió flanquejant	encebadors	Temperatura amplificació	mida del fragment (nb)	Eficiència
Ar3c7	LINE1	F:GGTTTGCTGTTTATAATT R:AAAACGTGCAAGTATACAA	50	331	1,83
	Alu Jo	F:GTTTTGTTTGTITTTAAATAGAGA R:TGTGAACATTTAAGATCTATT	54	242	1,96
	L1+L2	F:GAATAGATCTTAAATGTTTACACA R:ACAAAACCTAACAAATCCATTC	50	491	1,89
	CpGi	F:TCTGGTTAGTGTCTGCTCAAT R:GAGCACTCAGACACCCAGTAC	62	264	1,95
	<b>Ar3c7</b>	F:AAGGAGCAGCCACACCATAT R:GTGAACTATCTTCTACCAT	56	400	1,79
	Mir b	F:GTAGGAAGATAGTTCACAGTAC R:ACAGCCAAAACCAAATTGCCAA	60	244	1,85
	Alu sx	F:ACATGTATGTGTGGTTAATATA R:TTAGCATATTATCCTTAATAAGAA	54	358	1,84
	Alu Sp	F:CCATGACATCTCTCATATTATC R:CACAAATAGGAAATATGGACAC	60	342	1,96
Lv165	Alu sq	R:TATTTCTTTCCAAGACTTTG F:CAGGATCTACGGTCACAGA	60	385	1,93
	LINE2	F:GACATAAGGAATAGGAGGG R:CCCAAGTAACTGGCACCCAG	60	258	1,94
	CpGi	F:AGGGACTAGTGTCCCAGCA R:AAGAAAGCTAAAAGCCTGTC	60	263	1,95
	<b>Lv165</b>	F:GGGAAAATGCCAGTTCAG R:GCACCCCAAGTGTCAAAAAG	61	354	2,06
	Mir 3	F:GTCTCTAAGGGATGTG R:TGTAGAGCATGTCTGCCTAT	59	356	1,79
	CG rich region	F:AGGGACTAGTGTCCCAGCA R:AAGAAAGCTAAAAGCCTGTC	60	263	1,84
	Au104	Alu sx II	F:TTCTGAAGTTGCTGCTACT R:CAAAGAAATTAAGTAAATCGTG	58	372
Alu sx I		F:CAGTAATTTCTTTGTTCAGTA R:TTGCATTATTCGCATTCTTC	56	377	1,87
Alu sq		F:CTGAGCAGAGCCTGAGCCTG R:CCTGTCCGCCAGCCTCCA	70	228	1,83
CpGi		F:AAGGAGCAGCCACACCATAT R:GTGAACTATCTTCTACCAT	50	400	1,94
<b>Au104</b>		F:CTGCCTGTGTCATGCCATT R:TAAGGATGGACAGCCACAG	62	385	1,93
Alu sx		F:CGTACTGTGTTTGTGATGG R:GTCTTAACTGTGAACTCCTG	62	415	1,8
Alu sq		F:TTGAAAATTAGAATTGTGGGC R:GCTGAATTTCAAATTTACAGTGA	58	381	2,06



Element ALU	Regió flanquejant	encebadors	Temperatura amplificació	mida del fragment (nb)	Eficiència
Lv183	CG rich region	F:AAGGACCTGTCACTCCACTC R:GTAAAGTAGCATAGAAGGAGG	62	316	1,96
	Lv183	F:GCACTTGCCTTATCCCCA R:ATAGCGAGAACCCGTGTAG	61	379	1,83
	CpGi	F:AGGCTTTCCTCTACTATCAGATGA R:TGTCAGCCCTGGTTGGTTA	62	382	1,95
	Alu sp	F:GCCCTCTCCAGCACTTCTA R:GGTCCCTCTCCTAGAAATAG	62	374	1,93
	CG rich region	F:TCTGTGTCCCTCAAGTGCTTTC R:TGCAGCCCTTTCTCTGTG	66	263	1,86
Aj2c1	CG rich region	F:CCAGCCTCTCCGAATCTAA R:GCTAAATCATAAGCCGCAGAA	52	234	1,71
	Aj2c1	F:TTGCATCAGTTAACGTTTTCT R:ATATCAACCAACATTTAAGGGT	55	371	1,79
	Aj2c1 5'	F:TTGCATCAGTTAACGTTTTCT R:ATATCAACCAACATTTAAGGGT	55	172	1,96
	Aj2c1 3'	F:TTGCATCAGTTAACGTTTTCT R:ATATCAACCAACATTTAAGGGT	55	199	1,78
	CpGi	F:AACACTAGTGAACGAACT R:AGGAAGTTGTACCAGTCAA	54	153	1,8
	CpGi down I	F:TTGACTGGACAACCTTCT R:CCAAACTCGAGGATACAT	58	229	1,97
	CpGi down II	F:AGTCGCCAGAGATGTATC R:TGCCTCTAGCAGGAAAT	58	180	1,75
	Alu sx down I	F:TAATTCCTGCTAAGAGGCA R:TCTCTGACCACACAAAAATC	60	372	1,71
	Alu sx down II	F:CGGATTTTTGTGTGGTCAGA R:ACCCTAGCTAACATGGTGA	62	278	1,97
	Alu Jo down	F:ACTACTTTGGGCTGCTTTGT R:GAGCGGTGGCTCATATCT	65	350	1,85
AU60	Alu Sg	F:ATTATAGATGAGCACAATGTATAT R:TAACATTTAATAATAAAGCTGACC	58	370	1,96
	CpGi 5'	F:ATGAGGATCCGGGTGAGAGC R:ACCCTCTTGCCGCTGTATGG	67	242	1,87
	CpGi 3'	F:TCAGGTGAGGGTGGAAAACG R:AAGGACCACTGAGGGACAAG	64	252	1,83
	Alu sx	F:GGGACGGCCCGAGAGTTTT R:TCTTCTCCGTCCTATCTCAGA	58	370	1,92
	Au60	F:CTTGTGCAACATGGTCCTA R:CCTTAACTCAAGCACTGATGA	63	346	1,84
	Alu sg	F:GTCACCTCTTCTACTCT R:CTCAGGTTAAGAAAAGCAAC	58	358	1,96
	ZALU3	CpG up	F:CTGAGCAAGCCAGAATGA R:TTGCAAGAGGAGGAAGAC	65	139
CpGi 5'		F:CGCTCAACCAAGCACTCG R:CCCCTCTGTCTACTTTCATGG	64	134	1,9
Mir C		F:GCAGATGGCGGTCTTTTAT R:TATAGAGAGGGTGTCAAATCT	60	159	1,99
Alu Jb		F:ACCACATGTAACGCTATG R:AATATCCTTTGTAGTTTTAC	55	385	2,02
ZALU3		F:GCCACAGAACGAGACTCAG R:CGCCTGTAATGACGGATCACGA	64	376	2,04
Alu sx6		F:CAAGATGGCGAACTCAATC R:TTTTAGGCAGAGTCTTGCTC	60	195	2,26

**TAULA AI.11. ENCEBADORS EMPRATS PER GENERAR LES SONDES DEL NORTHERN BLOT**

La seqüència dels encebadors es mostra en direcció 5'-3'. F (encebador directe) i R (encebador invers). S'indica la temperatura d'amplificació segons l'encebador que estudiem i la mida del fragment amplificat.

ELEMENT ALU	GEN	REGIÓ (EXONS)	encebadors	Temperatura amplificació	mida del fragment (pb)
Aj2c1	DIEXF	3_4	F: GACAGTATTGTAGATGATGC R: CTGAACAGTGACTCGTGT	63	212
		6_7	F: AGACGCCGAAGCCAGAAGTT R: CCTCGAGGAGGCTGATGAA	63	142
		8_9	F: TCATTGGTGGAGAAGGAGAG R: ATGTGAGTCCAGGGGTAGT	63	155
		11_12a	CCTTCAAGGAGAGAAACAGT GGTGGCTCTCAGCATATTAC	63	151
		12.a	F: GCAGGAAGTGGTATTGGCA R: GGCTGGGGCAGAATAAGTTT	63	478
		12	F:TCCTGTTCCCTTCTGTGAG R:CCTGGAGCTCTGTCTTCT	63	169
		UTR002	F: GTCATTGTGAGAGGAGGAA R: GGGCTACAATTAACCTTGCTG	61	185

**TAULA AI.12. ENCEBADORS EMPRATS PER L'ASSAIG RACE**

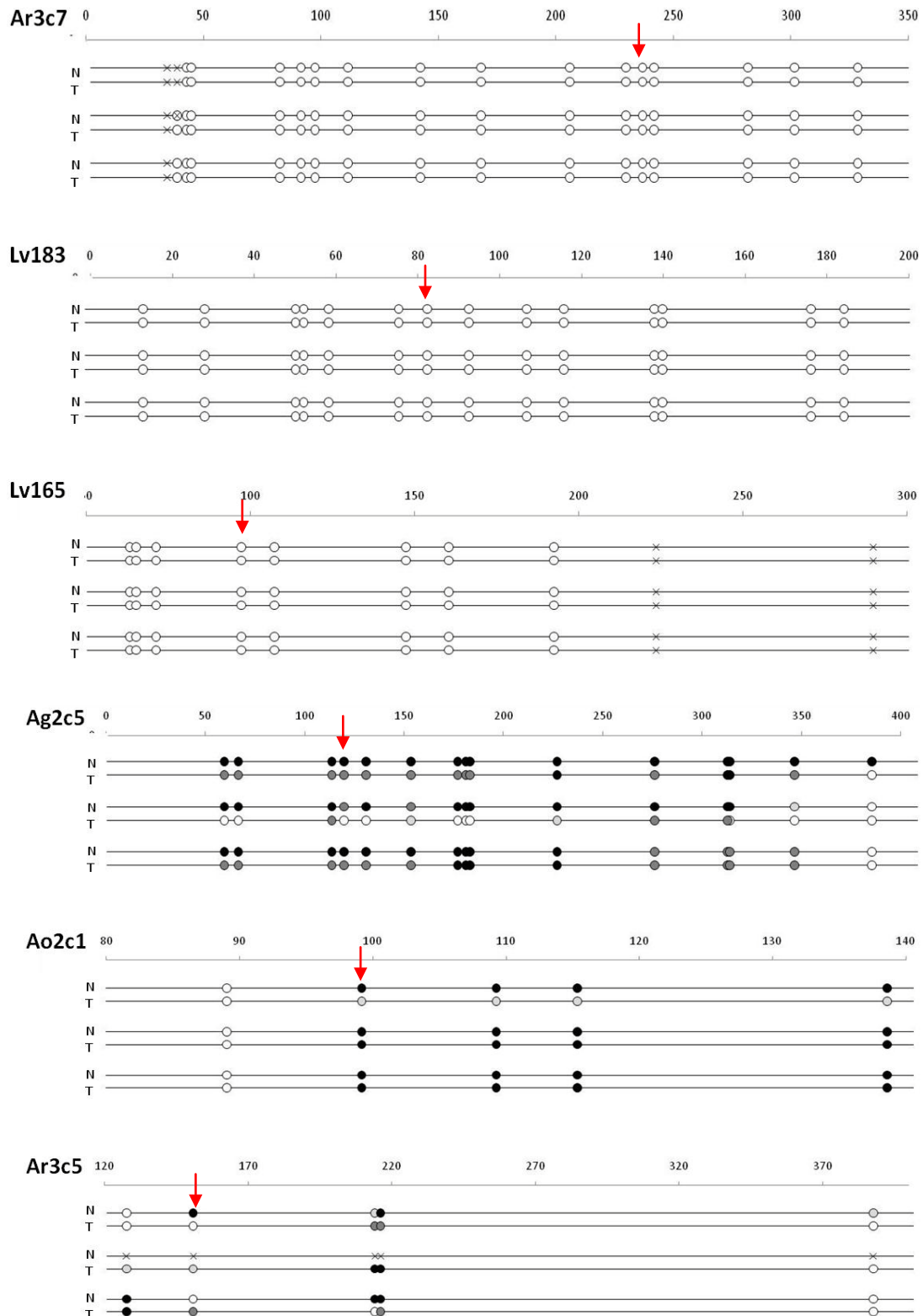
La seqüència dels encebadors es mostra en direcció 5'-3'. S'indica la temperatura d'amplificació segons l'encebador que estudiem.

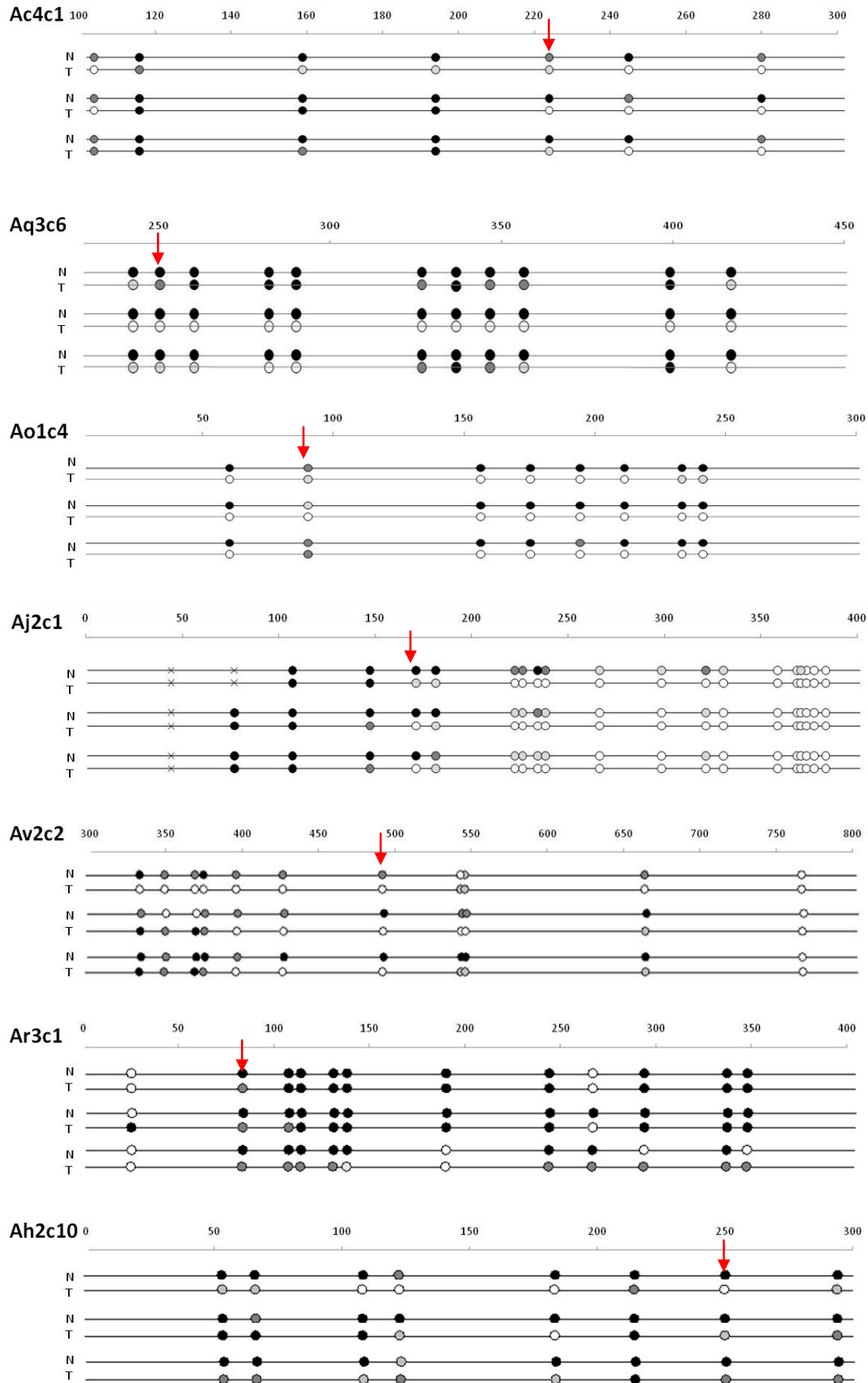
	REGIÓ	encebadors	Temperatura amplificació
RACE 5'	UPM mix	UPM long UPM short	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT CTAATACGACTCACTATAGGGC
	9R	C1orf107_GSP1-9R	GAGTTGATCTGGGCATCCTGAAGGGC
	UPM mix	UPM long UPM short	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT CTAATACGACTCACTATAGGGC
	7R	C1orf107_GSP1-7R	GAAGAGCTGCACCACCCGCAAAGCA
	UPM mix	UPM long UPM short	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT CTAATACGACTCACTATAGGGC
	3R	C1orf107_GSP1-3R	CCACACTGACATCGCTACCACCATC
RACE 3'	UPM mix	UPM long UPM short	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT CTAATACGACTCACTATAGGGC
	8F	C1orf107_GSP1-8F	CTGATGCAGAAGTGGGAGCATGTCTCTG
	UPM mix	UPM long UPM short	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT CTAATACGACTCACTATAGGGC
	12F	C1orf107_GSP1-12F	TGGTTGGTGTGGAGCGGGCGGCA
	UPM mix	UPM long UPM short	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT CTAATACGACTCACTATAGGGC
	12F2	C1orf107_GSP1-12F2	GCTCTCTGAAGGCTGATGACCACCTG
	UPM mix	UPM long UPM short	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT CTAATACGACTCACTATAGGGC
	12F3	C1orf107_GSP1-12F3	GGTCCTAGGCTCCCTTACACCTCT
	UPM mix	UPM long UPM short	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT CTAATACGACTCACTATAGGGC
	12F4	C1orf107_GSP1-12F4	GGCCACGTGGACCTGCACTGTTCTC

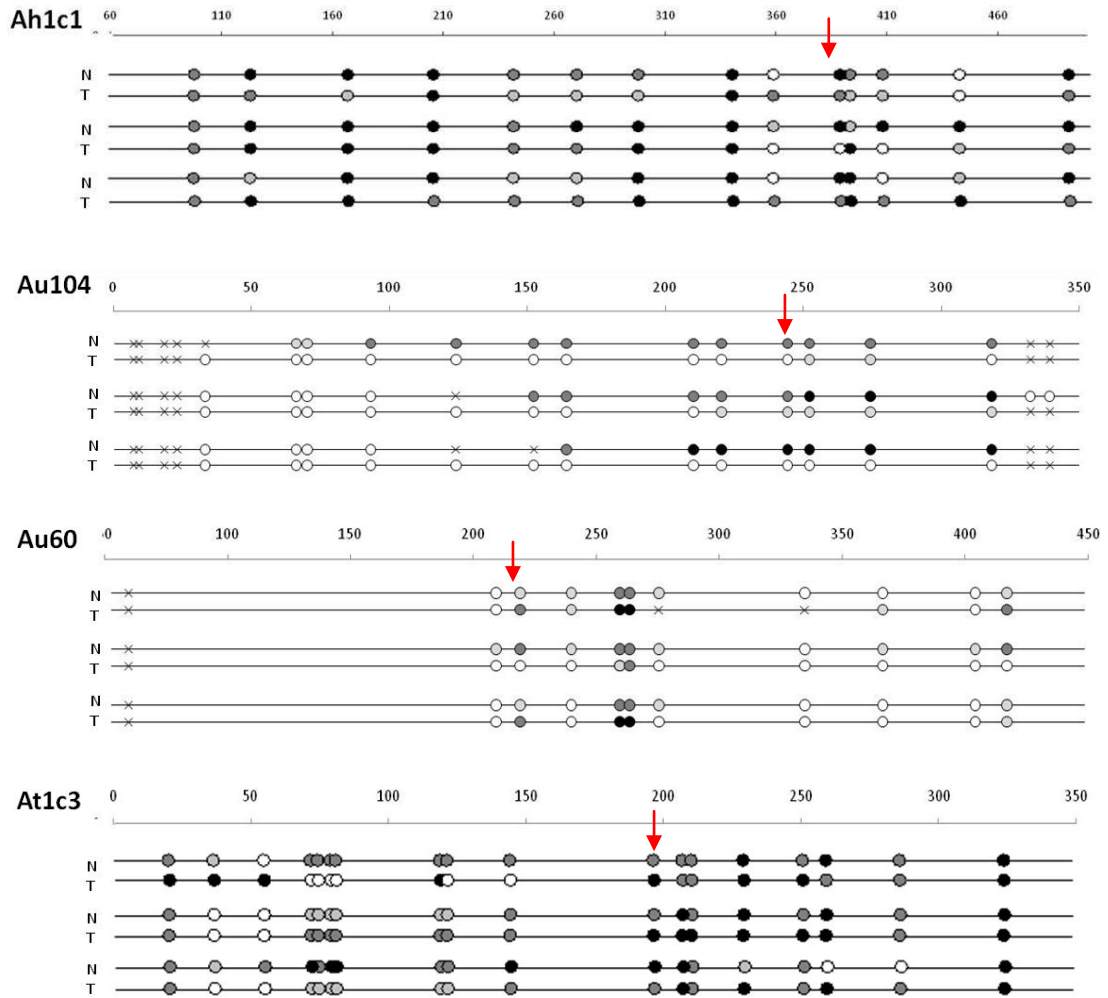
## ESTAT DE METILACIÓ

Marquem amb una fletxa vermella la posició de la CpG corresponent a la diana Smal.

**Figura AII.1. Estat de la metilació en els elements ALU en mostres mucosa normal de colon (N) i el tumor colorectal aparellat (T)**

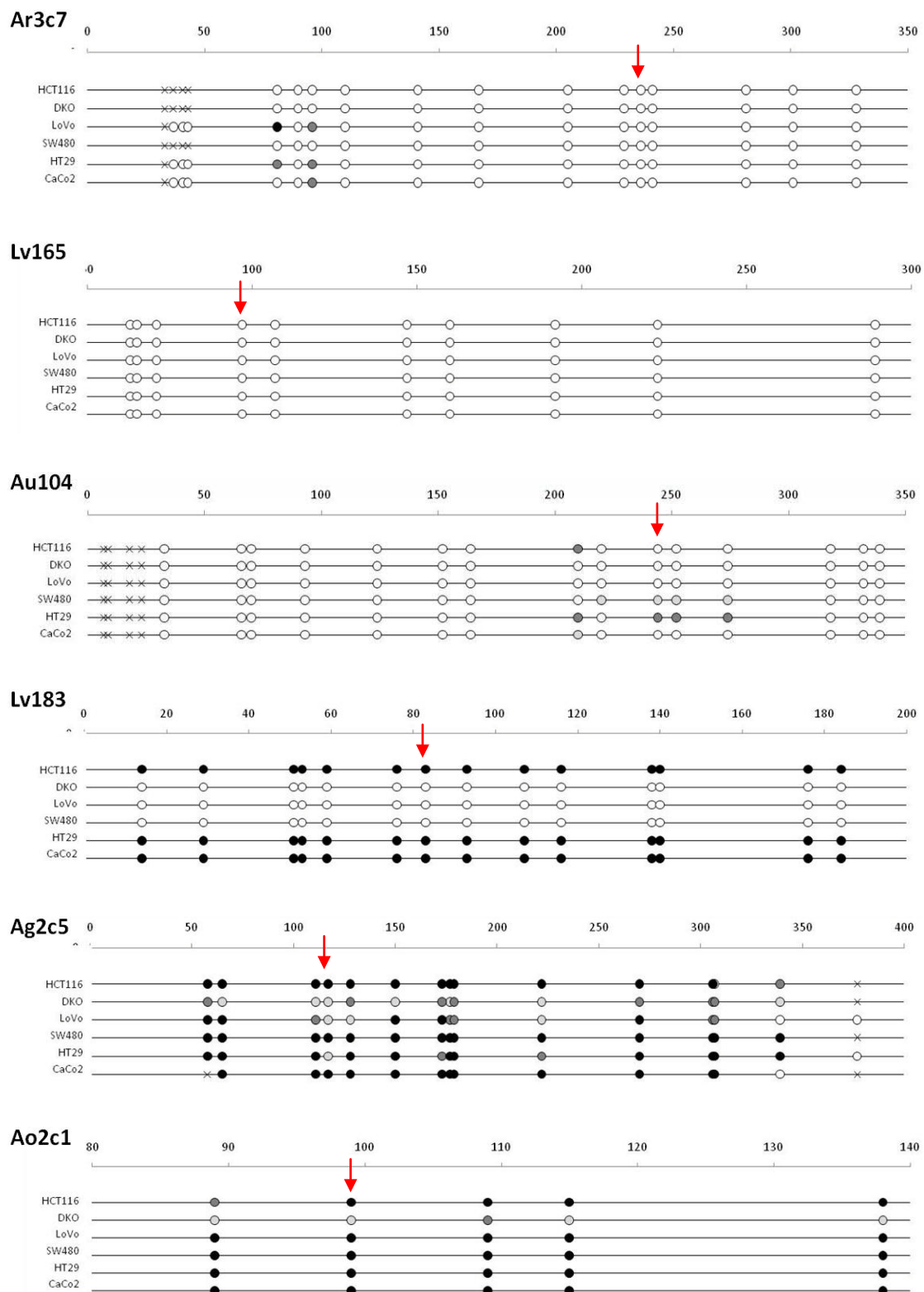


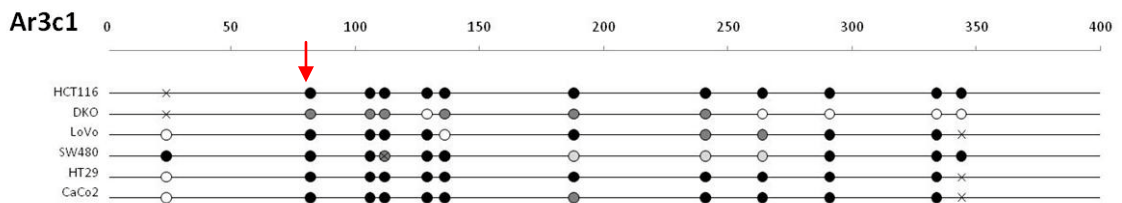
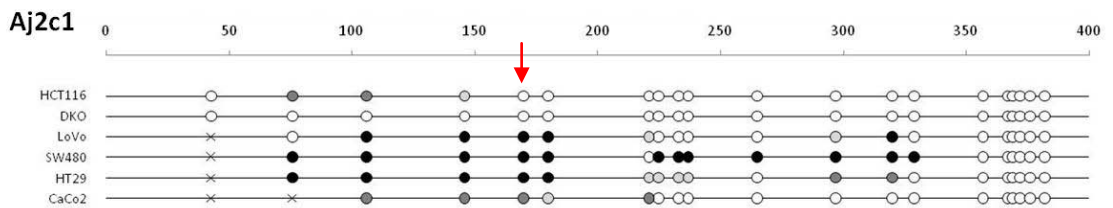
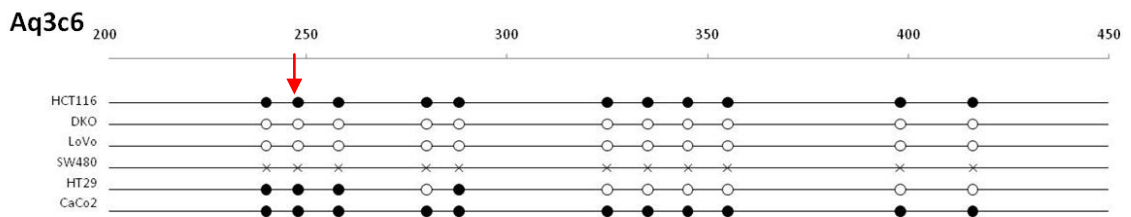
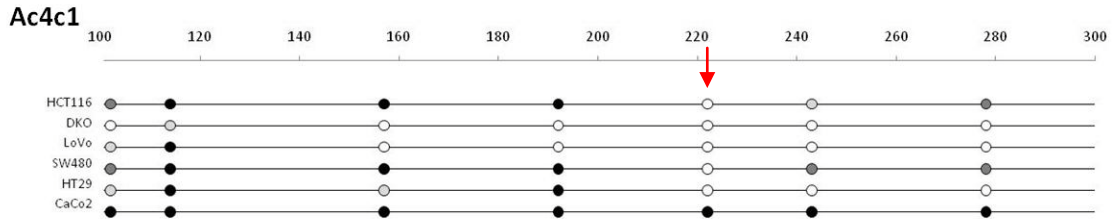
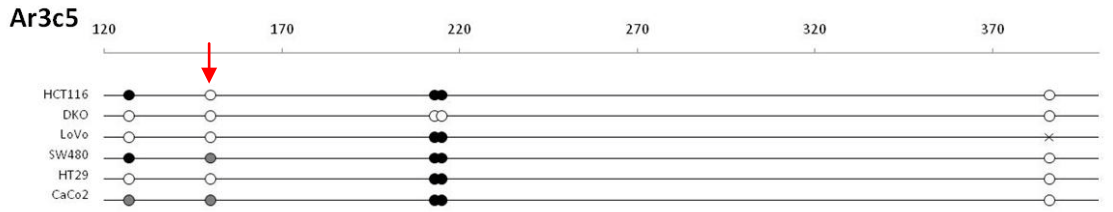




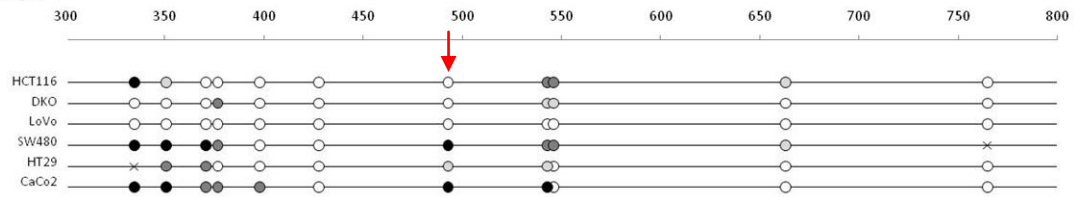
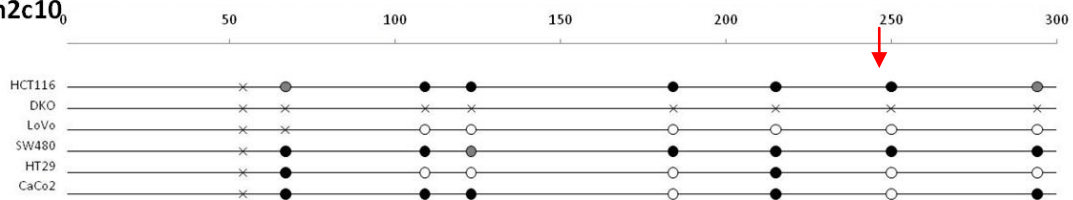
X no determinat, ○ 0% metilat, ◐ 25-50% metilat, ◑ 50-75% metilat, ● 100% metilat

**Figura AII.2. Estat de la metilació en els elements ALU en línies cel·lulars de càncer de còlon**

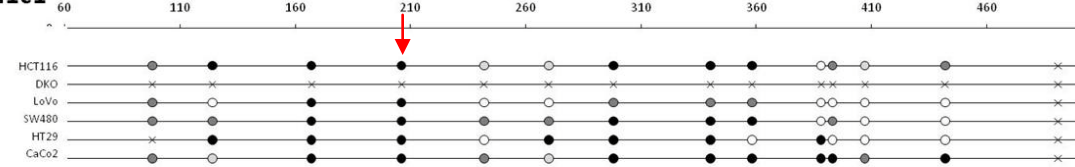




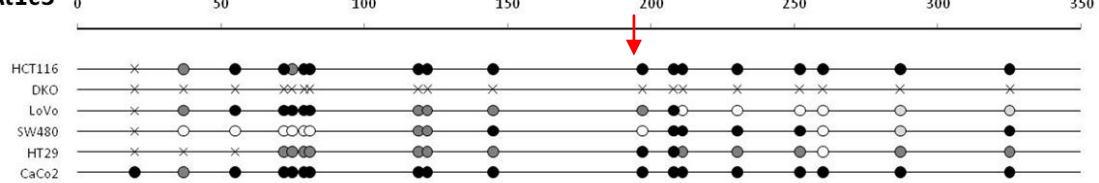
## Av2c2

Ah2c10<sub>0</sub>

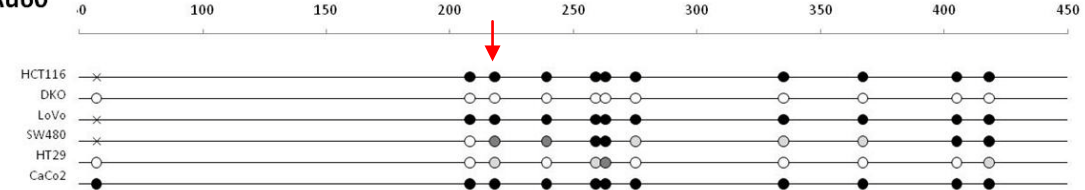
## Ah1c1



## At1c3



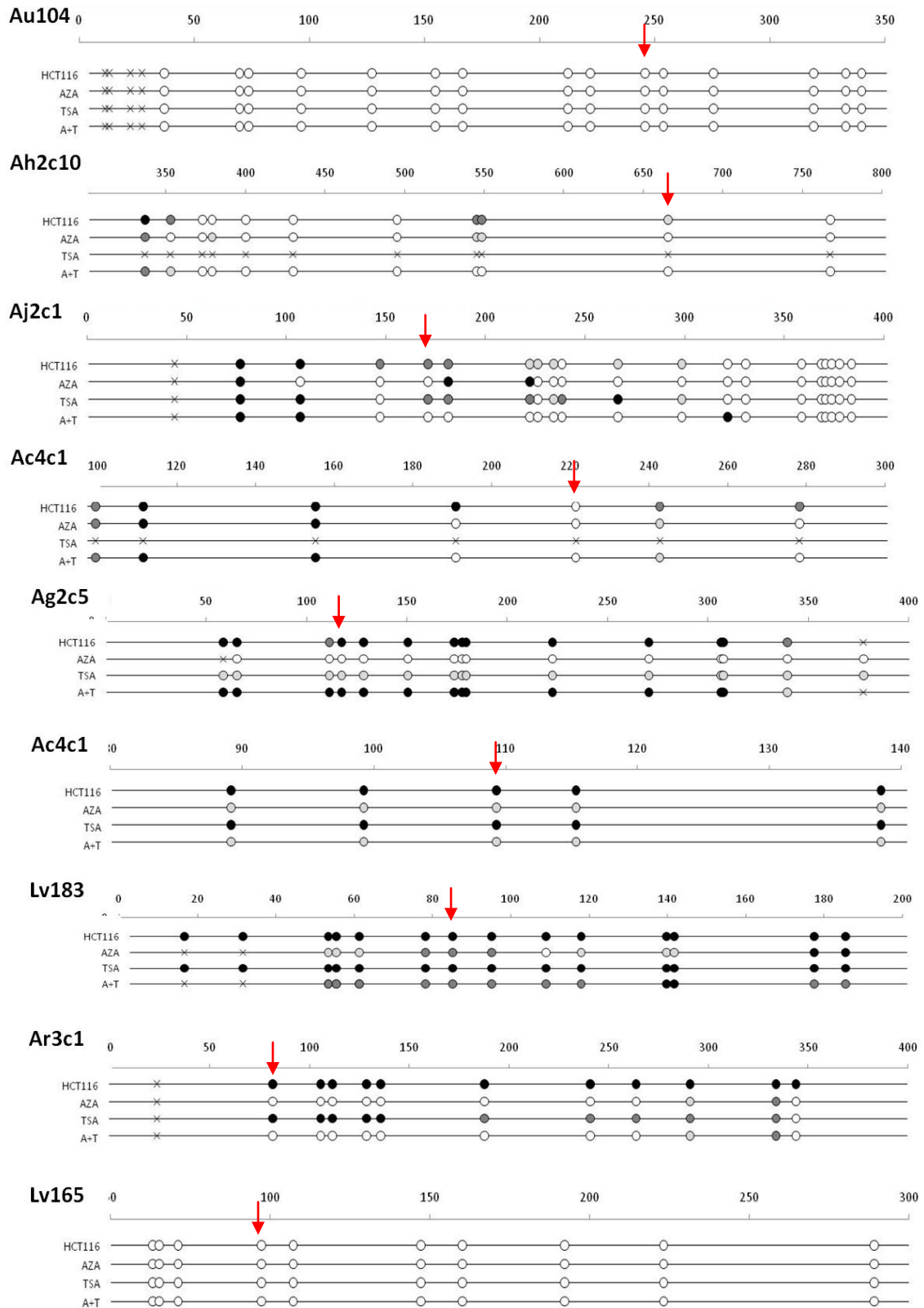
## Au60

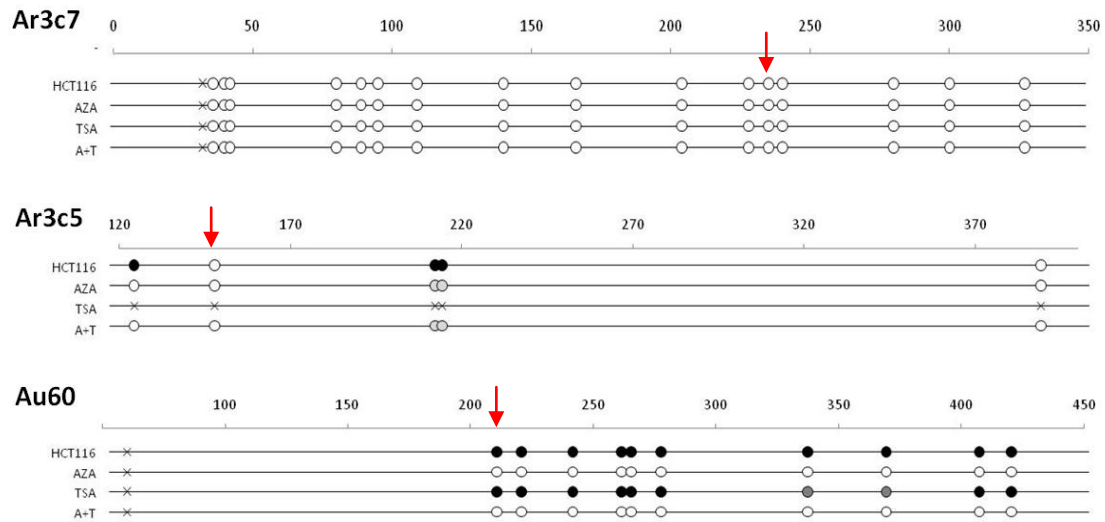


X no determinat, O 0% metilat, O 25-50% metilat, O 50-75% metilat, O 100% metilat



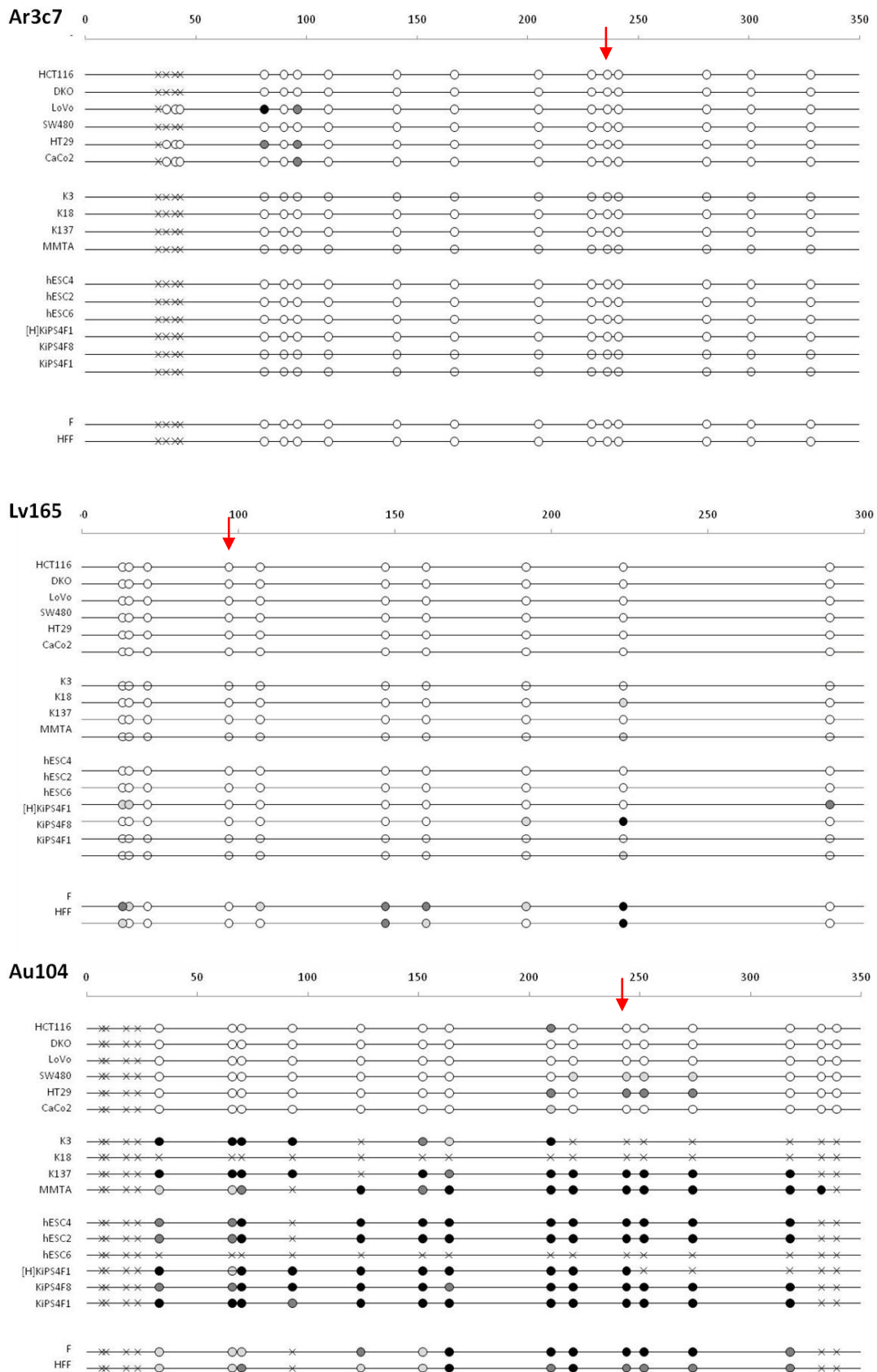
**Figura AII.3. Estat de la metilació en els elements ALU en HCT116 vs HCT116 tractat amb 5-azaC, Tsa, A+T**

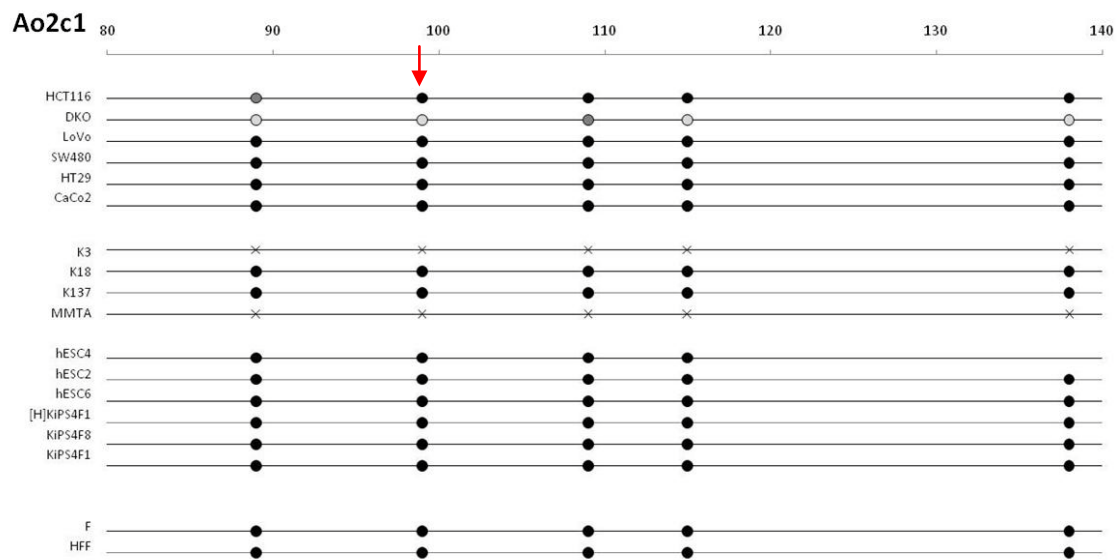
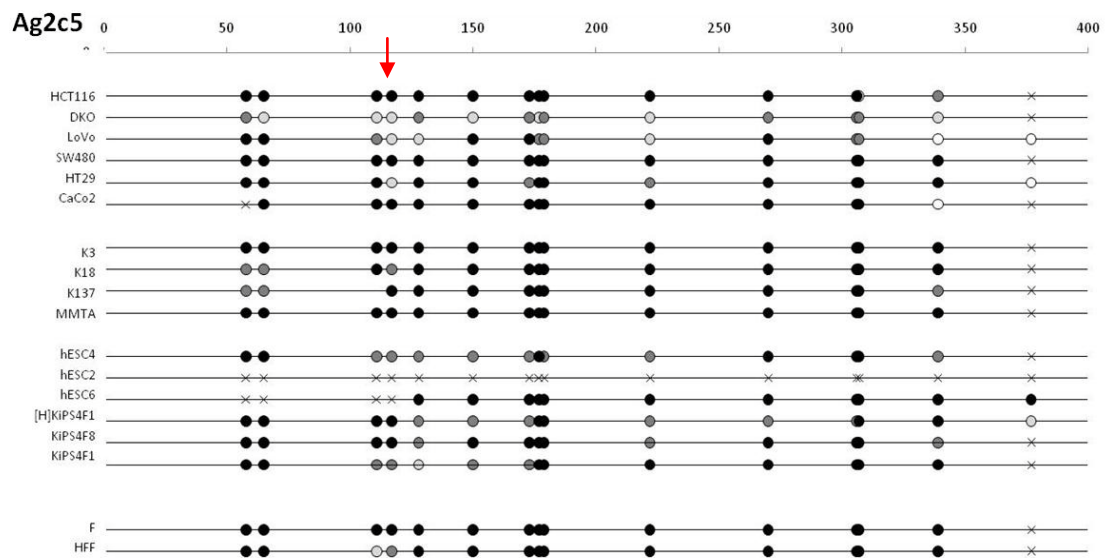
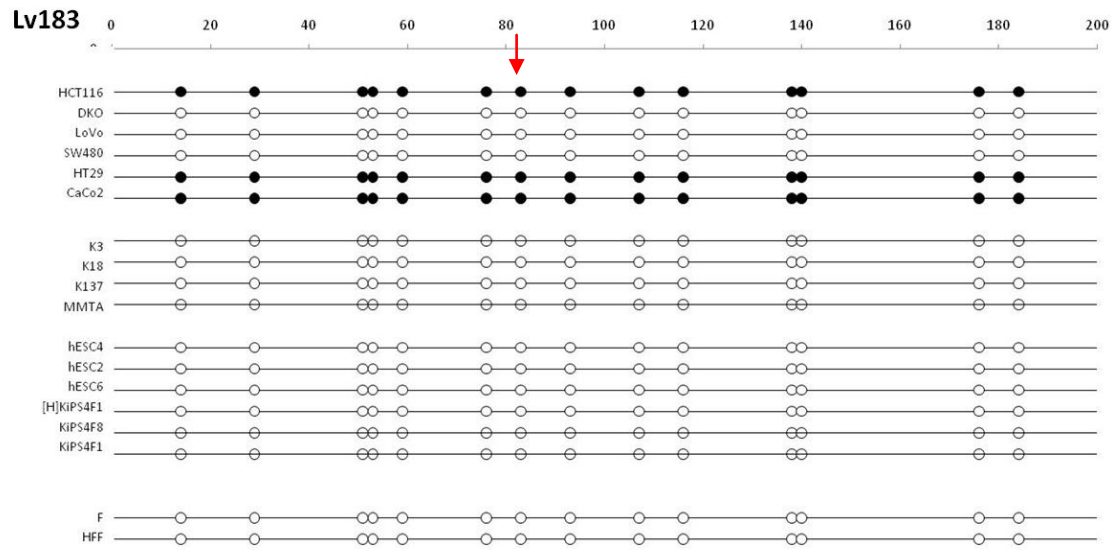




X no determinat, ○ 0% metilat, ○ 25-50% metilat, ● 50-75% metilat, ● 100% metilat

**Figura AII.4. Estat de la metilació en els elements ALU en diferents línies cel·lulars**





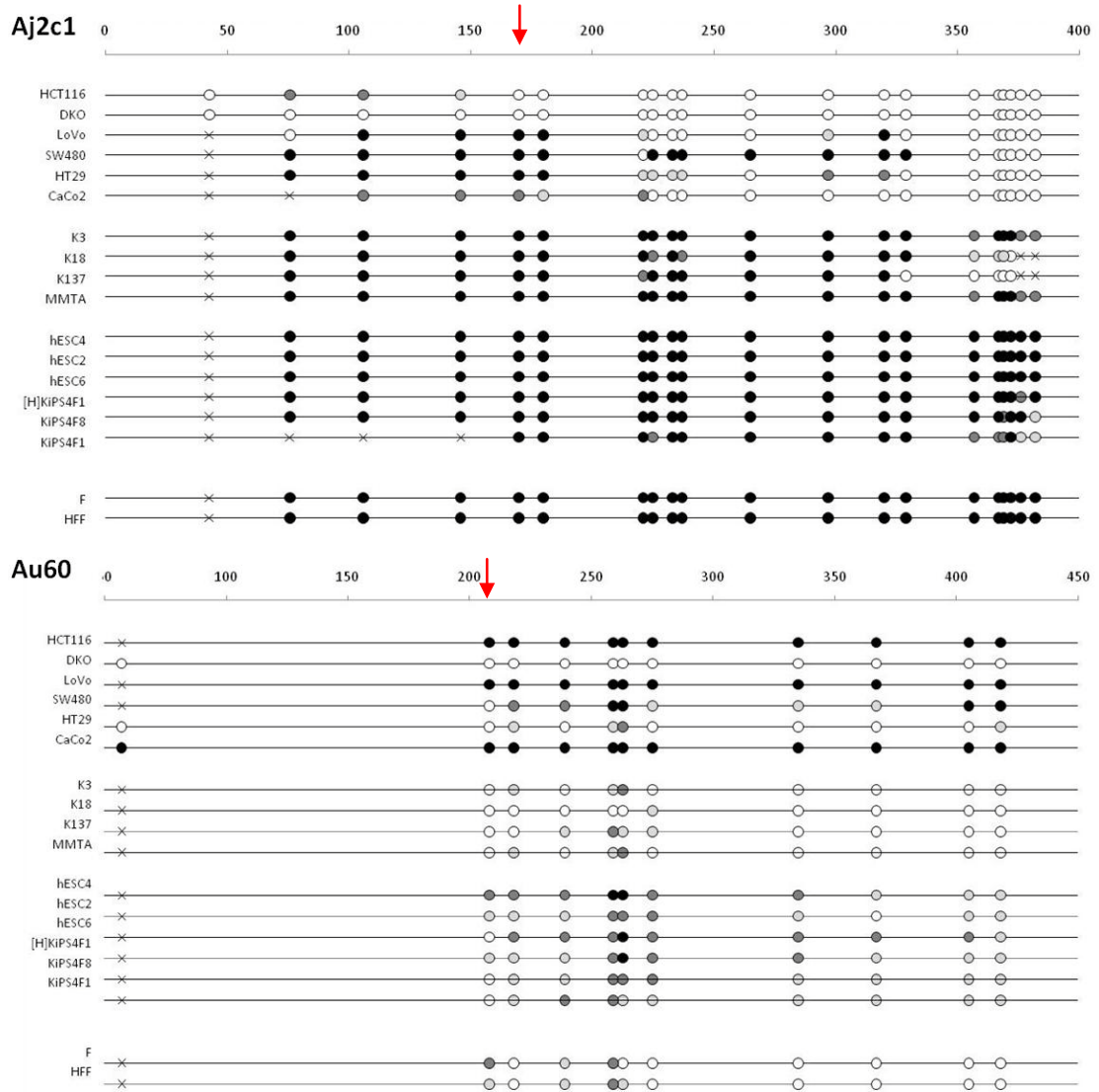


Figura AII.5. Estat de la metilació en l'element Alu ZALU3

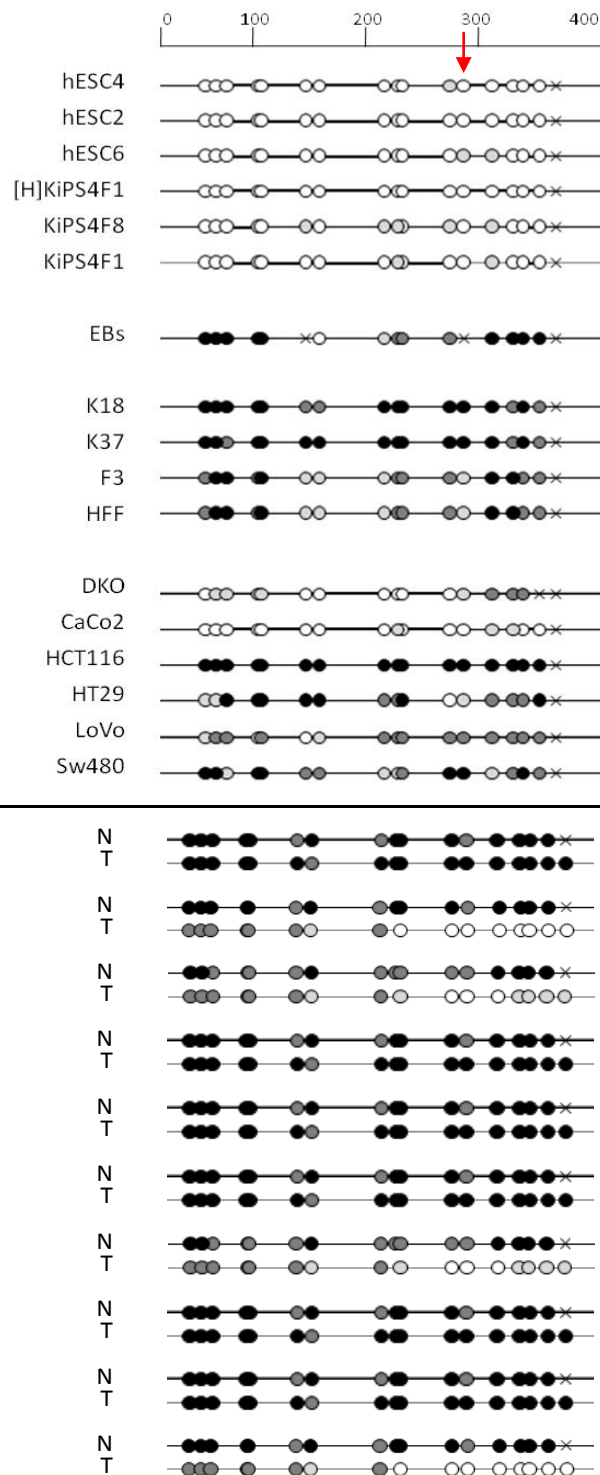
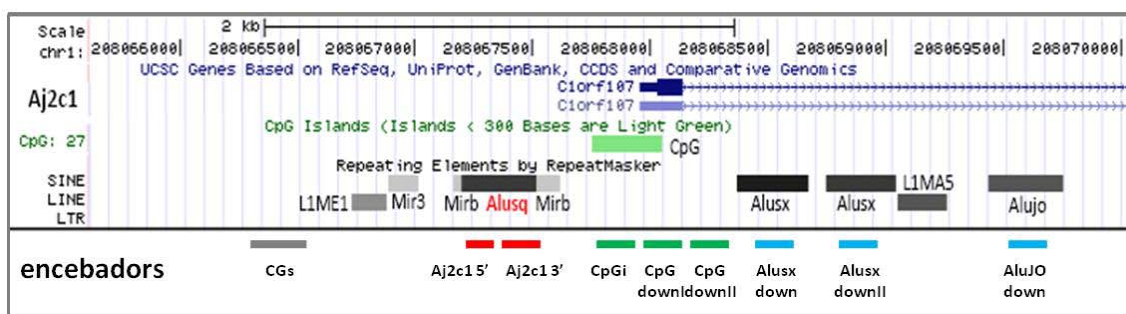


Figura AIII.1. Posició dels encebadors per CHIP de la regió promotora de *DIEXF*Figura AIII.2. Posició dels encebadors per CHIP de la regió promotora de *GLDC*