

Identificació i caracterització de nous factors associats als elements insulator de *Drosophila melanogaster*

Sergi Cuartero i Betriu

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Departament de Genètica

Facultat de Biologia

Universitat de Barcelona

TESI DOCTORAL

**Identificació i caracterització de nous factors associats
als elements insulator de *Drosophila melanogaster***

Sergi Cuartero i Betriu

Desembre del 2012

Programa de Doctorat de Genètica

Facultat de Biologia

Universitat de Barcelona

**Identificació i caracterització de nous factors associats als
elements insulator de *Drosophila melanogaster***

Memòria presentada per Sergi Cuartero i Betriu per obtenir el grau
de Doctor per la Universitat de Barcelona.

Treball realitzat sota la direcció de la Doctora Maria Lluïsa Espinàs
Janer a l'Institut de Biologia Molecular de Barcelona (IBMB-CSIC) i
l'Institut de Recerca Biomèdica de Barcelona (IRB).

La directora

La tutora

Dra M^aLluïsa Espinàs Janer

Dra Montserrat Corominas Guiu

Agraïments

Aquesta tesi no hagués existit mai, probablement jo no aspiraria a ser doctor, i —sens dubte— jo no hagués après ni una milèsima part de les coses que he après en els últims anys de no ser per què la Lluïsa me n'ha donat la oportunitat. L'agraïment és doncs, per sobre de tot, per a ella. Pel dia que va decidir que jo podia formar part del seu grup, i per tots els altres dies que ha treballat dur perquè la nostra feina tirés endavant. Perquè ha sabut combinar flexibilitat i exigència perfectament, perquè ha sabut corregir-me quan ho he necessitat i ha sabut ser comprensiva i oberta amb les meves propostes, incloent les més arriscades i que potser no han portat enlloc, però que m'han mantingut motivat i despert. Perquè s'ha arremangat sempre que ha estat necessari per ajudar-me en tot tipus d'experiments, i sóc conscient que això no és molt comú. Si hagués de tornar a fer una tesi, tornaria a fer-la amb la Lluïsa.

També he d'agrair al Ferran i al Jordi Bernués que sempre que ho he necessitat m'han ajudat. El Ferran sempre ha tingut la seva porta oberta per donar-me consells i el Jordi m'ha ajudat en tot: preguntes, cèl·lules, mosques, anticossos, plàsmids... No recordo moment en que, malgrat la seva falta de temps crònica, no m'hagi escoltat i parat tot per ajudar-me. Al Josep Portugal, que també m'ha donat bons consells sempre que li he demanat, li agraeixo el bon humor, i com no, que m'hagi deixat utilitzar el seu bany durant tots aquests anys tot i la meva insistència en oblidar-me d'anotar el meu nom al post-it i totes les vegades que me l'he deixat a 42°C i educadament m'ha recordat que un dia me'l carregaré.

Per fer tota la feina feta en aquesta tesi he necessitat l'ajuda de molta gent, que m'han ensenyat a treballar i s'han interessat per la feina. He de començar per agrair infinitament a l'Àlicia per multitud d'experiments i l'Ester Fuentes i l'Estefanía amb les mosques, els stocks i els politènics. A la Ujué per haver participat de la feina i haver-ho fet tant bé, sempre predisposada a

aprendre més i disposada a ajudar el que calgui, sempre amb el bon humor que la caracteritza. No tinc cap dubte que la teva tesi t'anirà molt bé. També li he d'agrair a la Sílvia Pérez, de qui vaig heredar el projecte, que m'ensenyés tot el que va poder en el poc temps que vam coincidir. Als que durant molt de temps han estat els meus companys d'U i amics, el Joan i el Romanx. El Joan sempre s'ha interessat pel meu projecte, ha proposat experiments i m'ha ensenyat a fer-ne, m'ha estimulat i m'ha motivat. No hi ha dubte que és una peça clau d'aquesta tesi. Amb el Romanx he après molt, hem comentat sempre els experiments i m'ha ajudat tant com ha pogut. Ha estat el meu company de taula, company d'experiments, professor d'alemany, i sobretot el meu amic. A tots dos gràcies pels bons moments que hem passat, per tots els vespres amb la música forta, pels nostres intents fallits de compartir menjar, pels experiments fins les tantes, ha sigut realment una gran època.

He de donar-li gràcies a tota aquella gent que d'una manera o altra han col·laborat amb la tesi com l'Oscar Reina i l'Evarist Planet amb els anàlisis bioinformàtics, el Herbert Auer, el David Fernández i l'Annie Rodolosse pel microarray i el CHIP-seq, el Francesc Canals per l'espectrometria de masses, l'Anna per desxifrar els dominis de les proteïnes, la Lídia Bardia per ensenyar-me a fer servir el confocal, l'Ainoa Olza per punxar els transgènics. A tota la gent de mosques que m'han ajudat i m'han ensenyat moltes coses: el Jordi Casanova, l'Andreu Casali, la Lara, el Carles, l'Héctor, l'Andrés, la Lídia, l'Ana i sobretot la Laura.

Li he d'agrair a tota la gent del laboratori, la que hi és i la que ja no, que hagin creat tant bon ambient. Començant per l'Olga, que m'ha resolt dos milions de dubtes i que si mai estic sol al lab i no sé com es fa alguna cosa, m'ha agafat el telèfon encara que fos un diumenge el vespre i com sempre, sap la resposta. La Mònica, la Sònia i l'Olga feu un molt bon equip i la vostra U sempre és la més alegre. Al Carles i l'Aleix els hi he d'agrair per la seva existència, jo no aguantaria una feina en la que no pogués comentar el futbol i la política cada dia. M'ho he passat molt bé amb vosaltres. Al Salva i al Tomás gràcies per l'alegria i pels bons

moments cantant Los caminos de la vida! I gràcies també tots els demés Marta Lloret, Marta Batlle, Marta Blanch, David, Martí, Pablito, Albert Vergoñós (que és de tot menys vergonyós), Marc Bataller, Bet, Sylvia Mansilla, Carolina, Anne, Eva, Johan, Milos, Joan Ponsà, Lucía, Albert i a tots els que hagueu estat al laboratori i que ens haguem creuat encara que sigui fugaçment.

He d'agrair també al Maxime i la Brigitte, els meus directors de projecte de Master, que em donguessin l'oportunitat d'entrar a la ciència i que confiessin en mi. I, com no, li he de donar les gràcies a tota la gent que m'ha rodejat tot aquest temps. Gràcies principalment per existir i per ser amics meus. Pexi, Pere, Pol, Joan, Lluís, Aleix, Xavi, Geri, Carles, Dani, Albert, Manex, Keko, Manolo, Leire, Pablo, Franc, Alberto, Markeletxe, Deivid. Sé que si algú llegeix això pensarà que exagero, però no. Realment sóc un afortunat i us tinc a tots vosaltres. Per sobre de tot però, li he d'agrair a l'Andrea tot el que hem passat junts i tot el que vindrà. Als meus pares i al meu germà (i a la Leire!), que m'heu donat suport incondicional en tot i fins a l'últim moment, gràcies.

A l'avi Joan, a l'avi David, a l'àvia Concepció

A l'àvia Carmen, companya en aquest viatge

ÍNDEX

1. INTRODUCCIÓ	1
1.1. L'emmagatzematge de la informació genètica al nucli cel·lular	3
1.1.1. La cromatina	3
1.1.1.1. Estructures de la cromatina	5
1.1.1.2. Remodelació de la cromatina	7
1.1.1.3. Variants d'histones	9
1.1.1.4. Modificacions post-traduccionals de les histones	11
1.2. La regulació de l'expressió gènica	15
1.2.1. La transcripció	17
1.2.1.1. La RNA Polimerasa	18
1.2.1.2. Els factors generals de transcripció	18
1.2.2. Els elements que regulen la transcripció	19
1.2.2.1. El promotor	19
1.2.2.1.1. Promotors pausats	21
1.2.2.1.2. Promotors alternatius	23
1.2.2.2. Els enhancers	25
1.2.2.3. Polycomb response elements	27
1.2.3. Els insulators	32
1.2.3.1. Insulators a Drosophila	34
1.2.3.1.1. Scs i scs'	34
1.2.3.1.2. Gypsy	35
1.2.3.1.3. Els insulators del Complex Bitòrax	36
1.2.3.1.3.1. Fab-7	39
1.2.3.1.3.2. Fab-8	40
1.2.3.1.3.3. Mcp	41
1.2.3.1.3.4. Organització i funcionament de la regió reguladora d'Abdominal-B	41
1.2.3.2. Proteïnes d'insulator	44
1.2.3.2.1. Su(Hw)	44

1.2.3.2.2. BEAF	45
1.2.3.2.3. CTCF	46
1.2.3.2.4. CP190	48
1.2.3.2.5. Models d'interacció entre els diferents tipus d'insulator	52
1.2.3.3. Els insulators des d'una perspectiva evolutiva	53
1.2.3.3.1. CTCF i els insulators a vertebrats	53
1.2.3.3.2. Els insulators com a organitzadors de la cromatina	56
2. OBJECTIUS	59
3. MATERIALS I MÈTODES	63
3.1. Materials	65
3.1.1. Primers	65
3.1.2. Plàsmids	66
3.1.3. Anticossos	67
3.1.4. Línies de Drosophila	68
3.1.5. Soques bacterianes	69
3.1.6. Línies cel·lulars	69
3.2. Mètodes	70
3.2.1. Cultius cel·lulars	70
3.2.2. Tècniques de treball amb DNA	71
3.2.2.1. PCR	71
3.2.2.2. Transformació de cèl·lules competents	72
3.2.2.3. Minipreparacions de DNA plasmídic	72
3.2.2.4. Maxipreparacions de DNA plasmídic	73
3.2.2.5. Extracció de DNA genòmic	73
3.2.3. Tècniques de treball amb RNA	73
3.2.3.1. Extracció d'RNA	73
3.2.3.2. RT-PCR	74
3.2.3.3. Síntesi de RNA de doble cadena	75

l'activitat insulator	
4.2.1. El complex proteic de CP190	108
4.2.1.1. Expressió induïda de la proteïna en cultiu cel·lular	108
4.2.1.2. Purificació del complex	110
4.2.1.3. Identificació dels components del complex	111
4.2.2. Anàlisi de les proteïnes identificades	112
4.2.2.1. Proteïnes d'insulator	112
4.2.2.2. NURF-38	116
4.2.2.3. Z4	117
4.2.2.3.1. Z4 colocalitza parcialment amb CP190 a	118
politènics	
4.2.2.3.2. Z4 afecta a l'activitat insulator	119
4.2.3. Caracterització i anàlisi funcional de lbf-1 i lbf-2	120
4.2.3.1. Interacció amb CP190	120
4.2.3.2. lbf-1 i lbf-2 són gens adjacents i proteïnes similars	121
4.2.3.3. lbf-1 i lbf-2 estan conservades dins del gènere	123
Drosophila	
4.2.3.4. lbf-1 i lbf-2 tenen un domini d'unió al DNA	124
4.2.3.5. Generació d'un mutant de lbf-1	126
4.2.3.6. Generació d'un mutant de lbf-2	127
4.2.3.7. Localització a cèl·lules S2	129
4.2.3.8. Localització a embrions	130
4.2.3.9. Unió a cromosomes politènics	132
4.2.3.10. Presència d'insulator bodies	138
4.2.3.11. Localització en cromosomes mitòtics	140
4.2.3.12. Alteracions en l'expressió gènica	141
4.2.3.12.1. RNAi a S2	142
4.2.3.12.2. Canvis d'expressió gènica	143
4.2.3.12.3. Co-regulació	145
4.2.3.13. Anàlisi exhaustiu de la localització de lbf-1, lbf-2	150

i CP190 al genoma	
4.2.3.13.1. Localització al genoma segons el tipus de regió	152
4.2.3.13.2. Posicionament sobre els gens	154
4.2.3.13.3. Ibf-1 i Ibf-2 colocalitzen quasi totalment entre elles i amb CP190	155
4.2.3.13.4. Ibf-1 i Ibf-2 colocalitzen parcialment amb les demés proteïnes d'insulator	156
4.2.3.13.5. Localització respecte els promotors	161
4.2.3.13.6. Ibf-1 i Ibf-2 s'uneixen a una nova seqüència consens	163
4.2.3.14. Activitat enhancer-blocking d'Ibf-1 i Ibf-2	169
5. DISCUSSIÓ	171
5.1. La regió reguladora d'Abdominal-B	173
5.1.1. El Fab-6 i l'iab6PRE	173
5.1.2. Transcripció als elements reguladors d'Abdominal-B	177
5.1.3. JIL-1 als insulators	179
5.2. Noves proteïnes d'insulator	183
5.2.1. El complex de CP190	183
5.2.2. Caracterització d'Ibf-1 i Ibf-2	186
5.2.2.1. Conservació i dominis	186
5.2.2.2. Localització del-lular d'Ibf-1 i Ibf-2	189
5.2.2.3. Canvis d'expressió gènica	192
5.2.2.4. Localització al genoma d'Ibf-1, Ibf-1 i CP190	195
6. CONCLUSIONS	203
7. BIBLIOGRAFIA	207

Índex de figures

Figura 1.- L'empaquetament de la cromatina	5
Figura 2.- Els cromosomes politènics	6
Figura 3.- Models de funcionament dels remodeladors de cromatina	8
Figura 4.- Histones canòniques i no-canòniques	10
Figura 5.- Modificacions covalents de les histones	12
Figura 6.- Elements del <i>core promoter</i>	21
Figura 7.- Model de control de la pausació de la polimerasa	22
Figura 8.- Exemple de gen amb promotors alternatius.	24
Figura 9.- Memòria epigenètica dels PRE/TRE al llarg del cicle cel·lular	31
Figura 10.- Les dues propietats definitòries dels insulators	33
Figura 11.- El Complex Bitòrax i els segments que controla	37
Figura 12.- El model dels Dominis	43
Figura 13.- Localització al genoma de les proteïnes d'insulator en relació als 5' i 3' dels gens anotats	50
Figura 14.- La hipòtesi de les tres sub-classes d'insulator	51
Figura 15.- Regulació del locus Igf2/H19	55
Figura 16.- Modelització del funcionament dels insulators	57
Figura 17.- Els llocs hipersensibles a DNAsa1 a l'iab-6	92
Figura 18.- Immunoprecipitacions de cromatina a la regió reguladora d'Abd-B	95
Figura 19.- Assaig d'enhancer-blocking en constructes transgènics amb el gen reporter mini-white	97
Figura 20.- La mutació de CP190 afecta a l'HS1 però no a l'HS2	98
Figura 21.- La mutació de CTCF afecta a l'HS1	99
Figura 22.- La mutació de JIL-1 afecta el funcionament de Fab-7 i HS1	102
Figura 23.- Immunoprecipitacions de cromatina d'H3S10p	103
Figura 24.- Transcripció al Fab-6 i Fab-7 a larva	105
Figura 25.- Transcripció al Fab-6 i Fab-7 a embrions	105
Figura 26.- Transcripció d'Abd-B i Fab-7 a cèl·lules en cultiu S2 i SG4	106

Figura 27.- Transcripció a larves wild-type i larves mutants de JIL-1 i CTCF	107
Figura 28.- Constructe per expressar CP190-TAP a cèl·lules	109
Figura 29.- Expressió de la proteïna de fusió CP190-TAP a cèl·lules	110
Figura 30.- Purificació del complex de CP190	111
Figura 31.- Co-immunoprecipitacions de Su(Hw) i BEAF	116
Figura 32.- Co-immunoprecipitació de NURF-38 amb CP190	117
Figura 33.- Co-immunoprecipitació de Z4 amb CP190	117
Figura 34.- Immunotincions de cromosomes politènics de larva wild-type i larva mutant de Chromator	119
Figura 35.- Interaccions entre Ibf-1, Ibf-2 i CP190	121
Figura 36.- Els gens CG8436 (Ibf-1) i CG9740 (Ibf-2) al genoma	121
Figura 37.- Similitud entre Ibf-1 i Ibf-2.	122
Figura 38.- Alineament de Ibf-1 i Ibf-2 amb els seus homòlegs a <i>Drosophila</i>	124
Figura 39.- BED finger	125
Figura 40.- Conservació del BED finger	126
Figura 41.- Nivell d'Ibf-1 al mutant Δ Ibf-1 en heterozigosi	127
Figura 42.- Immunolocalització de Ibf-1 i Ibf-2 a cèl·lules S2	129
Figura 43.- Immunolocalització de Ibf-1, Ibf-2 i CP190 a cèl·lules impactades	130
Figura 44.- Immunolocalització de Ibf-1, Ibf-2 i CP190 als embrions	132
Figura 45.- Immunotinció d'Ibf-1 i CP190 de cromosomes politènics	133
Figura 46.- Immunotinció d'Ibf-2 i CP190 a cromosomes politènics	134
Figura 47.- Immunotinció d'Ibf-1, Ibf-2 i CP190 a cromosomes politènics de larves mutants de CP190.	136
Figura 48.- Immunotinció d'Ibf-1 i CP190 a cromosomes politènics de larva mutant d'Ibf-2	137
Figura 49.- Immunotinció d'Ibf-2 a cromosomes politènics de larva mutant d'Ibf-1	137
Figura 50.- Insulator bodies a nuclis de cèl·lules de cervell	139

Figura 51.- Localització d'Ibf-1 i Ibf-2 a cromosomes mitòtics	140
Figura 52.- Baixada del nivell de proteïna en resposta a RNAi.	143
Figura 53.- Correlacions d'Spearman entre els tres knockdowns	146
Figura 54.- Valors simulats i observats de co-regulació	148
Figura 55.- Exemples de pics al genoma	151
Figura 56.- Localització segons la regió genòmica	153
Figura 57.- Localització dels pics sobre els gens	154
Figura 58.- Diagrama de Venn de les tres proteïnes del complex	155
Figura 59.- Solapaments de les Ibf-1, Ibf-2 i CP190 amb les altres proteïnes d'insulator	158
Figura 60.- Quadre de combinacions de proteïnes d'insulator.	159
Figura 61.- Posicionament d'Ibf-1, Ibf-2 i CP190 respecte altres proteïnes de la cromatina	160
Figura 62.- Localització sobre promotors	162
Figura 63.- Seqüències de DNA enriquides als llocs d'unió d'Ibf-1, Ibf-2 i CP190	164
Figura 64.- Percentatge de pics en cada ChIP-seq que contenen cadascun dels motifs	166
Figura 65.- Gel de retard amb el Motiu Ibf	166
Figura 66.- Gel de retard amb el Motiu Ibf amb competidors	167
Figura 67.- Gel de retard de diferents seqüències d'unió	168
Figura 68.- Ibf-1 és necessari per l'enhancer-blocking al Fab-7.	170
Figura 69.- Ibf-2 és necessari per l'enhancer-blocking al Fab-7 i Fab-8.	170

Índex de Taules

Taula 1.- Llista de primers utilitzats.	65
Taula 2.- Llista de plàsmids utilitzats.	66
Taula 3.- Llista d'anticossos utilitzats	67
Taula 4.- Identificació per espectrometria de masses de les proteïnes presents a les elucions de la purificació.	113
Taula 5.- Canvis d'expressió d'Ibf-1, Ibf-2 i CP190 en cadascun dels tres knock-downs.	144
Taula 6.- Gens significativament upregulats i downregulats	145
Taula 7.- Pics obtinguts al ChIP-Seq	150
Taula 8.- Nombre mitjà de Motif Ibf a cada pic en cada ChIP-seq i presència del Motiu Ibf en funció de la presència de CP190 i de Ibf1 i 2	168

1.- INTRODUCCIÓ

1.1 L'emmagatzematge de la informació genètica al nucli cel·lular

El DNA és la molècula on s'emmagatzema la informació necessària per construir i mantenir un ésser viu. Es troba a tots els organismes vius existents i a totes les cèl·lules dins de cada organisme. En el cas dels eucariotes es troba delimitat físicament dins la cèl·lula per un orgànel cel·lular, el nucli —a diferència del que passa en procariotes on no hi ha nucli. El nostre objecte d'estudi és el DNA eucariota, serà per tant important tenir sempre present la dimensió física i espacial que representa el nucli, no només contenidor passiu de la informació genètica sinó medi d'anclatge i organitzador de les activitats gèniques.

El DNA contingut en una sola cèl·lula humana estirat al seu màxim faria dos metres de longitud mentre que un nucli, tot i ser l'orgànel més gran d'una cèl·lula, fa de mitjana només sis micròmetres (Alberts et al., 2002), fent difícil d'entendre com hi cap tant DNA en un espai tant petit. Aquesta aparent incompatibilitat de tamanys ha estat resolta mitjançant un precís mecanisme d'empaquetament del DNA, que s'enrotlla al voltant d'unes proteïnes nuclears anomenades histones, formant unes unitats conegudes com a nucleosomes, que al seu torn també s'enrotllen i juntament amb més proteïnes d'unió al DNA i als propis nucleosomes formen el que s'anomena cromatina. Com es veurà a continuació, la cromatina no només compleix perfectament la seva funció d'empaquetament sinó que també és essencial a l'hora de regular els diferents processos bioquímics que ha de patir el DNA per transmetre la informació adequadament.

1.1.1 La cromatina

La unitat bàsica de la cromatina són els nucleosomes, formats per un octàmer on dos dímers H2A-H2B interaccionen amb un tetràmer H3-H4, quedant una estructura $2(H2A-H2B)-(H3)_2(H4)_2$ al voltant del qual el DNA hi dona

una volta i tres quarts cap a l'esquerra ocupant unes 146 pb aproximadament. Tant a l'entrada com a la sortida del nucleosoma hi ha la histona H1, tancant les dues voltes completes de DNA, uns 166 pb. Al DNA entre nucleosomes se l'anomena DNA linker, i la seva longitud és variable (Brown, 2001).

L'estructura dels nucleosomes és molt similar entre tots els metazous, essent les histones unes proteïnes altament conservades degut a la transcendència del correcte empaquetament del DNA. Són proteïnes de pH bàsic que comparteixen entre elles un domini globular conservat, el *histone fold domain*, que forma el centre del nucleosoma, i unes cues amino-terminals de 15-30 residus que s'extenen fora del nucli (Alberts et al., 2002). El *histone fold domain* o HFD està format per tres hèlix alfa separades per dos loops (Talbert and Henikoff, 2010).

Els contactes entre el DNA i les histones es donen cada deu nucleòtids aproximadament amb el sòlc menor cap a dins i sempre per la banda del grup fosfat a través de ponts d'hidrogen i interaccions electrostàtiques, i amb la deoxiribosa amb contactes no polars. La manca de contactes amb les bases nitrogenades és coherent amb la no especificitat de seqüència de les histones, podent interaccionar amb qualsevol seqüència. Les cues de les histones surten cap a fora i queden exposades podent patir modificacions post-traduccionals (PTMs), essent aquest un punt fonamental de la regulació de l'estructura de la cromatina i de processos com la transcripció o la replicació (Kornberg and Lorch, 1999).

De forma general s'ha vist que els nucleosomes reprimeixen la transcripció. Tant les interaccions DNA-histones al nucli del nucleosoma com les interaccions amb les cues de les histones són causants d'aquesta repressió. La repressió deguda a les interaccions histones-DNA poden ser contrarestades pels complexes remodeladors de la cromatina (Kornberg, 2007). D'aquesta manera les regions transcripcionalment actives acostumen a trobar-se en dominis amb poca densitat de nucleosomes, que a nivell experimental són detectables degut

a la seva elevada sensibilitat a la digestió amb DNAsa en contrast amb la resistència que ofereixen regions més compactades o amb més densitat de nucleosomes.

1.1.1.1 Estructures de la cromatina

Els nucleosomes i el DNA fan una fibra d'11 nm anomenada *beads on a string*. De totes maneres l'estructura secundària formada pel DNA i els nucleosomes units per la H1 és la fibra de 30 nm (Figura 1). Hi ha dos models per explicar la conformació exacta d'aquesta fibra: el model solenoid, on els nucleosomes es col·loquen l'un al costat de l'altre en una hèlix simple; i el model on es col·loquen en zig-zag formant una doble hèlix (Tremethick, 2007; Li and Reinberg, 2011). La cromatina es condensa durant la mitosi per tal de facilitar la separació de cromàtides durant l'anafase. En aquest punt els cromosomes poden ser observats integralment degut a l'alta condensació que arriba a fer 1400 nm (Alberts et al., 2002).

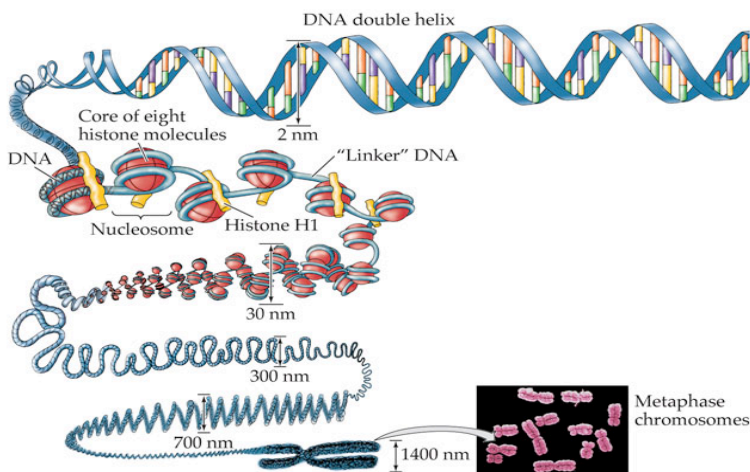


Figura 1.- L'empaquetament de la cromatina. Per tal d'encabir tot el DNA al nucli de cada cèl·lula, aquest es compacta a diferents nivells. En un primer nivell, al voltant dels octàmers d'histones formant els nucleosomes. Aquests es disposen formant la fibra de 30 nm, que es super-enrotlla fins a assolir la màxima compactació dels cromosomes mitòtics (Sadava et al., 2009).

A gran escala també són observables diferències entre la cromatina menys condensada (“oberta”, “activa”) i la més condensada (“silenciada”, “compactada”). Així, en cromosomes interfàsics es pot observar la cromatina oberta, anomenada eucromatina, i la condensada, anomenada heterocromatina. A l’eucromatina es troben la major part de gens transcripcionalment actius, mentre que l’heterocromatina es troba en forma de blocs de cromatina compactada i inaccessible per exemple als voltants del centròmer, al nuclèol o als telòmers. Sovint hi ha zones eucromàtiques que es silencien i es tornen heterocromatina, a aquesta se l’anomena facultativa en contraposició amb l’heterocromatina constitutiva, aquella que sempre està silenciada.

Un cas particular d’estructura cromosòmica és el dels cromosomes politènics (Figura 2). Aquests cromosomes gegants —unes 70-110 vegades més grans que un cromosoma mitòtic— han estat molt utilitzats en l’estudi de la genètica a *Drosophila* degut a la seva fàcil visualització al microscopi (que seria impossible en cromosomes interfàsics), majorment extrets de les cèl·lules de glàndula salival de larves, tot i que també es troben en altres teixits.

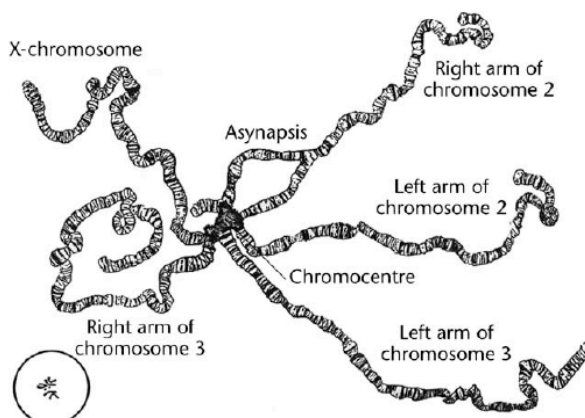


Figura 2.- Els cromosomes politènics. Cada cromosoma parental està aparellat amb el seu homòleg (sinapsi somètica). Hi ha regions on els dos homòlegs no estan aparellats (asinapsi). Tots els cromosomes estan units pel seu pericentròmer formant el cromocentre. A baix a l’esquerra hi ha un dibuix de cromosomes mitòtics d’ovari a la mateixa magnificació (Painter and Griffen, 1937).

El seu exagerat tamany és fruit d'un cicle cel·lular especial en el que només hi ha dues fases: S (síntesi) i G (intersíntesi). Després de la replicació les cadenes de DNA romanen unides i no se separen com en una mitosi normal, en un fenomen anomenat endoreplicació.

Els cromosomes politènics, mirats al microscopi, estan dividits per unes bandes més fosques i unes de més clares. Les bandes fosques, anomenades bandes, són riques en DNA i proteïnes, mentre que les clares, anomenades interbandes, ho són menys. El patró de bandes i interbandes és únic per cada cromosoma, permetent així el mapatge i identificació de cadascuna de les bandes. Tot i que la naturalesa molecular de bandes i interbandes no es coneix amb precisió, les bandes s'han associat tradicionalment a regions compactades i més aviat silenciades mentre que les interbandes correspondrien regions descompactades i transcripcionalment actives (Zhimulev et al., 2012).

1.1.1.2 Remodelació de la cromatina

L'estructura de la cromatina es pot modificar reclutant complexos proteics que mobilitzen o alteren el posicionament dels nucleosomes utilitzant ATP. Els remodeladors ATP-dependents pertanyen a la superfamília helicasa 2 (SF2). Entre la varietat de famílies que hi pertanyen, destaquen ISWI, SWI2/SNF2, BRM o CHD1 (Bouazoune and Brehm, 2006).

El complex més extensament estudiat és SWI/SNF, conservat des de llevat fins a humans i gran, d'uns 2MDa. Sembla que aquests complexos són reclutats per factors d'unió al DNA específics per la regulació de l'activitat transcripcional. La família CHD1 conté un cromodomini i un motiu d'unió al DNA. Normalment aquestes ATPases es troben en proteïnes que també contenen subunitat desacetilasa com NuRD o Mi-2, suggerint que utilitzen l'activitat remodeladora de la cromatina per desacetilar histones (Lusser and Kadonaga, 2003).

En general, els factors remodeladors de la cromatina poden exhibir diverses activitats (Lorch et al., 2001; Fan et al., 2003; Längst and Becker, 2004):

- catalitzen la mobilització dels nucleosomes en relació al DNA
- disrupció dels contactes DNA-histona
- transferència d'histones d'una cadena de DNA a una altra
- desasseblatge de nucleosomes
- generació de torsions superhelicals al DNA

Els mecanismes que s'han proposat com a model de funcionament d'aquests complexes es poden agrupar en dos principals (Figura 3). En un model (*twisting model*) el DNA rellicsa al voltant del nucleosoma i hi rota al voltant, substituint les bases que interaccionaven amb el nucleosoma per unes altres. Es mou per tant de base en base. En un model alternatiu (*loop recapture*), un loop petit de DNA es dissocia de l'octàmer i s'hi torna a unir en una altra posició del nucleosoma, establint així la direccionalitat de la rotació i propagant el loop (Lusser and Kadonaga, 2003; Längst and Becker, 2004).

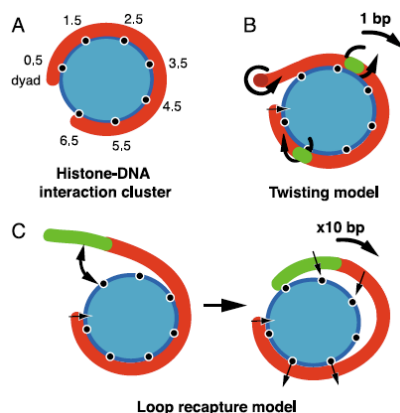


Figura 3.- Models de funcionament dels remodeladors de cromatina. A) Esquema de les interaccions DNA-histones, cada punt indica contacte entre DNA i histona anomenats *Superhelical Locations* SHL (0.5, 1.5, ..., 6.5). B) *Twisting model*. C) *Loop recapture model* (Längst and Becker, 2004).

1.1.1.3 Variants d'histones

Les histones H2A, H2B, H3 i H4 s'anomenen histones canòniques, són les majoritàries i la seva funció principal és l'empaquetament i la regulació de la transcripció. Els gens que les codifiquen solen estar disposats en clusters de repeticions i la seva transcripció se sol acoblar al moment de replicació, que és quan se les necessita per compactar el nou DNA sintetitzat. Ara bé, hi ha unes altres histones, les no canòniques, que difereixen en la seva seqüència i estructura de les canòniques i que porten a terme funcions diferents, com la reparació del DNA, recombinació meiótica, segregació de cromosomes, iniciació i terminació de la transcripció o empaquetament dels cromosomes de l'esperma (Brown, 2001; Talbert and Henikoff, 2010). Mentre que hi ha variants d'histona que són específiques de llinatge com les macroH2A o variants d'histona de l'esperma, algunes de les histones no canòniques són gairebé universals, estant presents a la majoria de llinatges eucariotes, i s'expliquen a continuació.

La variant centromèrica de la histona H3, CenH3 (CENP-A a humans; Cse4 a *Saccharomyces cerevisiae*; CID a *Drosophila*) és essencial per l'assemblatge del cinetocor. Les CenH3, que estan menys conservades que la canònica H3, tenen un 50-60% d'identitat amb la H3 al HFD i cap conservació a la cua N-terminal. Els nucleosomes amb CenH3 a *Drosophila* i a llevat induïx superenrotllaments positius del DNA que impliquen un enrotllament cap a la dreta, a diferència de l'enrotllament cap a l'esquerra de les histones canòniques. La seva transcripció està coordinada amb la replicació dels centròmers, de la fase S tardana a la G2 (Brown, 2001; Santaguida and Musacchio, 2009).

La H3.3 és una variant de la H3 de la qual només es diferencia per quatre aminoàcids. La seva transcripció pot ser independent de la replicació. És una variant de reemplaçament, es diposita als gens en transcripció, als promotors i als elements reguladors, així com durant les remodelacions de cromatina. La localització de la H3.3 als elements reguladors del genoma correspon a llocs hipersensibles a nucleases donat el procés dinàmic de

disrupció i reemplaçament de nucleosomes a aquests llocs (McKittrick et al., 2004; Schwartz and Ahmad, 2005).

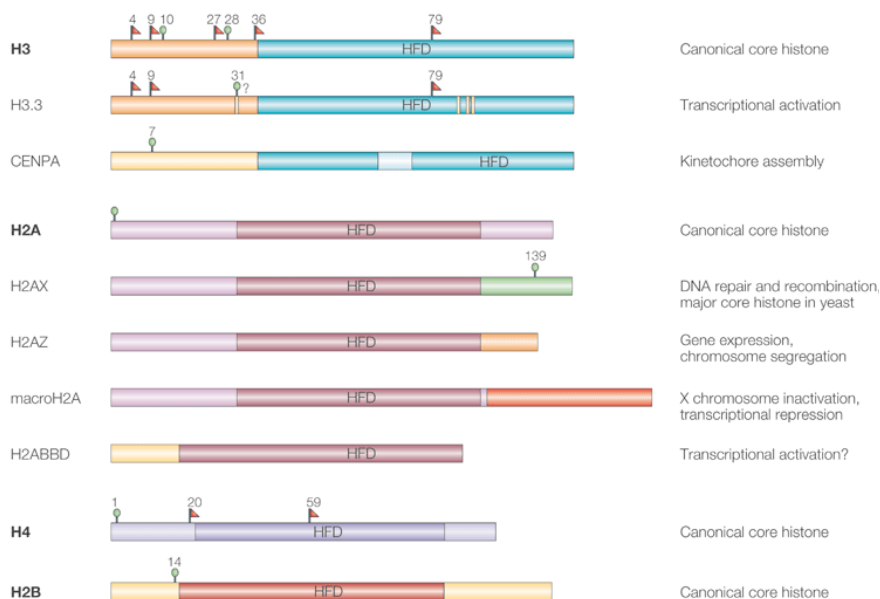


Figura 4.- Histones canòniques i no-canòniques. En banderes vermelles s'indiquen els llocs de metil·lació de lisines, i en verd les fosforilacions de serines. A la H3.3 en groc s'indiquen posicions que difereixen de H3. CENP-A té al seu HFD una regió que la porta al centròmer (blau cel), i una regió N-terminal diferent de les demés histones (groc fluix). A la dreta de cada variant se n'indica el procés amb el qual es relacionen (Sarma and Reinberg, 2005).

La H2A.Z és una variant de la H2A amb la que difereix amb el domini C-terminal que contacta amb la H3. Els nucleosomes amb H2A.Z es troben flaquejant les zones lliures de nucleosomes als llocs d'inici de la transcripció (TSS) a on promouen el reclutament de la RNA Polimerasa II. Si s'incorpora erròniament a dins del gen pot causar inici de transcripció críptica. Quan la RNA Pol II avança pel gen transcrivint, causa l'alliberament dels dímers H2A.Z-H2B i es reincorporen dímers H2A-H2B. A part de la seva contribució a la transcripció ha estat implicada en reparació del DNA; formació d'insulators; diferenciació de cèl·lules mare embrionàries o supressió d'RNA anti-sentit (Adam et al., 2001; Talbert and Henikoff, 2010).

La H2A.X és una altra variant de la H2A que s'hi assembla en l'HFD però hi difereix en una serina de la part C-terminal que pot ser fosforilada ràpidament produint la forma γ -H2A.x per les quinases ATM, ATR o DNA-PK. Aquesta fosforilació té lloc en resposta a trencaments de la doble cadena de DNA (*double-strand breaks* o DSB), en els que la γ -H2A.x ajuda a reclutar o retenir proteïnes de reparació del DNA i complexos remodeladors de la cromatina. Tot i que la seva funció en DSB està ben conservada, s'ha observat un origen filogenètic divers per aquest histona, per exemple a *Drosophila* és la H2Av, que prové de la H2A.Z que ha adquirit una serina que també pot ser fosforilada en resposta a DSB (Brown, 2001; Talbert and Henikoff, 2010).

1.1.1.4 Modificacions post-traduccionals de les histones

La cua N-terminal de les histones pot patir modificacions covalents com fosforilacions, ubiquitinacions, acetilacions o metilacions. Aquestes modificacions poden influir en les interaccions DNA-nucleosomes i per tant acabar afectant l'estructura de la cromatina, participant en processos com la transcripció, la replicació, la reparació del DNA o la progressió del cicle cel·lular. Les modificacions es donen de manera específica a cada lloc del genoma i de manera específica a nivell cel·lular, de manera que són fonamentals en la determinació de la identitat cel·lular i en el procés de diferenciació (Bhaumik et al., 2007).

L'acetilació es dona a lisines concretes de totes quatre histones, neutralitzant la càrrega positiva d'aquest residu. S'ha proposat que l'acetilació crea un estat obert i accessible de la cromatina a l'altre les interaccions DNA-histones (Shahbazian and Grunstein, 2007). La reacció la catalitzen les acetiltransferases (HATs), sovint presents en activadors transcripcionals com Gcn5/PCAF o CBP/p300. L'acetilació pot ser revertida per les desacetilases d'histones (HDACs). De la mateixa manera, molts repressors transcripcionals contenen activitat desacetilasa com mSin3a o NURD/Mi-2. Rpd3 conté activitat desacetilasa i s'associa a la forma elongativa de la RNA Pol II. A través d'aquesta

La fosforilació d'histones té un paper fonamental en la progressió del cicle cel·lular i en la resposta a estímuls externs i es pot donar a totes quatre histones, a residus serina i treonina. Els cromosomes es troben altament fosforilats en la serina 10 i 28 de la histona H3 durant la mitosi per l'acció de les Aurora quinases, molt ben conservades de llevat a humans (Koshland and Strunnikov, 1996). La família de quinases MSK/Jil-1 en canvi medien la fosforilació de la serina 10 durant la interfase, marca associada amb l'activació de la transcripció i la inducció de transcripció en resposta a estímuls externs (Jin et al., 1999; Wang et al., 2001; Regnard et al., 2011). La fosforilació de la H2A és una resposta a la senyalització de dany al DNA (Bhaumik et al., 2007).

La metilació es dona sobre residus lisina, arginina o sobre tots dos principalment a les histones H3 i H4. La metilació de lisines pot ser en forma mono, bi o tri sobre les lisines 4, 9, 27, 36 i 70 de la H3 i sobre la lisina 20 de la H4. Les diferents metilacions participen en processos i funcions diferents. A diferència de les HATs, que poden modificar diferents residus de la mateixa o de varies histones, les metiltransferases són altament específiques de substrat (Shilatifard, 2006; Bhaumik et al., 2007). Gairebé totes les metiltransferases d'histones caracteritzades contenen un domini SET. Set1 és una metiltransferasa que metila la lisina 4 de la H3, que és una marca que es troba a les regions 5' de gens activament transcrits. En canvi Set2 metila la lisina 36 de la H3, marca relacionada amb l'elongació ja que es troba tot al llarg del gen entre 5' i 3' en gens que s'estan transcrivint. Una metiltransferasa particular és Dot1, ja que no té domini SET. Dot1 catalitza la metilació de la lisina 79 de la H3, necessària pel silenciament dels telòmers (Lacoste et al., 2002).

Hi ha altres metilacions relacionades amb el silenciament. Per exemple la metilació de la lisina 9 de la H3 és característica de l'heterocromatina. La metiltransferasa Suv39 és qui la metila, i recluta el repressor transcripcional HP1, tot i que el mecanisme molecular de compactació de la cromatina per HP1 no està clar. La trimetilació de la lisina 27 de la H3 també està involucrada en el silenciament gènic. És característica dels dominis silenciats per Polycomb.

La desmetilació és duta a terme per dues famílies d'enzims, oxidases com LSD1 i hidroxilases de la família JmjC. Només aquesta última família d'enzims pot desmetilar trimetilacions. En general presenten especificitat, com Lid a *Drosophila* que només catalitza la desmetilació de H3K4me3 a H3K4 me2 (Schneider and Shilatifard, 2006).

El conjunt de les modificacions d'histones i la seva correlació amb certs processos o estats gènics van portar a alguns autors a plantejar-se si en realitat el que estaven intentant desxifrar no era sinó un codi, és a dir, un conjunt de senyals que són interpretades a través d'una equivalència coneguda i causen un efecte específic i coherent amb la senyal reconeguda. Així, Strahl i Allis van proposar la hipòtesi del codi de les histones que postula que els estats de les diferents modificacions de les histones, actuant de manera combinatòria o seqüencial en una o més cues d'histones, especifiquen i causen funcions úniques (Strahl and Allis, 2000). Aquesta hipòtesi es basa, entre altres premisses, en l'existència de totes les proteïnes amb dominis que reconeixen específicament les modificacions, com per exemple els bromodominis amb acetilacions, els 14-3-3 amb fosforilacions o els PHD fingers amb les metilacions.

1.2 La regulació de l'expressió gènica

El DNA conté la informació suficient que la cèl·lula necessita per sintetitzar proteïnes. Aquesta informació s'anomena *codificant* perquè l'ordre en què estan disposats els nucleòtids al llarg de la cadena de DNA és interpretat per la cèl·lula com l'ordre en què s'han de disposar els aminoàcids en un polipèptid. L'equivalència entre triplets de nucleòtids (codons) i aminoàcids és el que es coneix com a codi genètic, i és la base del Dogma Central de la Biologia. Aquest dogma postula que el gen, l'unitat bàsica d'informació del DNA, és transcrit a RNA missatger que es trasllada als ribosomes on es produeix la traducció d'informació i es sintetitzen les proteïnes necessàries pel desenvolupament i manteniment de l'organisme.

Però el DNA conté un altre tipus d'informació: aquella que indica quan, quant i com s'ha d'utilitzar la informació codificant. Aquest, anomenat per oposició DNA no codificant, ocupa la major la major part del genoma eucariota i el seu mode de funcionament sol ser resultat de la interacció amb proteïnes nuclears, tot i que també es dona el cas que interactui amb altres àcids nucleics, com una altra cadena de DNA o molècules d'RNA. De fet el DNA no codificant també es transcriu, en alguns casos *ad hoc* (la molècula transcrita és necessària per a dur a terme una funció), i en alguns casos com a subproducte conseqüència del pas de la màquinària transcripcional per una regió de DNA. En cap dels dos casos hi esdevindrà una traducció en base al codi genètic, ja que la informació que contenen no té cap relació amb l'ordre dels aminoàcids d'una proteïna. Al DNA no codificant la lògica interpretativa que hem de seguir per entendre l'ordre de les bases és completament diferent i més variable: de vegades té a veure amb l'afinitat i especificitat d'una sèrie de bases per part d'una proteïna (*binding sites*), de vegades té a veure amb la formació de *loops* en la molècula transcrita (*microRNAs* per exemple) i inclús de vegades pot tenir a veure amb la formació d'estructures alternatives de les hèlix de DNA (com en el cas dels *G-quadruplex*), entre molts altres casos descoberts i per descobrir.

En qualsevol cas, de tot el DNA no codificant, en aquest treball pretenem centrar-nos en aquell la missió del qual és controlar el nivell de transcripció dels gens. A l'hora de parlar de gens ens hi referim en un sentit ampli, no necessàriament codificant ja que inclou regions que transcriuen RNA no codificants, la definició que més s'hi aproximaria seria el gen com a "unió de seqüències genòmiques que codifiquen un conjunt coherent de productes funcionals potencialment sobreposats" (Gerstein et al., 2007). Aquesta puntualització és important per captar d'entrada la complexitat i la diversitat de l'estudi de la regulació de la transcripció, que com s'ha vist recentment és un fenomen que es dona a escala genòmica i no restringit a la fracció codificant (Djebali et al., 2012) com per exemple al genoma del ratolí on es transcriu un 63% del genoma mentre que només un 2% es tradueix finalment a proteïna (Carninci et al., 2005).

Ens referirem a aquesta part del DNA no codificant com a regions reguladores o DNA regulador, per distingir-lo d'altres tipus de DNA no codificant que no seran objecte del nostre estudi, com els transposons o regions cromosòmiques peculiars com els telòmers —malgrat també s'estiguin començant a descriure rols reguladors per a alguns d'ells (Stamatoyannopoulos, 2012).

Aquest DNA regulador es troba tant en regions intergèniques com dins dels gens i és essencial a l'hora de determinar el destí d'una cèl·lula. Pot haver-hi multitud de tipus cel·lulars diferents en un organisme i tots ells tenen un origen comú. La diferència fonamental entre uns i altres és el diferent programa transcripcional que han tingut al llarg del llinatge. Aquest programa transcripcional —espectre global de gens que s'expressen en una cèl·lula— és l'últim responsable de la identitat d'una cèl·lula i ve determinat tant per factors interns (identitat de la cèl·lula progenitora) com per factors externs: substàncies extracel·lulars que a través d'una cascada de senyalització modelen el programa transcripcional.

En qualsevol cas, l'acurada determinació de l'estat transcripcional d'un tipus cel·lular es modula a través d'una combinació de mecanismes en els quals el DNA regulador juga un paper essencial. Entre aquests mecanismes destaquen la participació de factors de transcripció sobre enhancers específics, el silenciament dut a terme per elements de resposta a Polycomb (PRE), l'aïllament de regions reguladores per insulators o diverses modificacions de les histones. Qualsevol error o malfunció d'aquests mecanismes pot tenir conseqüències fatals pel desenvolupament de l'organisme. Diversos desordres humans amb conseqüències negatives tant físiques com mentals com el síndrome de Rett, d'Angelmann, Prader-Willi, així com varis tipus de càncer són fruit d'alteracions en els sistemes de regulació genòmica (Egger et al., 2004).

A continuació s'explicaran amb més detall els principals elements de control de l'expressió gènica, però en primer lloc cal entendre bé com funciona aquest mecanisme tant finament regulat: la transcripció.

1.2.1 La transcripció

La transcripció comença quan la RNA polimerasa catalitza la producció d'RNA a partir d'una cadena motlle de DNA. Les polimerases són enzims grans que es solen compondre de més d'una desena de subunitats, tot i que quan estan actives sobre el DNA formen complex amb altres factors.

A cèl·lules eucariotes existeixen tres tipus diferents d'RNA polimerasa. A grans trets i amb algunes excepcions es podria dir que l'RNA polimerasa I transcriu els gens que codifiquen la majoria d'rRNA (RNA ribosomal); l'RNA polimerasa II és l'encarregada de transcriure els RNA missatgers (aquells que serviran de motlle per sintetitzar proteïnes en la traducció); i l'RNA polimerasa III pels tRNA (RNA de transferència) entre altres RNA reguladors petits.

El primer pas de la transcripció és la formació al promotor del complex de pre-iniciació (PIC), integrat principalment per la polimerasa i els factors generals de transcripció, que dóna pas al promoter melting que és quan les

cadena de DNA es separen per poder començar a transcriure, és la fase d'iniciació. L'avenç de la maquinària transcripcional al llarg del gen és el que es coneix com a fase d'elongació, a mesura que va avançant va sintetitzant la molècula de mRNA. Finalment la polimerasa arriba a l'extrem 3' del gen i s'allibera del DNA i de l'RNA, és la terminació.

1.2.1.1 La RNA polimerasa

La RNA polimerasa II és un complex compost per dotze subunitats, de la Rpb1 a la Rpb12, amb un pes molecular d'uns 0.5 MDa. El domini C-terminal (CTD) de la subunitat Rpb1, la més gran, té un motiu heptapèptid, Tyr-Ser-Pro-Thr-Ser-Pro-Ser, que es repeteix 45 vegades a *Drosophila*, 26 a llevat i 52 a humans (Chiba et al., 2010).

Tot i que les tres RNA polimerases tenen el mateix origen evolutiu i tenen una composició de subunitats similar, el CTD és particular de la Pol II. En efecte, aquest domini està directament relacionat amb els passos de la transcripció de l'mRNA, iniciació, elongació i terminació, així com el capping, l'splicing i el processament del 3' (Proudfoot et al., 2002). El processament de l'mRNA està, a diferència del que es creia fa temps, acoblat amb la transcripció espai-temporalment. El rol que juguen les diferents combinacions de modificacions bioquímiques del CTD es coneixen com a *CTD code* (Buratowski, 2003). Les modificacions principals i més estudiades fins al moment són les fosforilacions de les series 2, 5 i 7 (Egloff et al., 2012). Les serines 5 i 7 estan fosforilades per l'activitat quinasa del TFII al voltant del TSS, mentre que la serina 2 està fosforilada durant l'elongació pel factor P-TEFb (*positive elongation factor b*).

1.2.1.2 Els factors generals de transcripció

Tot i que l'RNA polimerasa II és l'enzim que sintetitza l'RNA d'un DNA motlle, no és capaç per si sola de reconèixer el promotor basal. Aquest és un procés que requereix uns factors addicionals que es coneixen com a Factors

Generals o Basals de Transcripció, que són TFIIA, TFIIB, TFIID, TFIIE, TFIIF i TFIIH (*Transcription Factor for RNA Polymerase II A-H*).

El requeriment de cadascun d'ells és dependent de la composició del promotor. En promotors amb caixa TATA la Pol II necessita els TFIIB, TFIID, TFIIE, TFIIE i TFIIH. No obstant, aquests factors no són suficients per començar la transcripció en promotors amb DPE (Lewis et al., 2005).

TFIID és un complex que comprèn la TBP i una dotzena de TAFs. Hi ha múltiples punts d'interacció amb els elements del promotor: TBP s'uneix a la caixa TATA; TAF1 i TAF2 reconeixen l'Inr; TAF6 i TAF9 interaccionen amb el DPE. TFIIB interacciona amb TBP i assisteix en el recrutament de la polimerasa al promotor. TFIIB s'uneix al promotor a través del reconeixement específic de les seqüències BRE de manera TBP-dependent. TFIIA assisteix al recrutament de TBP a la caixa TATA. TFIIE, TFIIF i TFIIH actuen suseqüentment a la unió de TFIID i TFIIB al promotor basal i medien en el procés de separació de cadenes de DNA al inici de la transcripció (Juven-Gershon and Kadonaga, 2010).

1.2.2 Els elements que regulen la transcripció

1.2.2.1 El promotor

El promotor core o basal és la mínima seqüència de DNA suficient per dirigir la iniciació de la transcripció per la maquinària de la RNA polimerasa II, mentre que s'anomena promotor proximal a la zona que cobreix uns 250 nucleòtids a una banda o l'altra del TSS (Butler and Kadonaga, 2002). El promotor core o basal típicament inclou el lloc d'inici de la transcripció (TSS) i s'extén 35 nucleòtids més enllà ja sigui upstream o downstream. Hi ha doncs ocasions on un promotor basal pot ser de només 40 nucleòtids. S'hi troben amb freqüència diversos motius de seqüència: la caixa TATA, l'element Inr, el TFIIB recognition element (BRE) o el downstream core promoter element (DPE). Cadascun d'ells té la seva funció específica, i cap d'ells es troba a tots els promotors: no hi ha cap element de promotor que sigui universal.

La caixa TATA s'anomena així perquè conté una seqüència consens TATAAA, tot i que amb certa variabilitat. S'ha estimat que a *Drosophila* aproximadament un 43% de 205 promotors basals contenen una caixa TATA (Kutach and Kadonaga, 2000). La TBP (TATA-Binding-protein) és la proteïna que més sovint s'uneix a la caixa TATA, que té dos TBP-related factors (TRF1 i 2, a *Drosophila*). TRF1 és capaç d'unir-se a la caixa TATA mentre que TRF2 no (Butler and Kadonaga, 2002).

A l'element Inr és on es troba l'inici de transcripció. A *Drosophila* té una seqüència consens T-C-A₊₁-G/T-T-C/T (Kutach and Kadonaga, 2000). Normalment la transcripció s'inicia al nucleòtid A₊₁ però no és sempre el cas. Tant el factor general de transcripció TFIID com la RNA pol II són capaços d'interaccionar de manera específica amb l'element Inr, així com altres proteïnes d'unió a DNA com YY1 (Weis and Reinberg, 1997).

L'element DPE es situa a entre +28 i +32 de la posició A₊₁ de l'Inr, i sol ser a promotors TATA-less. La seqüència consens a *Drosophila* és A/G₊₂₈-G-A/T-C/T-G/A/C. La distància entre l'Inr i el DPE és tant important que si es decala d'un sol nucleòtid baixa dràsticament el nivell de transcripció basal així com la unió de TFIID, que sembla que és qui reconeix aquest element. A més s'ha estudiat la seva freqüència al genoma de *Drosophila* i s'ha vist que és tant comú com ho és la caixa TATA (Kutach and Kadonaga, 2000). Sovint s'estableix una relació mutuament excloent entre DPE i TATA. És el cas per exemple dels gens homeòtics a *Drosophila* (Juven-Gershon and Kadonaga, 2010). S'ha vist que la majoria d'aquests gens són TATA-less i tenen DPE. Així, un factor de transcripció regulador d'aquests gens com *Caudal* és específic de DPE.

El BRE és un element reconegut específicament per la TFIIB i que es troba immediatament upstream de la caixa TATA (Lagrange et al., 1998). La seva contribució a la transcripció, negativa o positiva, no està del tot clara (Butler and Kadonaga, 2002).

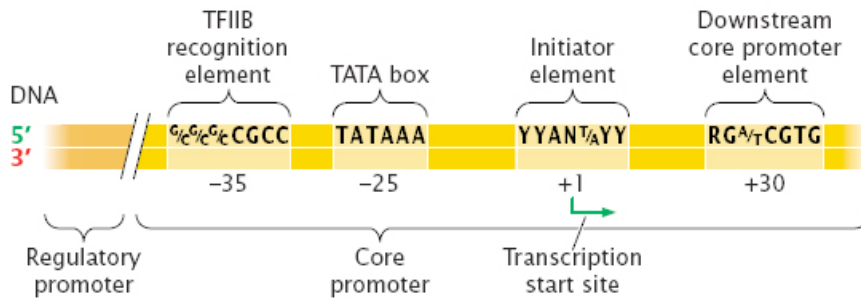


Figura 6.- Elements del core promoter. El promotor basal és la zona que envolta el TSS, a diferència del promotor proximal, que abarcaria també zones reguladores més allunyades (*Regulatory promoter*). El TFIIIB recognition element o BRE; la caixa TATA; l'*initiator element* o Inr; i el *Downstream core promoter element* o DPE (Clancy, 2008).

1.2.2.1.1 Promotors pausats

NELF (Negative elongation factor) és un factor format per quatre subunitats (A, B, C/D i E) que com el seu nom indica actua negativament sobre l'elongació. P-TEFb és una quinasa composta per Cdk9 i Ciclina T que actua positivament sobre l'elongació. DSIF està compost per Spt4 i Spt5 i té un rol dual sobre l'elongació: s'uneix a la Pol II i hi incorpora varis factors al complex d'elongació.

Ben aviat després de l'iniciació de la transcripció DSIF i NELF s'uneixen a la Pol II i la pausen uns 25-50 nucleòtids downstream del TSS. Es tracta d'un pas limitant generalitzat anomenat *promoter-proximal pausing*, pausació proximal al promotor. En aquest punt és la serina 5 i no la serina 2 del CTD de la Pol II la que es troba fosforilada. Quan P-TEFb fosforila la serina 2 es dissocia NELF de la Pol II i s'allibera de la pausació, reemplantant l'elongació. P-TEFb també fosforila NELF-E i Spt5, la subunitat gran de DSIF. Això fa que DSIF passi d'activador a repressor mitjançant el recrutament del complex Paf1 i Tat-SF1 al complex elongador. Tots aquests factors faciliten la continuació de l'elongació (Chiba et al., 2010).

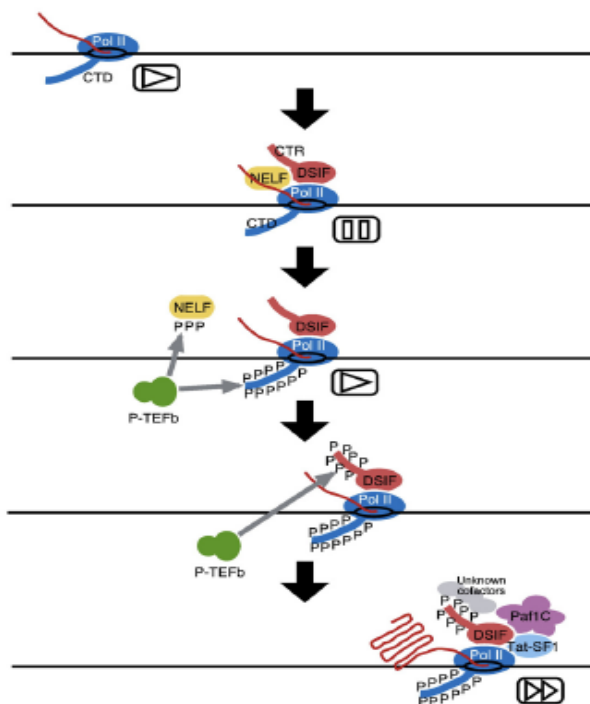


Figura 7.- Model de control de la pausació de la polimerasa. Just després de l'inici de la transcripció, DSIF i NELF s'uneixen a la Pol II i la pausen. Quan P-TEFb fosforila el CTD de la Pol II i a DSIF, es reprèn l'elongació i s'uneixen factors com Paf1C o Tat-SF1 (Chiba et al., 2010).

Aquest és un mecanisme generalitzat a la majoria de gens, tot i que la posició precisa de pausació o l'intensitat varia d'un gen a l'altre. Aquesta sembla principalment regulada a través del control de l'activitat de PTEFb. El recrutament de PTEFb a un gen és específica i dependent de senyals, com per exemple heat-shocks, que alteren dràsticament la distribució al genoma de PTEFb en segons (Lis et al., 2000). Els estudis fets a escala genòmica mostren com la polimerasa pausada es troba a més d'un 30% dels gens, molts d'ells relacionats amb el desenvolupament i els *immediate-early genes* (gens de resposta ràpida a estímuls extracel·lulars, com els gens de resposta a heat-shock). En canvi, no es va trobar a gens expressats ubiqüament de metabolisme i creixement cel·lular (Zeitlinger et al., 2007).

Alguns estudis suggereixen una relació directa entre pausació i regió lliure de nucleosomes, mostrant com una depleció de NELF comporta la pèrdua de les regions lliures de nucleosomes a loci que normalment tindrien la polimerasa pausada (Gilchrist et al., 2008). S'han proposat tres rols principals per la pausació (Chiba et al., 2010): el primer seria el de restringir els nivells de transcripció d'un gen; el segon seria donar resposta ràpida a gens en estat pausat, ja que el pas de recrutament de complexes i iniciació ja haurien estat completats i la cromatina es trobaria en un estat obert gràcies a la presència de la polimerasa i per tant l'elongació es donaria de manera immediata al rebre una senyal extracel·lular; i tercer seria la relació de la CTD de la Pol II amb el processament del mRNA, de manera que la pausació serviria de *checkpoint* per aquest correcte processament.

1.2.2.1.2 Promotors alternatius

Amb freqüència trobem gens que contenen més d'un promotor que dona lloc a un transcrit alternatiu dins del mateix gen. S'ha estimat que a *Drosophila* un 14% dels gens tenen promotors alternatius, amb una mitjana de 2.43 promotors/gen pels gens amb més d'un promotor (Zhu and Halfon, 2009). Aquests promotors alternatius afegeixen una capa més regulació, ja que sovint estan diferentment utilitzats tant a nivell temporal com espacial. Un exemple d'expressió diferencial de promotors alternatius a nivell temporal és la del gen *hunchback*. Els dos promotors de *hunchback* es troben a 3 kb de distància, els seus transcrits tenen un exó sencer diferent, i un es transcriu només a embrions primerencs mentre que l'altre ho fa també a adults (Bender et al., 1988).

Un altre exemple ben estudiat, en aquest cas d'expressió diferencial a nivell espacial, és el dels transcrits generats a partir del gen homeòtic *Abdominal-B* del Complex Bitòrax. Aquest gen genera dues proteïnes anomenades *m* (*morphogenetic*) i *r* (*regulatory*) que tenen rols diferents com el seu nom indica, i provenen de transcrits iniciats a partir de promotors alternatius. Així com la *m* s'expressa als parasegments 10-13, la *r* només ho fa al 14 i té un rol regulador

sobre l'expressió de l'm i d'altres gens homeòtics (Casanova et al., 1986; Delorenzi and Bienz, 1990).

Per tant es pot dir que l'ús de promotors alternatius té una doble funció regulatòria: permet l'expressió de més un transcrit a partir d'un mateix gen, i a més permet que l'expressió d'aquests pugui estar diferencialment regulada en funció del teixit i de l'estadi de desenvolupament. Els mecanismes involucrats en la selecció del promotor usat en cada cas han estat poc estudiats fins al moment, tot i que sí que s'ha vist que hi ha diferències a nivell de composició dels elements del promotor basal segons la posició d'aquest respecte els altres promotors (Zhu and Halfon, 2009). A vertebrats hi ha molts exemples descrits de com aquesta expressió diferencial segons el promotor afecta al desenvolupament, i per tant de com perturbacions en l'ús de promotors alternatius condueixen a diverses malalties (Ayoubi and Van De Ven, 1996). Un exemple és el del gen *Dclk1* (Figura 8) durant el desenvolupament del cervell a mamífers (Pal et al., 2011). Aquest gen, important per la migració neuronal i la neurogènesi, té dos promotors que generen transcrits diferents segons l'estadi del desenvolupament. El primer genera les isoformes 1 i 2 a embrions, mentre que el segon genera les isoformes 3 i 4 a adults, que manquen d'un domini d'unió als microtúbuls. La desregulació d'aquest patró porta a un desenvolupament aberrant del cervell.

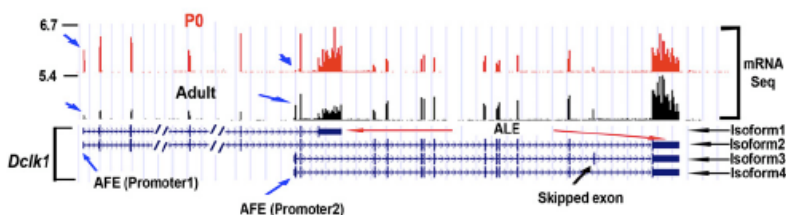


Figura 8.- Exemple de gen amb promotors alternatius. Nivells d'expressió dels diferents transcrits del gen *Dclk1* durant el desenvolupament del cervell a mamífers. Per mRNA seq es detecten diferències substancials entre el transcrits expressats a embrió (P0) i a adult: les isoformes 3 i 4 només s'expressen a adults. Una desregulació d'aquest patró condueix a un desenvolupament incorrecte del cervell (Pal et al., 2011).

1.2.2.2 Els enhancers

Els enhancers (o enhancers transcripcionals) es defineixen com seqüències de DNA que regulen la transcripció d'un gen encara que n'estiguin a milers de bases de distància upstream o downstream i en qualsevol orientació (Fiering et al., 2000; Martin, 2001). Els primers enhancers descrits eren unes seqüències víriques repetitives de l'SV40 que eren capaces d'activar l'expressió d'un constructe amb el gen de la beta-globina de conill insertat a un cultiu de cèl·lules HeLa (Banerji et al., 1981). Es van adonar que aquesta activació era independent de l'orientació i l'emplaçament de l'enhancer dins del plàsmid. Aviat es van trobar els primers enhancers de mamífer, els del locus de la cadena pesada de la immunoglobulina humana, que a més eren específics de teixit (Banerji et al., 1983). Les estimacions actuals prediuen que hi ha més d'un milió d'enhancers al genoma humà (Heintzman et al., 2009). S'ha descrit que no només els gens codificants estan sota el control d'enhancers sinó també microRNAs (Sheng and Previt, 2011).

Donat que no hi ha una distància ni posició fixa a la que es trobin els enhancers respecte el promotor sobre el que actuen, el descobriment exhaustiu d'enhancers al genoma no és una tasca evident. Un dels mètodes més utilitzats és el mapatge de regions hipersensibles a la DNAsa ja que les regions reguladores, al tenir menys densitat de nucleosomes, es digereixen més ràpidament amb DNAsa que la resta. Un altre mètode és l'*enhancer-trap* que consisteix en integrar a l'atzar al genoma un gen *reporter* de manera que si la zona on s'integra està sota el control d'un enhancer això es podrà observar i seleccionar la seqüència de l'enhancer. A la vegada, el patró d'expressió del gen *reporter* integrat serà l'indicador del mode d'actuació de l'enhancer.

Amb l'emergència de les noves tecnologies de seqüenciació *high-throughput* i d'anàlisi *genome-wide* la tasca està esdevenint més eficient. Per exemple es pot aplicar la seqüenciació exhaustiva als experiments d'hipersensibilitat a la DNAsa i així obtenir un mapa genòmic de zones de

cromatina oberta (Hesselberth et al., 2009). Per anàlisi d'immunoprecipitació de cromatina (ChIP) s'ha vist que els enhancers estan enriquits amb alts nivells de H3K4me1 combinat amb nivells baixos de H3K4me3 (Heintzman et al., 2009), també la H3K27Ac s'ha vist correlacionada amb enhancers actius (Creyghton et al., 2010). Així, algunes acetil-transferases com p300 o ATAC s'han vist localitzant als enhancers (Krebs et al., 2011). De manera que la combinació de diverses marques a la cromatina seguida de seqüenciació ha esdevingut una manera de detectar possibles enhancers a escala genòmica (Cotney et al., 2012). També altres mètodes com el FAIRE (*formaldehyde-assisted isolation of regulatory elements*) que identifica noves regions reguladores que tenen un estat obert de la cromatina (Giresi and Lieb, 2009), o GRO-seq (*global run-on followed by high-throughput sequencing*) que ha revelat que els enhancers es transcriuen bidireccionalment (Wang et al., 2011).

S'han suggerit diferents hipòtesis sobre com un enhancer que pot estar afectant la transcripció d'un gen que està a varis centenars o milers de bases de distància (Palstra and Grosveld, 2012). N'hi ha que no impliquen contacte directe entre enhancer i promotor com el *polymerase tracking* (l'avenç de la polimerasa des de l'enhancer fins al promotor activaria aquest últim) o la propagació d'estructures de la cromatina que extendrien un estat obert fins al promotor. De totes maneres els models més acceptats són els que defensen que hi ha contacte (a través de proteïnes) entre enhancer i promotor. Aquests, per sobrepassar la distància existent entre ambdós elements, crearien un *loop* que deixaria fora tota la seqüència que queda entremig. Un dels casos més ben detallats és el del locus de la beta-globina en el que es demostra que el Locus Control Region (LCR) que faria d'enhancer es troba pròxim als promotors dels gens actius mentre que la cromatina que queda entremig així com els gens inactius queden aïllats pel loop, com es demostra mitjançant Chromosome Conformation Capture o 3C (Tolhuis et al., 2002).

La següent pregunta és com l'enhancer activa la transcripció un cop es troba pròxim o en contacte amb el promotor. Tot i que és una qüestió que

encara no està resolta, sí que se sap que l'enhancer recruta un complex proteic conegut com a enhanceosoma que conté com a mínim factors de transcripció i co-activadors (Palstra and Grosveld, 2012). Un mecanisme suggerit seria regulant la transició de la polimerasa pausada a la polimerasa activa al promotor proximal ja que s'ha vist que la deleció de l'LCR de la beta-globina redueix l'elongació mentre que manté l'ensamblatge del complex de pre-iniciació i la presència de la polimerasa al promotor (Sawado et al., 2003). Alternativament s'ha suggerit que l'enhancer seria l'encarregat de reclutar la polimerasa i factors de transcripció i entregar-los al promotor, fet que s'ha demostrat *in vitro* però que manca d'evidències convinents *in vivo*. En qualsevol cas el més probable sembla que l'acció dels enhancers vagi més enllà d'una sola activitat i tinguin una tasca seqüencial al llarg de la diferenciació cel·lular: des de mantenir el locus en un estat obert, reclutar els factors de transcripció específics de teixit i desenvolupament, i finalment ensamblar el complex de pre-iniciació al promotor del gen a través de loops de cromatina i alliberar la polimerasa pausada reclutant factors d'elongació (Palstra and Grosveld, 2012).

1.2.2.3 Polycomb response elements

Els gens del grup Polycomb (PcG) van ser descoberts a *Drosophila* com a repressors de gens Hox, que codifiquen uns factors de transcripció encarregats de determinar l'identitat cel·lular al llarg de l'eix antero-posterior. Subseqüentment es va anar descobrint que formen grans complexos multimèrics que s'encarreguen de silenciar l'estat d'expressió dels gens Hox (Kennison, 1995), però també s'han vist involucrats en varis processos cel·lulars com el control del cicle cel·lular (Oktaba et al., 2008), la senescència (Bracken and Helin, 2009) o la diferenciació (Pasini et al., 2007).

Les proteïnes del grup Polycomb formen tres complexos diferents (Schuettengruber et al., 2007). El PRC2 (Polycomb Repressive Complex 2) conté quatre components: E(z)(Enhancer of zeste); Esc (Extra Sex Combs); Su(z)12 (Suppressor of Zeste 12) i Nurf-55. E(z) té un domini SET que trimetila la lisina 27

de la histona H3 (Cao and Zhang, 2004). El PRC1 (Polycomb Repressive Complex 1) conté Pc (Polycomb); Ph (Polyhomeotic); Psc (Posterior Sex Combs) i dRing. Pc té un cromodomini que reconeix la H3K27me3. El tercer complex és PhoRC, i inclou Pho (Pleiohomeotic) i dSfmbt. Pho reconeix una seqüència específica al DNA, cosa que no fa cap subunitat ni de PRC1 i 2.

El manteniment de la memòria de l'estat transcripcional per part de Polycomb és un cas d'epigenètica. En efecte, si ens atenim a una definició clàssica d'aquest terme com la de Ptashne i Gann (Ptashne and Gann, 2002) que diu que l'epigenètica són "canvis en l'estat d'expressió d'un gen que no inclouen una mutació però que són heredats a cada divisió cel·lular en absència de la senyal que va iniciar el canvi", aquest és clarament el cas de Polycomb. El silenciament induït per Polycomb (establert pels gens *Gap* i *pair-rule* als inicis del desenvolupament) es manté amb les divisions cel·lulars, es transmet a les cèl·lules filles a la mitosi.

Al genoma, els elements als quals s'uneixen els complexos Polycomb s'anomenen PRE (Polycomb Response Elements) que poden fer fins a varis centenars de parells de bases i poden estar localitzats a varies kilobases de distància del promotor del gen que silencien (Ringrose and Paro, 2007). El seu efecte silenciador no depèn de si es troba upstream o downstream del gen ni de la seva orientació. Aquests elements contenen llocs d'unió consens de diversos factors de transcripció —alguns d'ells requerits per la unió del complexos Polycomb al PRE— com GAGA, *pleiohomeotic* (PHO), *pleiohomeotic-like* (PHO-like), *zeste* o *dorsal switch protein 1* (DSP1) (Morey and Helin, 2010). D'aquests, PHO i PHO-like s'han caracteritzat com a components del grup Polycomb. Els experiments fets amb elements PRE ectòpics són capaços de mantenir la repressió sobre el transgen al llarg de les generacions, i s'ha vist per immunolocalització que recluten els complexos Polycomb al lloc d'inserció a cromosomes politènics (Cavalli and Paro, 1999).

Val a dir que s'ha vist que els PRE no sempre són diferenciables d'uns altres elements del genoma, els Trithorax Response Elements (TRE). Aquests són els elements als que s'uneixen les proteïnes del grup Trithorax, identificades en un origen per la seva capacitat de contrarestar l'efecte de Polycomb: són considerats activadors transcripcionals que transmeten l'estat transcripcional al llarg de les divisions cel·lulars (Grimaud et al., 2006b). PREs i TREs es solapen al menys parcialment i les proteïnes tant d'un grup com de l'altre coincideixen en la seva localització al genoma tal i com revelen les immunoprecipitacions de cromatina (Ringrose and Paro, 2007).

Els PRE presenten una propietat particular que s'anomena *Pairing-sensitive silencing* i que consisteix en que la presència de dues còpies homòlogues d'un transgen que conté un PRE causen un silenciament més fort que si hi hagués una sola còpia. A més s'ha observat que un PRE pot silenciar un gen situat en trans que no tingui PRE si està situat al mateix lloc del cromosoma homòleg (Li and Müller, 2011). Això suggereix un contacte físic entre aquests dos elements al nucli, consistentment amb el model de loop que explicaria com un PRE silencia un promotor a varies kilobases de distància com és el cas del PRE bxd i el gen Ubx (Kahn et al., 2006). També hi ha casos on el contacte en trans es dona entre PREs no situats al mateix lloc del cromosoma homòleg com passa amb els PRE Mcp i Fab-7 del gen Abdominal-B (Muller et al., 1999). De fet s'ha observat per immunofluorescència de proteïnes PcG que als nuclis es formen uns pocs foci, contrastant amb la gran quantitat de PREs presents al genoma, suggerint que s'agrupen en uns quants punts focalitzats. Els gens Antennapedia i Abdominal-B, separats entre ells per varies megabases, colocalitzen al nucli quan els dos estan silenciats per Polycomb però no si algun dels dos està transcripcionalment actiu (Grimaud et al., 2006a; Lanzuolo et al., 2007). Un estudi recent (Li and Müller, 2011) demostra que la capacitat dels PRE d'establir interaccions en trans no és una propietat d'aquests elements sinó que ho farien a través d'elements insulator situats en la seva proximitat i dels que es coneix bé la capacitat d'unir diferents loci del genoma.

Una pregunta que encara no té resposta clara és com es manté l'estat silenciats per Polycomb d'un gen al llarg de les divisions cel·lulars, tenint en compte que durant la mitosi els complexos PcG es dissocien de la cromatina (Voncken et al., 1999). Com s'hi re-uneixen? Com identifiquen quins són els gens que s'han de silenciar? En un inici es creia que era a través de les marques en les histones que s'associen al silenciament induït per Polycomb, és a dir que el complex s'uniria als PRE que tinguessin per exemple la lisina 27 de la H3 trimetilada reconeixent-los com a regions a silenciar (Fischle et al., 2003). No obstant algunes observacions evidencien que això no seria suficient. Per exemple, els PRE solen estar deplecionats de nucleosomes, estant les marques associades a Polycomb als voltants del PRE i no al PRE en si (Papp and Müller, 2006). Així, la metilació de les histones al voltant del PRE seria una conseqüència de la presència del complex Polycomb, i no el reclam que portaria el complex al seu lloc. A canvi, s'ha proposat un altre model en el qual Polycomb silenciaria per defecte tots els PRE llevat d'aquells que es transcriuen, als quals qui s'hi uniria és Trithorax i per tant el gen estaria activat (Ringrose and Paro, 2007). Aquest model es va proposar arran de l'observació que la transcripció a través del PRE està correlacionada amb estat activat del gen. Això significaria que tot just després de la mitosi aquells PRE que no s'han de silenciar es transcriuen i eviten la unió de Polycomb. Però llavors, què és el que porta a alguns PRE a transcriure's i altres no? Aquest és un subjecte encara en debat, tot i que un candidat proposat és la variant d'histona H3.3, que es diposita principalment a regions transcripcionalment actives i que podria ser transmesa a través de la mitosi (McKittrick et al., 2004). No està clar però si és només una conseqüència de l'activitat transcripcional o seria per ella mateixa suficient per activar la transcripció a un locus. Altres candidats a actuar com a transmissors de l'estat del PRE a través de la mitosi són proteïnes d'unió al DNA que són capaces d'accedir-hi fins i tot en cromosomes mitòtics (com s'ha demostrat per Pipsqueak o GAGA) o també estructures conformacionals alternatives de la cadena de DNA com Z-DNA (Ringrose et al., 2003). MLL (ortòleg de Tritòrax a humans) s'ha demostrat que s'uneix als promotors en cromatina mitòtica

condensada, que és necessari per la reactivació ràpida de la transcripció al finalitzar la mitosi i que és necessària per la unió d'altres membres de Tritòrax als seus llocs durant la mitosi (Blobel et al., 2009). S'ha vist per FRAP que la unió de les proteïnes PcG a la cromatina mitòtica és 300 cops més estable que a interfase, i que la fosforilació de la H3S28 podria ser el mecanisme que controla la dissociació d'aquestes proteïnes mitosi (Fonseca et al., 2012).

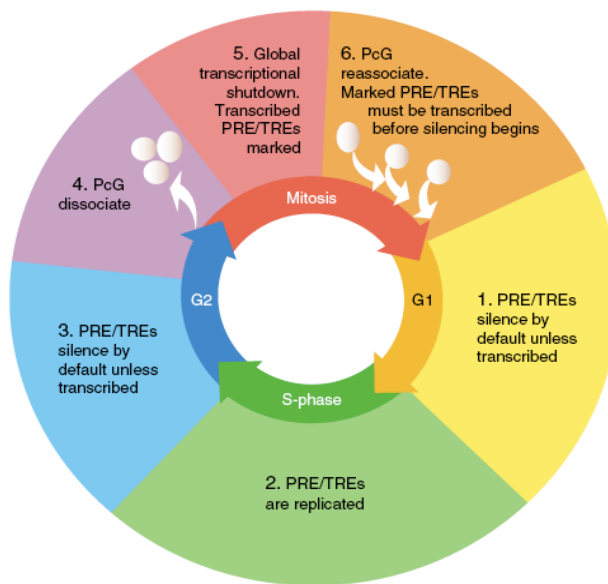


Figura 9.- Memòria epigenètica dels PRE/TRE al llarg del cicle cel·lular. 1) i 3) A la interfase, els PRE silencien per defecte, només no ho fan aquells que es transcriuen; 2) a la fase S, es repliquen amb la resta del genoma; 4) les proteïnes Polycomb i Trithorax es dissocien dels seus elements abans d'entrar en mitosi; 5) a l'entrar en mitosi, res es transcriu i d'alguna manera es marquen els PRE que no han d'estar silenciats ja que a l'acabar la mitosi (6) s'hauran de transcriure abans que les proteïnes PcG s'hi uneixin i els silencien (Ringrose and Paro, 2007).

1.2.3 Els insulators

La primera descripció d'un insulator ve del 1985, quan Udvardy, Maine i Schedl van examinar al detall el locus 87A7 de *Drosophila*, que conté els dos gens Hsp70 (Udvardy et al., 1985). Van observar que tant a una banda com a l'altra del locus hi havia uns segments de DNA particulars que eren hipersensibles a la digestió amb nucleases. Els van anomenar *specialized chromatin structure* (scs i scs') i van proposar que aquests dos elements delimitaven i aïllaven la regió de cromatina que es descondensa quan, a l'augmentar la temperatura, el locus es transcriu.

Des de llavors i fins a l'actualitat s'han anat descrivint moltes altres seqüències, no només a *Drosophila* sinó també a llevat i a vertebrats (Wallace and Felsenfeld, 2007; Bushey et al., 2008), que d'alguna manera delimiten la cromatina i separen espais discrets de regulació de l'expressió gènica. Anomenades *boundaries* (frontera o límit) o *insulators* (aïlladors), s'han definit clàssicament per presentar una de les dues següents propietats (Gerasimova and Corces, 2001):

—capacitat d'aïllar una regió de l'entorn cromosòmic (*barrier activity*): quan es troben flanquejant un transgèn l'aïllen d'efectes posicionals, com per exemple si el transgèn es troba en una regió heterocromàtica el mantenen al marge dels seus efectes repressors (Figura 10-A). Això permet l'expressió del transgèn independentment de la posició al genoma (Recillas-Targa et al., 2002; Barkess and West, 2012). Aquesta característica s'ha utilitzat per exemple en la generació de plàsmids destinats a teràpia gènica (Neff et al., 1997) evitant d'aquesta manera una possible inactivació del transgèn degut a l'entorn cromosòmic.

—activitat *enhancer-blocking*: quan en un constructe transgènic es col·loca una seqüència insulator en l'espai que hi ha entre un enhancer i el promotor d'un gen reporter, l'efecte activador de l'enhancer sobre el gen es veu

reduït o anul·lat, d'aquí el nom de *bloquejador de l'enhancer* (Figura 10-b). No silencia o desactiva l'enhancer ja que aquest encara pot activar un gen del que no estigui separat per un insulador. Aquest seria un efecte regulador que pot servir per assegurar el control temporal i espacial precís de l'acció dels enhancers sobre els gens que controlen (Vazquez and Schedl, 1994; Schweinsberg et al., 2004; Moon et al., 2005).

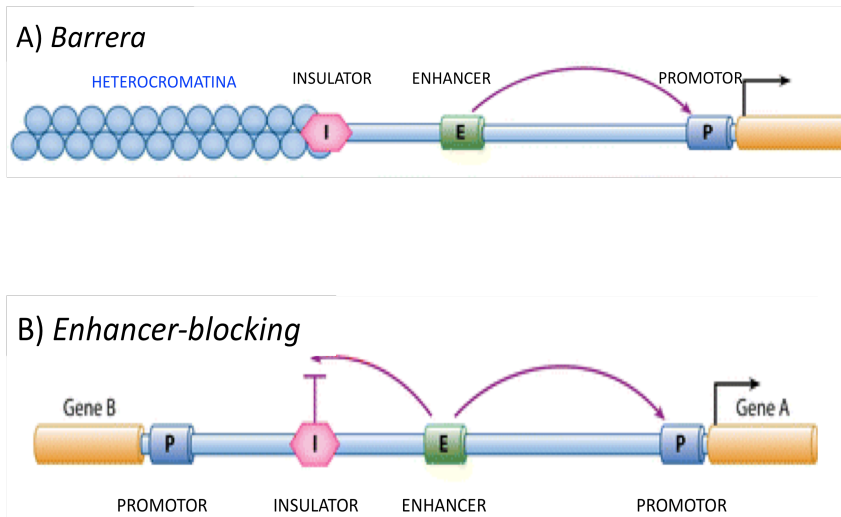


Figura 10.- Les dues propietats definitives dels insulators. A) Els insulators poden fer de barrera entre cromatina activa i cromatina silenciada. B) Un insulador situat entre un enhancer i un promotor és capaç de bloquejar l'efecte de l'enhancer (*enhancer-blocking*) (Vogelmann et al., 2011).

Aquesta última propietat suggereix que els insulators podrien tenir un paper important a l'hora d'organitzar el genoma en dominis funcionals, de manera que gens presents en un domini no es vegin afectats per seqüències reguladores d'un altre domini. De fet la definició dels insulators ha anat evolucionant amb el temps, passant de ser uns elements bloquejadors amb efectes a nivell local a tenir rellevància a nivell d'organització genòmica i cromosomal gràcies no només a la seva capacitat per separar dominis sinó també per reunir i posar en contacte diferents regions creant loops

cromosòmics, amb l'objectiu final que cada element i regió reguladora actuï de manera apropiada (Tolhuis et al., 2002; Dean, 2011; Erokhin et al., 2011).

Les seqüències insulator no tenen efecte per si soles sinó que actuen en conjunció amb unes proteïnes que les reconeixen i s'hi uneixen directament o indirectament. Mentre que a *Drosophila* se n'han descrit com a mínim sis (CP190, BEAF, Zw5, dCTCF, Su(Hw) i Mod(mdg4)), a vertebrats n'hi ha una (CTCF) i el complex de les cohesines amb qui interacciona. Cada proteïna interacciona amb una xarxa particular d'insulators normalment definida pel context i la localització genòmica, així com pel tipus cel·lular i la fase del cicle en que es trobi, podent-se solapar amb altres proteïnes al mateix insulator (Nègre et al., 2010; Schwartz et al., 2012). Les diferències funcionals i mecàniques entre els diferents tipus d'insulator en base a les proteïnes que s'hi uneixen és quelcom que encara no està clar.

La seva activitat s'associa a l'accessibilitat a la cromatina, per exemple interaccionant o reclutant factors de transcripció o factors remodeladors per posicionar, modificar o treure nucleosomes al seu voltant (Gaszner & Felsenfeld, 2006; West, Gaszner, & Felsenfeld, 2002; Emberly et al., 2008). Els insulators poden estar implicats en la compartimentalització de la cromatina interfàsica, a través d'interaccions amb compartiments del nucli com la làmina nuclear o els complexos del porus nuclear (NPC) a la perifèria nuclear, o altres estructures nuclears que regulin els estats d'activació dels gens mitjançant el posicionament de la cromatina (Guelen et al., 2008; Kalverda and Fornerod, 2010; van Bemmelen et al., 2010).

1.2.3.1 Insulators a *Drosophila*

1.2.3.1.1 scs i scs'

Són, com s'ha dit abans, els primers insulators descrits i flanquegen el locus 87A7 on hi ha gens de resposta a heat-shock (Udvardy et al., 1985). Cadascuna d'aquestes seqüències conté dos regions hipersensibles a la digestió

amb nucleases que rodegen un nucli central resistent a nucleases. L'scs i l'scs' difereixen en l'estructura de DNA i requereixen proteïnes diferents. A l'scs les seqüències hipersensibles són imprescindibles per la funció enhancer-blocking mentre que el nucli central resistent, ric en A/T és dispensable. La proteïna Zw5 s'uneix a una seqüència de 24 bp (Gaszner et al., 1999; Gerasimova and Corces, 2001) que té activitat enhancer-blocking. Mutacions en la seqüència a la que s'uneix Zw5 fan perdre la capacitat d'enhancer-blocking. Aquesta proteïna conté Zinc fingers i és essencial per la viabilitat cel·lular (Gaszner et al., 1999).

En canvi la seqüència de scs' té repeticions CGATA que interaccionen amb BEAF, proteïna responsable de l'activitat insulator de scs' (Zhao et al., 1995; Hart et al., 1997; Cuvier et al., 1998). En aquest cas mutacions en aquesta seqüència també causen la pèrdua d'activitat enhancer-blocking.

S'ha observat que BEAF i Zw5 interaccionen entre elles, i s'ha proposat un model en el que ambdues proteïnes interaccionarien portant al mateix lloc scs i scs' per tant fent un loop que aïllaria el locus 87A7 (Blanton et al., 2003).

1.2.3.1.2 Gypsy

Un altre insulator ben estudiat a *Drosophila* és el que hi ha al retrotransposó Gypsy. L'insulator fa 350 bp i està situat a la regió 5' transcrita però no traduïda upstream de l'inici de l'ORF del gag (Gdula et al., 1996). La inserció de l'insulator gypsy a les regions reguladores de gens causa fenotips mutants específics de teixit, degut a la inabilitat dels enhancers d'activar el promotor (Corces and Geyer, 1991; Holdridge and Dorsett, 1991; Jack et al., 1991). No inactiva l'enhancer ja que aquest manté la capacitat d'activar un promotor situat a l'altre costat (Cai and Levine, 1995; Scott and Geyer, 1995). També és capaç d'aïllar l'expressió d'un transgèn dels efectes de posició respecte l'entorn cromosòmic (Roseman et al., 1993) fins al punt que s'ha utilitzat per insertar transgens a espècies llunyanes com *Arabidopsis thaliana* (She et al., 2010).

Conté 12 còpies d'una seqüència de 26 bp a la que s'uneix la proteïna Su(Hw) (Suppressor of Hairy wing). La força de l'insulador depèn del nombre de repeticions del motiu bàsic de 26 bp, amb una relació lineal entre aquest nombre i la disminució de l'efecte de l'enhancer sobre el promotor (Spana et al., 1988; Scott et al., 1999). Com en els *scs* i *scs'*, l'insulador de *gypsy* conté una sèrie de tres zones hipersensibles a nucleases fortes (Gdula et al., 1996).

1.2.3.1.3 Els insulators del complex Bitòrax

Els gens homeòtics o gens Hox són uns gens molt ben conservats al llarg de l'evolució que codifiquen factors de transcripció que determinen la identitat dels segments de l'organisme durant el desenvolupament. Tots ells tenen un domini comú anomenat homeobox que és una seqüència de 180 bp que codifica la regió de la proteïna que s'unirà al DNA de manera específica. Els gens homeòtics presenten colinearitat, és a dir que l'ordre en que es disposen al llarg del cromosoma és el mateix ordre dels segments que determina cadascun. El control de l'expressió d'aquests gens a nivell temporal i espacial és molt important: una desregulació de qualsevol d'aquests gens pot portar a que un segment presenti la identitat d'un altre, fenomen conegut com a transformació homeòtica. Per tal de controlar amb precisió on i com es transcriuen, els gens homeòtics solen estar acompanyats d'intricades regions reguladores que contenen varis elements, entre ells insulators.

El complex Bitòrax és un cluster de gens homeòtics de *Drosophila* situat al cromosoma 3R que fa més de 300 Kb (Figura 11). Inclou els gens *Ultrabitorax* (*Ubx*), *Abdominal-A* (*abd-A*) i *Abdominal-B* (*Abd-B*), que són els encarregats de determinar la part posterior de l'organisme (de la meitat anterior se n'ocupen els gens del complex *Antennapedia*). La major part de les 300 Kb que el conformen són regions no codificants que contenen els elements reguladors dels gens, que ocupen l'1.4% de la seqüència del complex (Lewis, 1982; Karch et al., 1985; Martin et al., 1995).

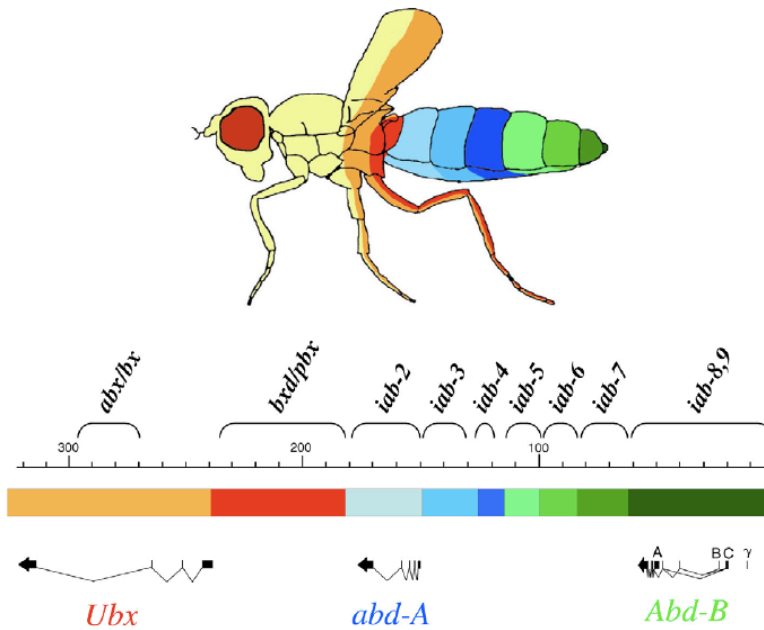


Figura 11.- El Complex Bitòrax i els segments que controla. Els tres gens homeòtics del Complex Bitòrax determinen la part posterior de l'organisme: des de la meitat posterior del tòrax fins a l'últim segment de l'abdomen. A nivell genòmic les nou regions reguladores que controlen la seva expressió estan disposades colinearment (des d'*abx/bx* fins a *iab-8,9*). Al dibuix, cadascuna d'elles es simbolitza amb el mateix color que el segment que determina (Maeda and Karch, 2006).

Els tres gens s'expressen de manera additiva al llarg de l'eix antero-posterior de l'embrió, o sigui que als segments més posteriors s'hi expressen tots tres (Celniker et al., 1989). *Ubx* es comença a expressar als segments més posteriors del tòrax, *abd-A* als primers segments de l'abdomen i *Abd-B* a la regió més posterior de l'abdomen. Estan regulats de manera acurada per nou dominis reguladors disposats en l'ordre dels segments que controlen. Els dominis *abx/bx* i *bxd/pbx* són responsables de l'expressió d'*Ubx* als para-segments 5 i 6 respectivament, que corresponen als segments T3 i A1. Els dominis *iab-2* a *iab-4* regulen *abd-A* als para-segments de 7 a 9, corresponent als segments A2, A3 i A4. Finalment els dominis *iab-5* a *iab-8,9* regulen l'expressió d'*Abd-B* als para-segments 10-14, que corresponen als segments abdominals A5-A9 (Boulet et al., 1991; Maeda and Karch, 2006).

La regulació d'Abd-B és la més complexa i la que més s'ha estudiat, està sota el control de quatre dominis que ocupen més de 100 Kb. Aquests quatre dominis s'anomenen iab (infra-abdominals) i estan separats entre ells per les regions Fab (Front-abdominals). A més, cada domini conté un enhancer (IAB) que activa Abd-B de manera específica al para-segment que li correspon. Els elements Fab s'encarreguen d'evitar que un enhancer actuï en un para-segment incorrecte, això ho fan delimitant cada domini.

A l'inici del desenvolupament els productes dels gens Gap i Pair-rule s'uneixen a uns elements anomenats iniciadors que es troben a cada iab i que actuen com a enhancers conferint l'expressió corresponent a Abd-B segons el segment on es trobi la cèl·lula. Quan acaba aquesta fase d'iniciació, els nivells de proteïnes gap i pair-rule decauen, i llavors qui s'ocupa de mantenir l'estat de cada iab segons la seva identitat són els anomenats elements de manteniment, encarregats de fer perdurar la informació posicional que els Gap i pair-rule han donat a l'inici del desenvolupament. Aquests elements de manteniment solen ser PREs, que com s'ha explicat abans també actuen com a TREs. La capacitat dels PREs de mantenir el seu estat (activat o reprimat) al llarg de les divisions cel·lulars sense alteració és el que permet que la identitat dels segments de l'organisme es mantingui immutable durant totes les etapes del desenvolupament. Així, si un domini està silenciats per Polycomb, l'enhancer o els elements que hi estiguin continguts no tindran efecte sobre l'expressió d'Abd-B (Mihaly et al., 1998; Maeda and Karch, 2006).

Per tal que el silenciament o l'activació d'un domini determinat pels seus elements de manteniment no s'extengui als dominis adjacents existeixen els Fab que actuen de barrera o bloquejador. En efecte es tracta d'insulators, i el seu correcte funcionament és crucial perquè no hi hagin errors d'identitat entre segments. Els Fab descrits clàssicament en la literatura són l'Mcp, el Fab-7 i el Fab-8 (Karch et al., 1994; Cavalli and Paro, 1998; Barges et al., 2000).

1.2.3.1.3.1 Fab-7

El Fab-7 és el Fab més ben conegut i un dels insulators més estudiats de *Drosophila* (Zhou et al., 1996; Schweinsberg et al., 2004). Es troba separant els iab-6 i iab-7, que controlen els parasegments 11 i 12 respectivament. Es va identificar inicialment degut a que la deleció Fab-7¹ a la que li faltava tot el Fab-7 sencer mostrava una transformació homeòtica on el PS11 es presentava com una duplicació del PS12. Els anàlisis genètics mostraven com els dominis iab-6 i iab-7 s'havien fusionat en un de sol i que això portava a una activació ectòpica de l'iab-7 al PS11 per part dels elements activadors de l'iab-6. Normalment l'iab-7 es troba silenciada a tots els parasegments anteriors al PS12. Es va veure que la deleció Fab-7¹ no només eliminava el boundary sinó també el PRE adjacent (Karch et al., 1994; Hagstrom et al., 1996).

Les delecions en les que només s'elimina el boundary i no el PRE difereixen amb Fab-7¹ en que la transformació homeòtica és més complexa: el PS11 és una mescla de guany de funció (transformació a PS12) i pèrdua de funció (transformació a PS10). En el primer cas es deu al mateix motiu que ja s'ha explicat: els elements activadors de l'iab-6 s'extenen a l'iab-7 al PS11 i per tant té fenotip de PS12. En canvi en el segon cas, la pèrdua de funció, el que passa és que com que el PRE de l'iab-7 encara és funcional i ja no hi ha un boundary que en delimiti el radi d'acció, el silenciament de l'iab-7 al PS11 es transmet a tot l'iab-6 (Maeda and Karch, 2006).

L'estructura de la cromatina del Fab-7 mostra tres regions hipersensibles a la digestió amb nucleases (HS1, HS2 i HS3) i una quarta de menor (HS4). Les grandàries d'aquestes regions van des de 400 bp (HS1) fins a la més petita, de 140 bp. Estan separades entre elles per regions resistents a les nucleases del tamany aproximadament d'un nucleosoma. Anàlisis de delecions i amb constructes transgènics mostren que la HS3 és on hi ha el PRE, mentre que el boundary ocuparia la regió d'1.2 Kb que va del HS1 a l'HS2, i hi ha controvèrsia

sobre la participació o no del lloc menor HS4 en la funció boundary (Hagstrom et al., 1996; Zhou et al., 1996).

1.2.3.1.3.2 Fab-8

L'element Fab-8 marca l'extrem de l'iab-7 (PS12) i el separa de l'iab-8 (PS13). A nivell d'estructura de la cromatina, conté dos regions hipersensibles a la digestió amb nucleases: HS1 d'uns 400 bp; i HS2 d'uns 150 bp (Barges et al., 2000). Es va demostrar que HS1 fa funció d'insulator mitjançant assaigs d'enhancer-blocking en constructes transgènics, mentre que HS2, que a més conté varis llocs d'unió de la proteïna *pleiohomeotic*, és un PRE. L'insulator impedeix que el silenciament que promou el PRE a l'iab-8 s'extengui a l'iab-7 i també el silenciï.

La deleció Fab-8⁴¹⁶ és una deleció en la que només falta l'insulator però no el PRE (Barges et al., 2000). En aquestes mosques s'observen transformacions homeòtiques tant de A7 a A8 (guany de funció) com de A7 a A6 (pèrdua de funció). En el cas del guany de funció els elements activadors de l'iab-7 s'extenen fins a l'iab-8 i per tant l'activen al PS12 i aquest presenta fenotip A8 quan normalment seria A7. En el cas de la pèrdua de funció el que passa és que l'efecte silenciador del PRE-8 al PS12 s'extén fins a l'iab-7 i aquest mostra un fenotip de PS11 perquè s'inactiva. És a dir, que al faltar l'insulator hi ha una fusió entre l'iab-7 i l'iab-8 i llavors s'estableix una competició a l'iab-7 al PS12 entre els elements activadors propis que es poden estendre fins a l'iab-8, i els elements silenciadors de l'iab-8. En canvi la deleció Fab-8⁶⁴ s'endua tant l'insulator com el PRE. En aquest cas només s'observa un fenotip de guany de funció de PS12 a PS13, perquè l'iab-8 perd la capacitat de silenciar i no hi pot haver doncs transformació A7 a A6 (Barges et al., 2000).

S'ha observat que la proteïna d'insulator dCTCF s'uneix al Fab-8 a través de varies seqüències consens que hi són presents. S'ha demostrat la necessitat pel correcte funcionament enhancer-blocking del Fab-8 la presència de dCTCF

tant amb constructes en embrions (Moon et al., 2005) com en assaigs amb cèl·lules S2 (Ciavatta et al., 2007).

1.2.3.1.3.3 Mcp

L'element Mcp (Miscadastral Pigmentation) es troba entre l'iab-4 i l'iab-5. El fenotip de la seva deleció és una transformació homeòtica de guany de funció on el PS9 apareix amb el fenotip del PS10 (Karch et al., 1994; Muller et al., 1999). És a dir, al PS9, on normalment només l'iab-4 està actiu, ho està també l'iab-5. De fet al PS9 wild-type no hi ha transcripció d'Abd-B, mentre que sí que es transcriu quan hi ha deleció d'Mcp. Això es pot interpretar com que, al perdre's la funció insulator que separava els dos dominis, els elements activadors de l'iab-4 al PS9 s'extenen fins a l'iab-5 i per tant es comporta com si fos el PS10. A la vegada també suggereix que aquesta deleció d'enduu, si n'hi ha, el PRE adjacent a l'insulator.

En efecte s'ha vist que l'insulator i el PRE estan molt junts, però són distingibles. El primer en ser identificat va ser el PRE (Busturia et al., 2001) que és una seqüència de 138 bp que conté llocs d'unió per a GAGA i Pleiohomeotic i que són essencials pel seu funcionament. Posteriorment es va identificar la seqüència que feia d'insulator (Gruzdeva et al., 2005), de 755 bp, que era capaç no només de funcionar com a enhancer-blocking sinó també de bloquejar el silenciament induït per un PRE adjacent.

1.2.3.1.3.4 Organització i funcionament de la regió reguladora d'Abdominal-B

Els insulators presents a cada Fab no només aïllen cada iab de la resta per tal que no es vegin afectats per efectes activadors o silenciadors aliens, sinó que també s'ha vist que organitzen l'estructura de cromatina de tota la regió reguladora (Cléard et al., 2006; Mihaly et al., 2006). Això ho fan a través de contactes llunyans entre elements (Muller et al., 1999) i amb la regió promotora d'Abdominal-B (Kyrchanova et al., 2008). Així aconsegueixen d'una banda formar

loops que deixin fora o aïllin dominis sencers, i d'altra banda posar en contacte elements presents al domini amb el promotor del gen que d'aquesta manera n'activin l'expressió de manera correcta (Sipos and Gyurkovics, 2005). Tot i que s'ha proposat un rol clau dels PRE del Bithorax a l'hora de mediar contactes intra-cromosomals (Lanzuolo et al., 2007) però anàlisis funcionals i d'interacció mostraven que els contactes són independents dels PREs (Li and Müller, 2011).

S'ha demostrat per exemple que el Fab-7 i el Fab-8 estan en contacte entre ells i amb el promotor d'Abd-B (Cléard et al., 2006), i que dCTCF és essencial per a que es produeixin aquests contactes (Kyrchanova et al., 2008). Així sorgeix la idea que per a aquesta organització de la cromatina a través de contactes entre diferents elements hi juguen un paper fonamental les proteïnes d'unió a insulators, que d'alguna manera són elles qui medien aquests enllaços. En efecte tant a Mcp, Fab-7, Fab-8 com a la regió promotora d'Abd-B s'hi ha detectat per CHIP presència de proteïnes d'insulator com CP190 o dCTCF (Moon et al., 2005; Holohan et al., 2007; Mohan et al., 2007).

Com hem explicat, els insulators tenen activitat enhancer-blocking, són per tant bloquejadors del contacte entre un enhancer i un promotor si s'hi troben entremig. Com s'ho fan doncs els enhancers més distals de la regió reguladora d'Abdominal-B per atènyer el promotor si hi ha insulators entremig? S'han descrit uns altres elements presents en la proximitat del Fab-7 i del Fab-8 que s'anomenen PTS o Promoter Targeting Sequence (Zhou and Levine, 1999; Chen et al., 2005) .

Aquests elements serien capaços de contactar amb una altra seqüència que es troba prop del promotor (PTE o Promoter Targeting Element) sobrepasant l'efecte bloquejador dels insulators. O sigui que aquests elements també contribuirien a la correcta organització de la cromatina de la regió per tal que cada iab es posi en contacte amb el promotor sense interferir amb el silenciament d'altres iab (Lin et al., 2010).

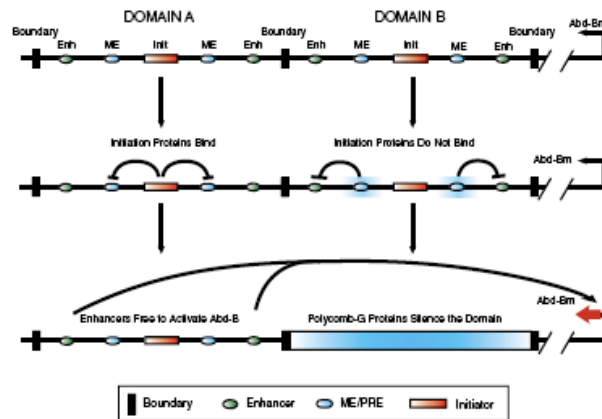


Figura 12.- El model dels Dominis. Cada domini funciona autònomament per dirigir l'expressió d'Abdominal-B. L'estat de cada domini ve determinat en un inici pels elements iniciadors, a qui li prenen el relleu els elements de manteniment (PRE/TRE). Quan un domini està silenciats, els insulators (boundaries) impedeixen que es silencii el domini adjacent (Mihaly et al., 2006).

Totes les dades existents sobre el funcionament de la regió reguladora d'Abdominal-B han estat sumariades en l'anomenat Model dels Dominis (Mihaly et al., 2006), segons el qual cada domini controla autònomament l'expressió de d'Abd-B en un para-segment (Figura 12). Cadascun d'aquests dominis és un conjunt modular dels elements necessaris per controlar aquesta expressió, que comença a l'inici del desenvolupament com a resposta a les proteïnes Gap i pair-rule que s'uneixen als elements iniciadors, que transmeten la senyal d'activació a la resta d'elements del domini. Entre aquests hi ha els elements de manteniment que s'encarreguen de fer perdurar la senyal més enllà de la decaiguda dels Gap i pair-rule. Si no hi ha hagut unió dels Gap i pair-rule el domini es queda silenciats i no afecta l'expressió del gen. Per tal que aquest silenciament no s'extengui al domini adjacent hi ha els elements Fab, que juntament amb els PTS posaran en contacte el domini amb el promotor d'Abdominal-B.

1.2.3.2 Proteïnes d'insulator

1.2.3.2.1 Su(Hw)

Suppressor of Hairy-wing és una proteïna de 109 KDa que conté 12 zinc fingers i que s'uneix directament al DNA reconeixent una seqüència consens GCATACTTTT (Adryan et al., 2007). Es va identificar en un principi perquè una deleció al gen que la codifica suprimeix els fenotips causats per una inserció del retrotransposó Gypsy (Spana et al., 1988). Es va descobrir que s'uneix a banda i banda del retrotransposó, en el que es va identificar com a insulator de Gypsy, i es va veure que és capaç de conferir aïllament respecte dels efectes posicionals de la inserció al genoma. De la mateixa manera, és necessària perquè els insulators de Gypsy aïllin d'efectes posicionals un transgen amb un mini-white (efecte barrera) i també pel bloqueig de l'acció d'un enhancer sobre un promotor si hi ha l'insulator de Gypsy enmig (Roseman et al., 1993) i que aquest enhancer continua actiu podent activar un gen que estigui a l'altra banda (Scott and Geyer, 1995). Es va veure també que s'uneix a centenars de llocs a cromosomes politènics, el que suggeria que a part d'aïllar el retrotransposó Gypsy també tenia altres llocs d'unió al genoma amb altres funcions (Harrison et al., 1989).

Mitjançant immunoprecipitacions de cromatina s'ha vist que s'uneix al genoma preferentment a regions intergèniques, tot i que també es troba en part a gens, sobretot a introns (Ramos et al., 2006; Adryan et al., 2007). Tot i trobar-se majoritàriament a regions intergèniques, en mutants de Su(Hw) s'observen greus alteracions de l'expressió gènica de manera generalitzada. Su(Hw) s'uneix a l'insulator de gypsy juntament amb una altra proteïna, Mod(mdg4) amb qui interacciona directament a través d'un domini Leucine zipper i amb qui coincideixen pràcticament a totes les bandes de politènics. El domini BTB de Mod(mdg4) és essencial per la funció insulator (Golovnin et al., 2007). Totes dues formen insulator bodies, els corpuscles nuclears que s'observen quan es fa immunotinció (Ghosh et al., 2001).

De fet s'han detectat interaccions entre Su(Hw) i elements de la làmina nuclear (Van Bommel et al., 2010) i del NPC (Nuclear Pore Complex) i per això s'ha suggerit que Su(Hw) crea interaccions amb el nucli que ajuden a organitzar i particionar el genoma (Kalverda and Fornerod, 2010). Tant Su(Hw) com Mod(mdg4) es troben en extractes de la matriu nuclear (Byrd and Corces, 2003).

1.2.3.2.2 BEAF

La proteïna BEAF-32 (Boundary Element Associated Factor 32KDa) va ser identificada per la seva unió a l'insulator scs' de l'hsp70 al locus 87A7. S'uneix a la seqüència CGATA i a cromosomes politènics forma centenars de bandes que coincideixen amb les interbandes del DAPI. En constructes transgènics es mostra necessària per l'activitat enhancer-blocking (Hart et al., 1997).

Es van intentar caracteritzar més proteïnes que s'uneixen als llocs on s'uneix BEAF, i es va trobar que DREF s'uneix a la mateixa seqüència consens. DREF és un regulador transcripcional de gens involucrats en la fase S del cicle cel·lular, quan es produeix la replicació. Es va demostrar que DREF i BEAF competeixen pel mateix lloc d'unió, tot i que DREF no tingui funció de proteïna d'insulator (Hart et al., 1999). De fet tant una com l'altra tenen el mateix domini d'unió al DNA, un tipus especial de zinc finger (BED zinc finger) que sembla provenir evolutivament de proteïnes de transposons. Aquest zinc finger va ser precisament descrit a partir de BEAF i DREF i d'aquí el nom (Aravind, 2000).

Anàlisis a nivell genòmic dels llocs on s'uneix BEAF mostren un enriquiment prop de gens reguladors del cicle cel·lular i de la segregació. En mutants de BEAF s'observen defectes del cicle cel·lular i de la segregació, i que els nivells d'expressió d'aquests gens es veuen alterats. Es creu que BEAF els regula a través de l'impediment de la disposició en aquests llocs de la metilació de la H3K9 (Emberly et al., 2008).

Per CHIP on ChIP s'han detectat fins a 1820 llocs d'unió de BEAF al genoma, dels quals el 85% es trobaven a menys de 300 pb de distància d'un TSS (Jiang et al., 2009) cosa que d'entrada representa una diferència substancial amb el que s'observa a una altra proteïna d'insulator com Su(Hw) que està majoritàriament a regions intergèniques. A la majoria s'aquests gens on hi ha unió de BEAF hi ha també dimetilació de la H3K4, presència d'RNA Pol II i de la variant d'histona H3.3. Això és coherent amb l'observació que BEAF es troba a les interbandes dels cromosomes politènics, suggerint que té forta predilecció pels gens actius, o dit d'una altra manera, forma part dels elements que participen de l'activació gènica. Això es corrobora en mutants de BEAF, on la majoria d'aquests gens veuen disminuïda la seva transcripció (Jiang et al., 2009).

Una peculiaritat de BEAF és que mostra un fort enriquiment a als anomenats *head-to-head genes*, aquells gens que tenen un TSS pròxim —menys d'1Kb de distància— i que es transcriuen en direccions oposades (Yang et al., 2012). A més, aquells gens head-to-head on hi ha BEAF tenen tendència a no estar co-regulats, a diferència de molts head-to-head sense BEAF. El genoma de *Drosophila* és particularment dens i la quantitat de head-to-head es molt elevada en relació als vertebrats (Yang and Yu, 2009). El fet que a vertebrats no hi hagi homòleg de BEAF ha fet suggerir la hipòtesi que BEAF ha sorgit com una proteïna necessària a *Drosophila* per controlar aquesta alta densitat gènica i permetre l'expressió diferencial de gens head-to-head que comparteixen o tenen molt a prop les regions reguladores (Yang et al., 2012).

1.2.3.2.3 CTCF

CTCF va ser inicialment descoberta en vertebrats quan s'estudiaven proteïnes que s'unien a les regions hipersensibles a nucleases de la seqüència flanquejant el gen *c-myc* de pollastre (Lobanenkov et al., 1990). Mitjançant anàlisis de gels de retard es va veure que s'unia preferencialment a la seqüència CCCTC, d'aquí el seu nom CCCTC-binding factor, CTCF. D'alguna manera l'expressió de *c-myc* es veia alterada quan faltava CTCF. Posteriorment ja va ser

caracteritzada com a proteïna d'insulator, comprovant la seva necessitat per l'activitat enhancer-blocking de varis insulators de vertebrats (Bell et al., 1999).

No va ser fins uns anys més tard que se'n va identificar un homòleg a *Drosophila*, anomenat dCTCF o simplement CTCF (Moon et al., 2005). Com a vertebrats, la CTCF de *Drosophila* té 11 zinc fingers amb els que pot reconèixer la mateixa seqüència de DNA que el seu homòleg. Es va comprovar que té activitat enhancer-blocking, essent essencial pel funcionament del Fab-8. Mitjançant immunotincions de cromosomes politènics s'ha vist que colocalitza extensament amb CP190 però poc amb Su(Hw) o Mod(mdg4).

Juntament amb CP190, CTCF es troba unit a sis dels insulators que separen els dominis del Complex Bitòrax. La manca de CTCF causa alteracions en el patró d'expressió d'Abdominal-B i en conseqüència transformacions homeòtiques (Mohan et al., 2007). De fet s'ha arribat a proposar que a l'emergència dels metazous bilaterals (Bilateria) —que van patir la gran diversificació durant l'explosió càmbrica— hi va jugar un paper essencial CTCF, sense la qual l'organització dels gens homeòtics no hagués tingut lloc (Heger et al., 2012). En efecte CTCF és present a la major part de clusters de gens homeòtics dels llinatges que en tenen, tenint el paper d'organitzador de les regions reguladores que els controlen. Anàlisis de CHIP on ChIP mostren que tot i que es troba molt present a regions intergèniques també es mostra present amb freqüència a les immediacions de TSS (Smith et al., 2009). Així, es situa en una situació a mig camí entre el cas de BEAF, molt gènic, i Su(Hw), clarament intergènic. D'aquesta manera s'ha proposat que CTCF sigui un tercer tipus d'insulator diferenciat dels altres dos (Gerasimova et al., 2007; Bushey et al., 2009). CTCF s'ha trobat particularment enriquit als extrems de dominis silenciats que contenen la marca H3K27 trimetilada, i una depleció de CTCF produeix disminucions d'aquesta marca a aquests dominis, recalcant la importància que té a l'hora de mantenir els dominis de Polycomb, tal i com succeeix als iab del Bitòrax (Van Bortle et al., 2012).

1.2.3.2.4 CP190

CP190 (Centrosomal Protein 190) és una proteïna de 1096 aminoàcids que pesa 190 KDa, que té motius Zinc finger, un motiu BTB, una senyal de localització nuclear i una senyal de localització al centrosoma (Oegema et al., 1995; Whitfield et al., 1995). Precisament va ser identificada per primera vegada a causa de la seva presència al centrosomes, com a resultat d'una purificació de complexos proteics d'extracte d'embrions de *Drosophila* que s'unissin a una columna d'afinitat de microtúbuls (Kellogg and Alberts, 1992).

Durant la interfase, CP190 es troba dins dels nuclis. Quan la cèl·lula entra en mitosi i hi ha el trencament de l'embolcall nuclear localitza als centrosomes, fins la telofase, quan torna al nucli (Oegema et al., 1997). Al nucli s'uneix a cromosomes politènics a un gran nombre de loci (Whitfield et al., 1995). No sembla que tingui cap rol essencial als centrosomes ja que en mutants de CP190 no veu alterada ni l'organització dels microtúbuls, ni la formació de centrosomes ni la mitosi en general. És més, la letalitat causada per mutants nuls de CP190 pot ser rescatada per una forma truncada de la proteïna que és incapaç de localitzar als centrosomes, suggerint que la funció essencial de CP190 és la que es dona al nucli durant la interfase, restant una incògnita què hi fa als centrosomes (Butcher et al., 2004).

No obstant, tot i que la mitosi es vegi inalterada en mutants de CP190, sí que s'han observat defectes en l'expansió axial, un procés de distribució dels nuclis dins l'embrió de *Drosophila* al llarg de l'eix antero-posterior que és dependent d'actina i miosina. Més concretament s'ha vist que la manca de CP190 afecta a l'organització de la miosina restant inalterada l'actina. El mateix mutant de CP190 que no localitza a centrosomes però sí al nucli no pot rescatar aquest defecte, suggerint que hi ha alguna connexió entre la presència de CP190 al centrosoma i la formació de l'esquelet d'actina/miosina als embrions (Chodagam et al., 2005).

La primera indicació que CP190 podria estar participant de la funció insulator al nucli ve de la troballa que forma complex amb Su(Hw) i Mod(mgd4) a l'insulator de Gypsy, i que a més n'és essencial pel correcte funcionament d'aquest (Pai et al., 2004). Es va trobar que colocalitzava extensament amb aquestes dues proteïnes a cromosomes politènics i també als insulator bodies, que desapareixien en mutants de CP190. Tot i que no s'unia directament al DNA de l'insulator de Gypsy, sí que ho feia a un altre insulator com el del locus *yellow-achaete*. Es va proposar un model en que totes tres proteïnes juntes eren capaces de formar ultra-estructures de la cromatina com loops per tal s'aïllar regions.

Posteriorment es va veure que CP190 també s'associa amb CTCF i colocalitzen tant a cromosomes politènics com als insulators del complex bitòrax (Gerasimova et al., 2007; Mohan et al., 2007). A més, CTCF també es troba als mateixos insulator bodies en els que colocalitzen Cp190, Su(Hw) i Mod(mdg4). CP190 és essencial per, tal i com passa amb CTCF, l'activitat enhancer-blocking del Fab-8. S'ha vist que la unió de CP190 a la cromatina no només canvia segons el tipus cel·lular sinó que pot canviar com a resposta a estímuls externs com heat-shocks o la recepció d'eccidisona (Wood et al., 2011).

Un anàlisi a tot el genoma de la localització de CP190 utilitzant CHIP on CHIP mostrava que més de la meitat dels llocs d'unió es trobaven a regions gèniques i una gran proporció d'aquests a les proximitats de TSS de gens actius (Bartkuhn et al., 2009). D'aquests llocs un 23% coincideixen amb llocs d'unió de CTCF. Tant els llocs prop de TSS com els intergènics mostren una clara depleció de nucleosomes. Dels llocs compartits amb CTCF hi ha un fort enriquiment en límits de regions riques en H3K27 trimetilada.

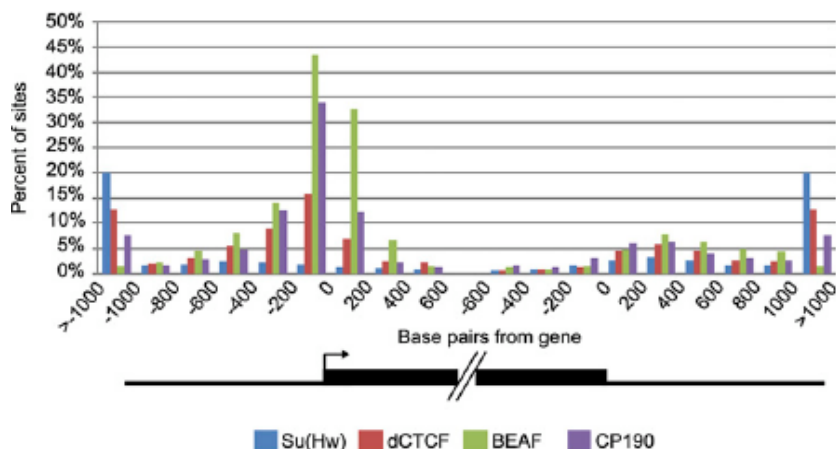
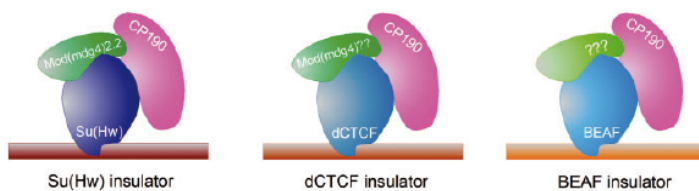


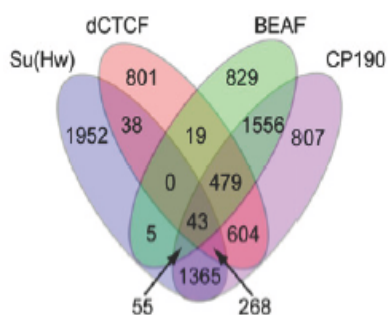
Figura 13.- Localització al genoma de les proteïnes d'insulator en relació als 5' i 3' dels gens anotats. En fragments de 200 pb es mostra una normalització dels gens mostrant 1Kb amunt i avall del gen i 600 pb a partir i fins als inicis i final de gens. Su(Hw) és la que més presència té a regions intergèniques, mentre que BEAF la més sobrerrepresentada al 5' dels gens i CTCF té un paper intermig (Bushey et al., 2009).

A més de les colocalitzacions a politèncics de CP190 amb Su(Hw) i CTCF (però no entre elles) s'ha descrit la forta colocalització amb BEAF, que té molta més tendència a unir-se a TSS i per tant tampoc colocalitza fortament ni amb CTCF ni amb Su(Hw). Així s'ha proposat un model (Bushey et al., 2009) en el que CP190 seria la proteïna en comú als tres tipus d'insulators —BEAF, CTCF i Su(Hw)— cadascun amb la seva xarxa de seqüències d'unió, però totes tres amb un mecanisme de funcionament similar que implica formació d'estructures de cromatina i organització dins del nucli dels cromosomes i que involucraria a CP190 en un paper central en la formació d'aquestes xarxes. Aquest paper passaria per la interacció de CP190 amb totes tres a través del seu domini BTB (Oliver et al., 2010) i també la formació de dímers de CP190 que permetrien l'agrupament en complexos d'insulators distants (Bonchuk et al., 2011).

A)



B)



C)

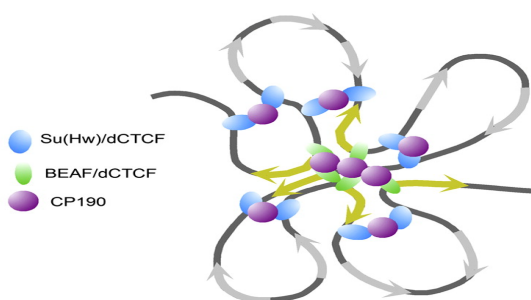


Figura 14.- La hipòtesi de les tres sub-classes d'insulator. A) CP190 seria una proteïna comú a tots tres tipus, cadascun d'ells determinat per la proteïna que s'uneix al DNA: CTCF, Su(Hw) o BEAF. Mod (mdg4) també en formaria part en el cas de Su(Hw) i potser de CTCF B) Diagrama de Venn de localització totes quatre proteïnes. Totes tenen una part compartida amb CP190 i una part pròpia, però no compartida entre elles. C) Loops de DNA causats per interaccions de proteïnes d'insulator, amb CP190 fent de pont entre elles (Bushey et al., 2009).

1.2.3.2.5 Models d'interacció entre diferents tipus d'insulator

Darrerament gràcies als avenços en tècniques genome-wide s'han intensificat els estudis de posicionament i combinatòria de les proteïnes d'insulator al genoma, aportant noves dades i hipòtesis sobre la seva funció i classificació (Nègre et al., 2010; Hou et al., 2012; Schwartz et al., 2012; Van Bortle et al., 2012). Partint de la hipòtesi de les tres sub-classes d'insulator amb CP190 com a peça comú (Bushey et al., 2009), Nègre i col·laboradors (Nègre et al., 2010) han proposat una classificació alternativa en la que només hi hauria dos tipus diferents d'insulators: la Classe I serien aquells que es troben enriquits a promotors, introns, 5'UTRs, finals de transcripció de gens i també regions intergèniques (CP190, CTCF, BEAF); i la classe II serien tots els insulators associats a Su(Hw), molt més presents a regions intergèniques que la Classe I. La Classe I tindria algun rol en els promotors adjacents divergents, on les seves proteïnes s'hi troben enriquides, i totes dues Classes es troben enriquides en promotors alternatius. L'altra classificació alternativa proposada (Schwartz et al., 2012) defensa que l'agrupació en dues o tres classes implica un llinar de discriminació arbitrari entre propers o no a TSS i postula que enlloc de simplificar s'han de tenir en compte totes les combinacions que es troben al genoma (no totes les combinacions hi tenen lloc) i agrupen els insulators en setze classes.

A més en aquest estudi es remarca una significativa correlació entre presència d'insulator i depleció de nucleosomes, així com una forta presència als límits de regions enriquides en H3K27 trimetilada. Aquesta última observació es veu també verificada en l'estudi de Van Bortle i col·laboradors (Van Bortle et al., 2012), en la que es proposa que els insulators són fonamentals per mantenir restringits es dominis silenciats de H3K27 trimetilada, observant que en RNAi de proteïnes d'insulator els nivells d'aquesta marca d'histones es veu reduïda en els dominis. A l'estudi de Schwartz i col·laboradors (Schwartz et al., 2012) s'observa que la depleció de proteïnes d'insulator no porta a grans canvis d'expressió gènica utilitzant RNAi en cèl·lules S2.

1.2.3.3 Els insulators des d'una perspectiva evolutiva

Amb l'excepció de CTCF, totes les proteïnes d'insulator descrites fins a l'actualitat es troben només a insectes. Mentre que la presència d'homòlegs de CP190 i Su(Hw) està repartida entre tots els insectes, Mod(mgd4) no ha estat trobat més enllà de Drosòfílids i mosquits i BEAF només té homòlegs a *Drosophila* (Schoborg and Labrador, 2010). El cas de BEAF ha estat extensament estudiat perquè la seva presència es correlaciona amb una característica de *Drosophila* com és l'alta densitat gènica. La presència d'aquesta proteïna a les regions reguladores que hi ha entre gens pròxims i divergents, molt freqüents a *Drosophila*, ha suscitat una possible explicació evolutiva de la seva conservació a *Drosophila* però no en altres espècies (Yang et al., 2012).

CTCF és una proteïna molt conservada: no només es troba a *Drosophila* i a vertebrats, sinó a tot el grup de metazous Bilaterals, tant deuterostomats com protostomats, amb algunes excepcions com alguns nematodes o els platihelminths. O sigui que l'origen de CTCF es troba com a mínim a l'ancestre comú de deuterostomats i protostomats (Heger et al., 2012). També les seqüències d'unió de CTCF es troben ben conservades. El fet que per una banda CTCF és una peça clau de la regulació de gens homeòtics a vertebrats i a *Drosophila*, i per altra la forta correlació entre la presència de CTCF i clusters de gens homeòtics ha fet proposar als autors (Heger et al., 2012) que CTCF va ser essencial per a l'emergència dels Bilaterals durant l'explosió càmbrica ja que sense aquesta proteïna no s'haguessin pogut regular les separacions entre dominis característiques dels clusters de gens Hox.

1.2.3.3.1 CTCF i els insulators a vertebrats

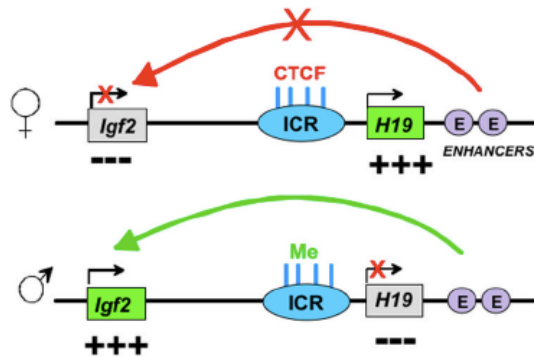
El primer insulator descrit a vertebrats es troba al locus de la beta-globina de pollastre (Chung et al., 1993). A 5' del gen hi ha un lloc hipersensible a nucleases (HS4) que no només ha demostrat tenir activitat enhancer-blocking sinó que a més protegeix el gen de la beta-globina d'una regió silenciada adjacent. Fa de barrera entre una zona hiperacetilada, hipersensible a DNAsa i

una zona hipoacetilada i resistent a DNAsa. Aquest insulador conté un lloc d'unió de CTCF, essent CTCF suficient per conferir activitat enhancer-blocking (Recillas-Targa et al., 2002). També hi ha un altre insulador amb unió de CTCF al cantó 3' del gen, suggerint la idea que els dos insulators flanquegen el locus de manera a preservar l'estat de cromatina obert i la correcta expressió del gen.

Un altre exemple ben estudiat i on el rol de CTCF és essencial és el locus *Igf2/H19* de ratolí (Arney et al., 2006). En aquest locus hi ha dos gens activats pel mateix enhancer però que tenen empremta genètica: l'*Igf2* s'expressarà de la còpia paterna mentre que *H19* ho farà de la materna. El gen *Igf2* potencia el creixement fetal i placentari, mentre que *H19*, que genera un RNA no codificant, retarda aquest creixement. Hi ha un insulador que bloqueja l'activació de l'enhancer sobre el promotor de *Igf2* només en la còpia materna (Figura 15-a). Aquest insulador té llocs d'unió de CTCF que són fonamentals: si s'eliminen hi ha expressió bial·lèlica de *Igf2* perquè l'insulador deixa de funcionar a la còpia materna. Per tal que l'insulador no funcioni de manera normal al cromosoma patern aquest es troba metilat, de manera que CTCF no s'hi pot unir i l'enhancer activa l'*Igf2* (Yang and Corces, 2011). En humans el funcionament d'aquest locus és el mateix, i el malfuncionament de l'insulador pot portar al síndrome de Beckwith-Wiedeman, on només s'expressa *Igf2* i no *H19* ja que presenta microdelecions del lloc d'unió de CTCF. Aquest síndrome provoca sobre-creixement del fetus i alt risc de formació de tumors (Herold et al., 2012).

El funcionament de CTCF al locus *H19/Igf2* implica la formació de loops de cromatina (Figura 15-b). Un candidat a mediar aquests loops junt amb CTCF és el complex de la cohesina (Wendt et al., 2008). De fet, d'estudis genome-wide s'ha observat una extensa colocalització entre els dos. S'ha demostrat que la cohesina s'uneix a CTCF i és necessària per formar interaccions de cromatina a distància. El model és que CTCF s'uneix al DNA específicament i la cohesina, que està a més implicada en cohesió de les cromàtides germanes, confereix capacitat d'unió entre diferents loci (Gause et al., 2008; Parelho et al., 2008; Rubio et al., 2008; Wendt et al., 2008).

A)



B)



Figura 15.- Regulació del locus *Igf2/H19*. A) A l'al·lel matern CTCF impedeix l'activació d'*Igf2*, mentre que al patern la metilació de l'insulator no permet la unió de CTCF i *Igf2* es pot transcriure. B) Model de funcionament de l'insulator a partir d'experiments 3C. L'insulator contacta físicament amb l'enhancer impedit l'activació d'*Igf2* en l'al·lel matern només si hi ha CTCF (Wallace and Felsenfeld, 2007).

La mutació de CTCF és letal a embrions de ratolí, i multitud d'estudis n'han demostrat la rellevància en molts processos: desenvolupament embrionari, desenvolupament neuronal, diferenciació hematopoiètica, desenvolupament de la retina, regulació de microRNAs, etc (Wallace and Felsenfeld, 2007; Herold et al., 2012).

A nivell genòmic la seva localització reflexa el que succeeix a *Drosophila*, és a dir que tant es troba prop de TSS com a regions intergèniques (Kim et al., 2007). També s'han estudiat les interaccions gèniques medades per

CTCF mitjançant ChIA-PET, que va revelar més de 1800 interaccions (Handoko et al., 2011). Aquestes reflexaven separacions entre dominis reprimits de H3K27 trimetilada i dominis actius de H3K4 trimetilada. A més, molts dels llocs de CTCF coincideixen amb llocs associats a la lamina (LADs), reforçant la idea de CTCF com a organitzador genòmic (Guelen et al., 2008).

1.2.3.3.2 Els insulators com a organitzadors de la cromatina

Les observacions tant a vertebrats com a *Drosophila* indiquen que els insulators, mitjançant les proteïnes que s'hi uneixen, estableixen contactes entre ells i amb altres elements reguladors (promotors, enhancers) i és a través d'aquests contactes que finalment s'obté l'expressió gènica pertinent. Així, a través únicament d'una funció *arquitectònica* acabarien per donar lloc a una regionalització de les regions cromosòmiques que contenen elements reguladors que explicaria els efectes observats des dels inicis del seu estudi: el bloqueig de l'acció d'un enhancer sobre un promotor; l'efecte *barrera*, delimitant l'extensió de l'estat epigenètic de la cromatina; o finalment la divisió d'una mateixa regió reguladora en dominis independents malgrat ser adjacents.

Aquesta explicació ve reforçada per les interaccions dels elements insulator amb elements del nucli, com la lamina, la matriu nuclear o els complexos del porus nuclear. A través d'aquestes interaccions es poden crear punts d'anclatge que, coherentment amb les hipòtesis de regionalització del nucli, ajudin a crear aquests dominis de cromatina d'acord amb el seu estat transcripcional. A gran escala, les agrupacions dins del nucli són observables i podrien ser el que coneixem com a *insulator bodies*.

Diversos estudis recents han proposat que la cromatina al nucli està compartimentalitzada físicament (Dixon et al., 2012; Hou et al., 2012; Sexton et al., 2012). Aquesta organització d'alt ordre (*higher-order chromatin organization*) sumaria una nova capa de regulació als també descrits diferents tipus de cromatina segons les marques epigenètiques (Filion et al., 2010).

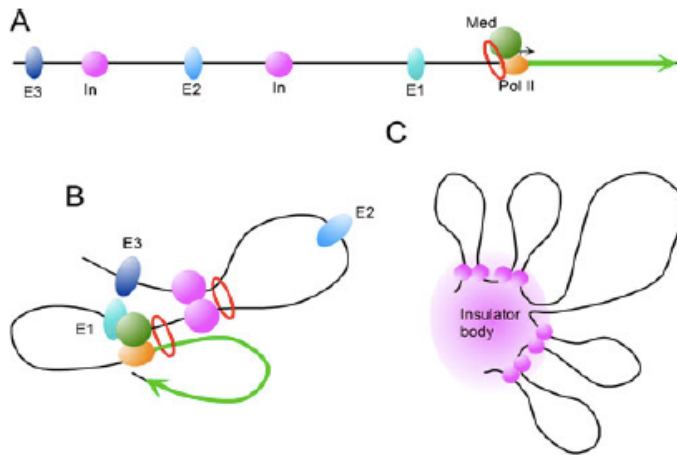


Figura 16.- Modelització del funcionament dels insulators. A) Un gen eucariota típic amb varis enhancers que el poden activar (E1, E2, E3), dos insulators per delimitar-los (In), i un inici de transcripció amb la Pol II i el complex Mediator (activador transcripcional). B) els anells vermells, que en vertebrats serien de cohesina, i les proteïnes d'insulator estableixen contactes que deixen fora els enhancers inapropiats (E2 i E3) i permeten l'activador per part de l'apropiat (E1). C) finalment, agrupacions de regions amb un mateix estat transcripcional serien portades al mateix lloc a través dels insulators i es formen els insulator bodies (Hou and Corces, 2012).

Els compartiments físics —dominis— poden incloure varis tipus de cromatina definida per la presència de proteïnes específiques i marques epigenètiques, amb la cromatina activa preferencialment situada als extrems i la silenciada a l'interior (Hou et al., 2012). Aquests dominis, o dominis topològics (Dixon et al., 2012), es defineixen com regions molt extenses, de l'ordre de megabases, de cromatina que interacciona amb ella mateixa. Els dominis acaben de manera abrupta en regions petites que funcionen de límit entre un domini i altre. S'ha observat, tant a vertebrats com a *Drosophila*, que hi ha un clar enriquiment de les proteïnes d'insulator a les regions que delimiten els dominis topològics (Dixon et al., 2012; Hou et al., 2012). Això suggereix que els insulators poden tenir un paper important a gran escala en la compartimentalització de la cromatina al nucli i la delimitació de dominis topològics.

A vertebrats el complex de les cohesines ja ha demostrat la seva capacitat de posar en contacte diferents fibres de cromatina. Sembla clar que un

dels mecanismes, si no el principal, de funcionament dels insulators a vertebrats és posant en contacte diferents regions reguladores a través de la unió de CTCF al DNA i de la interacció d'aquest amb les cohesines, que creen una xarxa de contactes intercromosomals de llocs CTCF. A *Drosophila* no s'ha demostrat la participació de la cohesina en insulators ni de cap complex similar, per tant s'ha d'investigar el mode de funcionament de proteïnes com CP190, per qui s'ha proposat un rol d'engranatge entre tots els tipus d'insulator, i s'ha de veure si interacciona amb noves proteïnes que puguin donar més pistes sobre com funcionen els insulators de *Drosophila*.

2.- OBJECTIUS

El nostre centre d'interès són els elements insulator i la seva contribució a la regulació de l'expressió gènica utilitzant com a organisme model *Drosophila melanogaster*. Els objectius d'aquesta tesi són:

- Anàlisi de la regió reguladora d'Abdominal-B: estudiar l'element regulador *Fab-6* i el seu rol com a insulator, possibles mecanismes de funcionament i proteïnes involucrades.

- Identificació de proteïnes implicades en l'activitat insulator: analitzar el complex proteic de CP190 i caracteritzar-ne els components, amb especial ènfasi en aquells que no han estat descrits i estudiar-ne la funció.

3.- MATERIALS I MÈTODES

3.1 Materials

3.1.1 Primers

Taula 1.- Llista de primers utilitzats. Cada primer es mostra en direcció 5' -> 3' i se n'indica l'experiment per al que s'ha utilitzat.

Primers	Seqüència 5' - 3'	Experiment
tubulina f	ACCTGAACCGTCTGATTGGC	RT-PCR
tubulina r	GCAGAGAGGCGGTAATCGAG	RT-PCR
act f	GCGCGTTACTCTTTCACCA	RT-PCR
act r	ATGTCACGGACGATTTACAG	RT-PCR
rp49 f	GTGCGAGTGCCGAGTTAAT	RT-PCR
rp49 r	TGGTTGCTGCTATCCCAATCTC	RT-PCR
pET 8436 U	AAACTCGAGATGCCCGAAAGAAGT	Generació d'anticossos
pET 8436 L	AAAGGATCCTTAATCGCGTAGG	Generació d'anticossos
pET 9740 U	AAACTCGAGATGTCGCAAGC	Generació d'anticossos
pET 9740 L	AAAGGATCCCTACTCGCATTC	Generació d'anticossos
WLE6	TCATCCTGGAGACGCTTAACT	Comprovació mutant Δ RB1
Vha 5'	ATAACAAACCAGGCGCAA	Comprovació mutant Δ RB1
CP190-R1 Upper	TAATACGACTCACTATAGGGCGCCTAGT CCAAGGGGAATAC	RNA doble cadena
CP190-R1 Lower	TAATACGACTCACTATAGGGAACGATT GAGCCACCTGAT	RNA doble cadena
CG8436-R1 U	TAATACGACTCACTATAGGGAACAAATT CTGGACGATGCTG	RNA doble cadena
CG8436-R1 L	TAATACGACTCACTATAGGGTCTTTGCT TTCGTCCAGGT	RNA doble cadena
CG9740-R1 U	TAATACGACTCACTATAGGGTCGGTATT AAGGTGGAGGCT	RNA doble cadena
CG9740-R1 L	TAATACGACTCACTATAGGGTCGGTCAA GTAGCTACCAAG	RNA doble cadena
LacZ U	TAATACGACTCACTATAGGGATGACCAT GATTACGCCAAGC	RNA doble cadena
LacZ L	TAATACGACTCACTATAGGGCAATTTCCA TTCGCCATTGAC	RNA doble cadena
CP190 Upper	AAACTCGAGATGGGTGAAGT	Clonatge CP190-TAP
CP190-C Lower	AAAGATATCTAGCTCCTCCTTCGCCGC	Clonatge CP190-TAP
BS Motif U	TCTGTGTTTTGACTATTTTACATTTGACCA TGGAT	EMSA
BS Motif L	TCTGATCCATGGTCAAATGTAATAAGTC AAAACA	EMSA
BS Fals U	TCTGTGTTTTGACTGGCCTTCTGTTGACC ATGGAT	EMSA
BS fals L	TCTGATCCATGGTCAACAGAAGGCCAGT CAAACA	EMSA
BS PolyT U	TCTGGAATAACATAGAATAACATAGAAT AACA	EMSA
BS PolyT L	TCTGTGTTATTCTATGTTATTCTATGTTAT TC	EMSA
BS Nostre 3x U	TCTGCTATTTTACATTTTATTGACATTTTA CATTGAGCTTACTCTTCTACATTTTACATT	EMSA

BS Nostre 3x L	TCTGAATGTAAATGTAGAAGAGTAAGC TCAATGTAAAATGTCAATAAAATGTAAA ATAG	EMSA
BS CTCF U	TCTGCATTTATTACATTGGTTTGTGTCTTT GCTTATCGCAACTAC	EMSA
BS CTCF L	TCTGGTAGTTGCGATAAGCAAAGACACA AACCAATGTAATAAATG	EMSA
Fab-8 U	CTTTGACGAGTTTTCCAAGC	Clonatge Casper-EnhW-Fab-8
Fab-8 L	GGGCTACCAGTGCCTGGCC	Clonatge Casper-EnhW-Fab-9
HS1 U	TTTATCTTGCATGGCTGC	ChIP
HS1 L	GAAGATGACTAAATCCAATGC	ChIP
HS2 U	ATTCAAATTCGGTTAGGTGC	ChIP
HS2 L	ATCTACTGGGTGCGGAGG	ChIP
Mcp U	TGCATTCTCATTTTCCCGC	ChIP
Mcp L	ACAACCCACCAGAGGACG	ChIP
F8 U	CAGACATCTTCGGTTCATCC	ChIP
F8 L	TGAATGCTGTTTTGGAATGC	ChIP
Fab7 U	TTGCGCTGGTGTGCGTGC	ChIP
Fab7 L	TCAAGCTGTGTGGCGGGG	ChIP
Sap5	GTGCGATAGGATTGCTGC	ChIP
SapM	GTTGGGGTACACGCAAGC	ChIP
Abd-B C	GGGTAAGGATAGGCGAACATGT	RT-PCR
Abd-B T	TTACGGCTCCGGTACTACG	RT-PCR
Fab7X C	TATGACGTCCGTTGCTCGTC	RT-PCR
Fab7X T	AATCGAGGGATCTTGGGAGG	RT-PCR
Fab6U (101RT2U)	TTATCTTGCAT	RT-PCR
Fab6L (101RT2L)	GCTAAACCCGATTTGCTT	RT-PCR
Fab7U (84U)	TCTGTGCTTCCCGTG	RT-PCR
Fab7L (84L)	GAAAGTGCACCGCCAATAA	RT-PCR
CG8436 U	AAAGATATCCAATGCCCGAA	Clonatge
CG8436 L	AAACTCGAGATCGCGTAGGG	Clonatge
CG9740 U	AAAGGTACCATGTCGCAAGC	Clonatge
CG9740 L	AAACTCGAGCTCGATTACG	Clonatge
3 Pin	ACCCTTAGCATGTCCGTGG	Comprovació mutant GSV17
8436L2	CGCCCGAAATCTTGGAT	Comprovació mutant GSV17
9740 5'U	CATCGGTTGAATTTACAGCAGC	Comprovació mutant GSV17

3.1.2 Plàsmids

Taula 2.- Llista de plàsmids utilitzats.

pCasper-EnhW	Generat per S. Pérez-Lluch. Conté l'enhancer del gen White
pCasper-EnhW-Fab7	Generat per S. Pérez-Lluch. Conté el Fab7 entre l'Enhancer del white i el mini-white
pCasper-EnhW-iab6HS1	Generat per S. Pérez-Lluch. Conté el Fab6 entre l'Enhancer del white i el mini-white
pCasper-EnhW-iab6HS2	Generat per S. Pérez-Lluch. Conté el PRE6 entre l'Enhancer del white i el mini-white

CP190-CTAP	En un vector pMK-CTAP hi hem clonat el CDS sencer de CP190
pET-CP190	Hem insertat el CDS de CP190 (que codifica des de l'aminoàcid 125 fins al final de la seqüència) a un pET14b per expressar la proteïna amb un tag d'histidines i purificar-la per fer anticossos
pET-CG8436	Hem insertat tot el CDS de Ibf-1 a un pET14b per expressar la proteïna amb un tag d'histidines i purificar-la per fer anticossos
pET-CG9740	Hem insertat tot el CDS de Ibf-2 a un pET14b per expressar la proteïna amb un tag d'histidines i purificar-la per fer anticossos
pCasper-Fab8	Conté el Fab-8 entre l'Enhancer del white i el mini-white. El Fab-8 s'ha amplificat amb els primers Fab-8 U i Fab-8 L
LD32555	Plàsmid provinent del <i>Drosophila Genomics Resource Center</i> que conté el cDNA de Ibf-1 taguejat a un HA
paC5.1-CG9740	Plàsmid que conté el cDNA de Ibf-2 taguejat a un V5

3.1.3 Anticossos

Taula 3.- Llista d'anticossos utilitzats. Per cada anticòs se n'indica l'animal en que es va generar i d'on el vam obtenir.

Primaris	Animal	Origen
HA	Rata	Roche (12CA5)
BEAF	Ratolí	Hybridoma Bank
NURF-38	Conill	Cedit per Paul Badenhorst
Z4	Rata	Cedit per Roman Kessler
V5	Ratolí	Invitrogen (R960-25)
CP190	Conill	Generat al laboratori
Ibf-1	Rata	Generat al laboratori
Ibf-2	Rata	Generat al laboratori
Cenp-C	Conill	Cedit per Olga Moreno-Moreno i Mònica Torras-Llort
GFP	Conill	Invitrogen(A-6455)
Tubulina	Ratolí	SIGMA (A2547)
Actina	Ratolí	Chemicon (MAB3408)
H3S10 fosforilada	Conill	Cell signaling (#9701)
Polycomb	Conill	Santa Cruz (sc-25762)
GAGA	Conill	Generat al laboratori

Secundaris	Fabricant
Anti-conill HRP	Amersham (NA 934)
Anti-rata HRP	Jackson (712-035-150)
Anti-ratolí HRP	Jackson (715-035-150)
Anti-conill cy2	Jackson (711-225-152)
Anti-rata cy3	Jackson (712-165-150)
Anti-rata cy2	Jackson (712-225-150)

Els anticossos que s'han generat al laboratori s'han fet seguint el protocol descrit a (Sambrook et al., 2001). Els anticossos de Ibf-1 i Ibf-2 s'han generat a partir de la seqüència sencera del CDS clonada a un pET14b, i la proteïna purificada s'ha injectat a rates. En canvi l'anticòs de CP190 s'ha generat amb la part de la proteïna entre els aminoàcids 125-1096 i s'ha fet a conill.

3.1.4 Línies de *Drosophila*

- *white*⁻ (w1118): Mutació al gen *white* que elimina la proteïna.
- CP190^{P1}: Mutant de CP190 (Pai et al., 2004)
- CP190^{P11}: Mutant de CP190 (Pai et al., 2004)
- CP190^{H31.2}: Mutant de CP190 (Pai et al., 2004)
- CTCF^{Y+6}: Mutant de CTCF (Gerasimova et al., 2007)
- PBac{RB}CG8436^{e03576}: Inserció d'un element pBac(RB) al gen *Ibf-2*.
- P{GSV6}GS16482: Inserció d'un element GSV6 al gen *Ibf-2*.
- JIL1^{Z60}: Mutant de JIL-1 (Zhang et al., 2003)
- RNAi CG8436 (VRDC 12526): Expressa l'RNAi de *Ibf-1* sota el control d'un UAS.
- B+P7: Conté el Fab-7 entre l'enhancer del *white* i el mini-*white*.
- Fab8: Conté el Fab-8 entre l'enhancer del *white* i el mini-*white*.
- Fab6: Conté el Fab-6 entre l'enhancer del *white* i el mini-*white*.
- PRE-6 (HSiab62): Conté el PRE-6 entre l'enhancer del *white* i el mini-*white*.

3.1.5 Soques bacterianes

- Per a amplificar plàsmids utilitzem **DH5 α** : *deoR endA1 gyrA96 hsdR17* ($r_{\text{K}}^- m_{\text{K}}^-$) *recA1 relA1 supE44 thi-1 F'[traD36 proAB⁺ lac1^q lacZ Δ M15]*
- Per a expressar proteïnes utilitzem **BL21 (DE3)**: *BF⁻ dcm ompT hsdS* ($r_{\text{B}}^- m_{\text{B}}^-$) *gal λ (DE3)*

3.1.6 Línies cel·lulars

En aquest treball hem treballat sempre amb la línia S2 o SL2 de cèl·lules de *Drosophila*, a partir de la qual hem generat línies transfectades tant transitòries com estables. Són cèl·lules d'origen embrionari tardà (20-24 hores) que creixen en semi-adherència als flascons de medi. En algun experiment hem utilitzat la línia SG4, que és també d'origen embrionari i semblants a les S2, però tenen programes transcripcionals diferents. En particular, per al nostre interès, el gen Abd-B s'expressa mentre que en S2 està silenciada.

3.2 Mètodes

3.2.1 Cultius cel·lulars

3.2.1.1 Manteniment

Les cèl·lules S2 (o SL2) es creixen a 25°C sense CO₂ amb medi *Schneider's insect Medium* (Sigma) suplementat amb 100 µg/ml d'estreptomicina (Gibco), 100 unitats/ml de penicil·lina (Gibco) i un 10% de sèrum fetal boví (FBS) (Gibco), el qual s'ha d'inactivar 30 minuts a 56°C. Es mantenen en flascons Corning de 75 cm² a una densitat de 1-8·10⁶ cèl·lules/ml en un volum de 5-6 ml, fent dilucions 1/6 o 1/8 cada 3-4 dies. I, per obtenir grans quantitats de cèl·lules per transfectar es prepara un flascó de 175 cm² fent una dil·lució 1/6 en un volum final de 30 ml a partir de cèl·lules bastant crescudes i es deixa créixer durant 3-4 dies.

Per congelar les cèl·lules es fa créixer un flascó de 75 cm² fins a una densitat de 5·10⁶ cèl·lules/ml. Per recollir les cèl·lules es transfereix el cultiu a un falcon de 15 ml i es centrifuga 2 minuts a 1300 rpm a temperatura ambient. Llavors s'elimina el sobrenedant i es resuspèn el sediment cel·lular amb 900 µl de FBS. Es passa a un criotub i es va afegint gota a gota 100 µl de DMSO. El procés de congelació ha de ser lent. Per això, primer es guarda el criotub durant 2 hores a -20°C, després tota la nit a -80°C i, finalment es guarda en nitrogen líquid. Per descongelar-les s'agafa un criotub guardat en nitrogen líquid i es descongela ràpidament entre les mans. Llavors, es passa a un falcon de 15 ml amb 5 ml de medi a temperatura ambient, es centrifuga 2 minuts a 1300 rpm a temperatura ambient i es resuspèn el sediment cel·lular amb 5 ml de medi. Finalment, es passa a un flascó Corning de 75 cm² i es deixa créixer durant 3-4 dies.

3.2.1.2 Transfecció

Per transfectar transitòriament els DNA plasmídics fem servir el mètode de fosfat càlcic. Aquest facilita la seva entrada dins de la cèl·lula mitjançant la

formació de precipitats de DNA-fosfat. El fet que la transfecció sigui transitòria fa que s'hagi d'analitzar en un període de 24 a 72 hores. Per obtenir suficient cèl·lules per transfectar es fan créixer a una densitat d'aproximadament $6 \cdot 10^6$ cèl·lules/ml en un volum de 30 ml en flascons de 175 cm^2 . Un cop quantificada la concentració exacta mitjançant un comptatge amb la càmera de Neubauer, s'inoculen $2 \cdot 10^6$ cèl·lules en plaques de cultiu de 60 mm^2 (Corning) en un volum final de 5 ml i s'incuben 24 hores a 25°C . Per cada placa es transfecten $10 \mu\text{g}$ de DNA en un volum final de $50 \mu\text{l}$.

Al dia següent de plaquejar les cèl·lules es preparen en eppendorfs les quantitats de DNAs necessàries en un volum final de $125 \mu\text{l}$ i a cadascun es va afegir 1 ml de CaCl_2 0.25 M esterilitzat per filtració amb un filtre de $22 \mu\text{m}$. En un tub falcon de 15 ml s'afegeix 1 ml de HeBS 2x (NaCl 250 mM, KCl 9 mM, Na_2HPO_4 1.5 mM, glucosa 10 mM, HEPES 50 mM pH 7.1) esterilitzat també per filtració. Mentre es vorteja suaument s'afegeix gota a gota la barreja de DNA- CaCl_2 sobre el HeBS 2x per produir precipitats de tamany petit i aspecte blanquinós. Aquests precipitats s'incuben durant 35 minuts a 25°C i, transcorregut aquest temps, s'afegeixen $850 \mu\text{l}$ per placa. Si fem una transfecció transitòria incubem les plaques 48 hores.

3.2.2 Tècniques bàsiques de treball amb DNA

3.2.2.1 PCR

A les PCR clàssiques hem fet servir sempre unes condicions estandard a les que hem hagut de fer variacions segons els oligos (augmentant o disminuint la temperatura d'hibridació, i en algun cas variant lleugerament la quantitat de magnesi) o variacions segons la llargada de l'amplicó (a raó aproximadament d'1 kb per minut d'elongació).

En una reacció estandard hi posem el DNA (de l'ordre de nanograms), MgCl_2 a 2 mM, els dNTP a 0.4 mM, el tampó a 1X, els primers a $0.5 \text{ pmol}/\mu\text{L}$, i la polimerasa a $0.05 \text{ U}/\mu\text{L}$. Generalment hem utilitzat la Taq Polimerasa Biotools.

Els cicles i temperatures estàndard que utilitzem:

2 min a 95°C

--

1 min a 95°C – Desnaturalització

1 min a 55 °C – Hibridació

1 min a 72°C – Elongació

x 30 cicles

--

10 min a 72°C

3.2.2.2 Transformació de cèl·lules competents

Incubar les cèl·lules amb 5 uL de la lligació durant 20 minuts en gel. Xoc tèrmic de dos minuts a 42°C en un bany. Recuperar les cèl·lules en 400 uL de medi LB sense antibiòtic durant 10 minuts. A la flama, plaquejar en una placa d'LB amb antibiòtic. Deixar tota la nit en una estufa a 37°C.

Medi LB

10 grs/L triptona

5grs/L extracte de llevat

10 grs/L NaCl

1 mM NaOH

3.2.2.3 Minipreparació de DNA plasmídic

Inocular 4 mL de medi LB amb antibiòtic amb una colònia aïllada. Deixar créixer tota la nit a 37°C i en agitació. Passar 1,5 mL a un eppendorf i centrifugar a 9000 rpm durant un minut. Treure el sobrenadant amb pipeta Pasteur i resuspendre en 100 uL de GTE vortjant. Afegir 200 uL de solució de NaOH 0,2N-SDS 1% i desar en gel durant cinc minuts. Afegir 150 uL d'acetat de sodi 3M, agitar i desar en gel cinc minuts més. Centrifugar cinc minuts a 14000 rpm, passar sobrenadant a un eppendorf nou. Afegir 450 uL de fenol/cloroform, vortejar i centrifugar 2 minuts a 14000 rpm. Recuperar la fase aquosa (superior) i afegir-hi 1 mL d'etanol fred. Invertir do o tres cops i centrifugar cinc minut a 1400 rpm. Treure el sobrenadant i rentar el pèlet amb 300 uL etanol 70%. Centrifugar 3 minuts a 14000 rpm, descartar el sobrenadants i deixar assecar el pèlet. Resuspendre'l amb 50 uL de TE 1X i 1 ug d'RNAsa, 15 minuts a 37°C.

TE 1X

10 mM Tris HCl pH 8
1mM EDTA

GTE

50 mM glucosa
10 mM EDTA
25 mM Tris HCl pH 8

3.2.2.3 Maxipreparació de DNA plasmídic

Inocular 500 mL de medi LB amb antibiòtic amb un minicultiu. Incubar tota la nit en agitació a 37°C. Centrifugar a 4500 rpm. Descartar el medi i extreure i purificar el DNA amb Nucleobond. Deixar assecar el pèlet de DNA i resuspendre en 500 uL de TE 1X.

3.2.2.4 Extracció de DNA genòmic

S'ha de posar una mosca en un eppendorf i congelar-lo. Llavors se li afegeixen 40 uL de buffer SB i 0.4 uL de proteïnasa K 20 mg/mL. Es tritura amb un pal blau d'eppendorfs. S'incuba 30 minuts a 37°C i després dos minuts a 95°C per inactivar la proteïnasa. D'aquí se'n fa una dilució 1/5 per utilitzar-lo en PCR.

3.2.3 Tècniques bàsiques de treball amb RNA**3.2.3.1 Extracció d'RNA**

Si fem l'extracció de larves, en necessitem 15, que posem en un dounce amb Nitrogen líquid i esclafem mentre el nitrogen s'evapora, de manera que les larves queden pulveritzades, i procedim a la purificació amb Trizol. En el cas de les cèl·lules partim d'un cultiu d'un flascó petit que centrifuguem cinc minuts a 300 g i n'extreiem el medi i procedim al Trizol.

Per purificar, afegim 1 mL de Trizol que homogenitzem amb la pipeta, tot seguit hi afegim 0.2 mL de cloroform, que sacsegem enèrgicament, i centrifuguem a 12000 g 15 minuts a 4°C. Extreiem la fase aquosa i la precipitem

afegint 0.5 mL d'isopropanol, que centrifuguem a 12000 g durant 10 minuts a 4°C. Després treiem l'isopropanol, hi posem 1 mL d'etanol 75% i centrifuguem a 7500 g cinc minuts a 4°C. Finalment es resuspèn el pèlet amb 100 µL d'aigua mili-Q. Addicionalment es pot fer un rentat extra de l'RNA amb l'RNAeasy Kit de Qiagen.

3.2.3.2 RT-PCR

Es tracta de, en un primer pas, retrotranscriure *in vitro* l'RNA a DNA per, en un segon pas, poder-lo amplificar i així estimar les quantitats relatives de diferents molècules d'RNA. La real-time PCR o PCR quantitativa funciona de manera que hi ha un reactiu, el SybrGreen, que s'uneix al DNA de doble cadena —per tant als amplicons en una PCR— i que emet fluorescència. Aquesta és detectada pel mateix termociclador i s'utilitza per conèixer la quantitat d'amplicó generat al cap dels cicles. L'estimació de les quantitats relatives de DNA inicial sorgeixen de la idea que, donats un nombre de cicles iguals per a diferents mostres, aquelles que donen més fluorescència partien de més DNA inicial, i viceversa. Per tant no es tracta d'una estimació de la quantitat de DNA de la que s'originava cada mostra sinó d'una estimació relativa entre mostres d'un mateix experiment. És per això que sovint s'utilitzen gens de transcripció constant i ubíqua (housekeeping) com a referència i els resultats d'un experiment es donen com a mesura de la relació dels gens d'interès (target) a un gen de referència housekeeping.

La retrotranscripció la vam fer amb el kit Transcriptor First Strand cDNA Synthesis Kit de Roche. Utilitzem 1 µg d'RNA, un primer oligo-dT (que s'uneix a tots els RNA poly-A) 2.5 µM, buffer de transcripció 1X, inhibidor d'RNAsa 20 U, mix de dNTP 1mM, i 10 U de Transcriptasa Reversa. Ugem amb aigua fins a 20 µL i deixem la reacció 1 hora a 50°C. Després inactivem l'enzim esclafant a 85°C cinc minuts i ho guardem en gel.

Per fer la real-time PCR utilitzem el kit Light Cyclor 480 Sybr Green I Master. Fem les reaccions en un volum de 10 µL, dels quals 1 µL és de cDNA, 0.5

uL de cada primer, 5 uL de mix de SybrGreen i 3 uL d'aigua. Ho posem en plaques de 12 x 8 i utilitzem la màquina Light Cycler 480 de Roche. El programa estandard és el següent:

5 min a 95°C
--
10 segons a 95°C – Desnaturalització
10 segons a 60 °C – Hibridació
10 segons a 72°C – Elongació
x 45 cicles

3.2.3.3 Síntesi de RNA doble cadena

L'RNA de doble cadena s'ha utilitzat per fer knock-downs d'un transcrit a cèl·lules S2 mitjançant RNAi. Per a sintetitzar-lo, s'escull una regió de 150-300 pb d'una regió exònica del gen a eliminar i es fa una PCR convencional amb uns oligos que incorporen als seus extrems el promotor de la RNA polimerasa bacteriana T7. Amb l'amplicó de DNA es fa la transcripció in vitro seguint el procediment per defecte del kit Megascript T7 d'Invitrogen, inclòs el pas de TurboDNAsa per eliminar el DNA motlle restant.

3.2.4 Tècniques bàsiques de treball amb proteïnes

3.2.4.1 Extracció proteica

Si extreiem proteïnes de cervell de larva, disseccionem 15 larves III amb el mateix temps de vida i amb el mateix genotip i n'aïllem els cervells amb les pinces en PBS1X-NP40 0,05%. Aquests cervells els posem en 200 uL de PLB 1X-beta-mercaptoetanol 2% els posem a 95°C durant tres minuts. Tot seguit els pipetegem amunt i avall per disgregar i centrifuguem a 14000 rpm 3 minuts. Recuperem el sobrenadant, soniquem cinc minuts al 30% en un sonicador Branson.

Si extreiem proteïnes de cèl·lules S2, hem d'agafar un flascó petit a mig créixer i extreure'n el contingut, que posem en un Falcon de 5 mL. Centrifuguem

a 300 g cinc minuts i llencem el sobrenadant, que és el medi. Les cèl·lules, al pèlet, les resuspenem amb PLB1X-beta-mercaptoetanol 2%, posem a 95°C cinc minuts, pipetegem bé amunt i avall, i soniquem al Branson cinc minuts al 30%.

PLB 5X

0.125 M Tris pH 6.8

22% glicerol

5% SDS

Blau de bromofenol (punta d'espàtula)

3.2.4.2 Electroforesi en gel d'acrilamida

L'electroforesi en gel d'acrilamida separa proteïnes segons el seu pes molecular. La separació té lloc en condicions desnaturalitzants amb SDS i a través d'una matriu de poliacrilamida (PAGE). El gel es compon d'una matriu apiladora, al 5% d'acrilamida, i una separadora al 10%.

Totes les mostres es carreguen en tampó de càrrega (PLB) i amb beta-mercaptoetanol, havent bullit cinc minuts a 95°C. Es corren a través del gel amb tampó Laemmli 1X.

Gel apilador

acrilamida:bisacrilamida 29:1 5 %

Tris-HCl 0.125 M pH 6.8

SDS 0.1 %

6 µl PSA 10 %/ml solució

3 µl TEMED/ml solució

Gel separador

acrilamida:bisacrilamida 29:1 10-18 %

Tris-HCl 0.4 M pH 8.7

SDS 0.1 %

6 µl PSA 10 %/ml solució

3 µl TEMED/ml solució

Tampó Laemmli

1% SDS

2M glicina

250mM Tris base

3.2.4.3 Western Blot

Un cop hem separat les proteïnes amb un gel d'acrilamida (SDS-PAGE), les transferim del gel a una membrana de nitrocel·lulosa. Es transfereix durant dues hores a 4°C a 80V en buffer de transferència. Després la membrana es bloqueja amb solució de bloqueig durant una hora a temperatura ambient o tota la nit a 4°C. Després es fan tres rentats amb PBS-tween 0,1%. S'incuba la membrana amb anticòs primari durant una hora i es fan tres rentats més de deu minuts amb PBS-tween 0,1%. S'incuba durant una hora amb l'anticòs secundari conjugat amb peroxidasa i es torna a rentar cinc cop cinc minuts amb PBS-tween 0,1%.

Per a revelar afegim sobre la membrana unes gotes de la solució reveladora (ECL Western Blotting System de GE Healthcare) de manera que recobreixi tota la membrana, durant un minut. S'asseca la membrana i s'exposa cinc minuts a un film d'autoradiografia Curix RP2 (Agfa). El film es revela amb un HyperProcessor d'Amersham.

Tampó de transferència

12 grs Tris Base
12 grs glicina
20 mL SDS 10%
800 mL metanol
Volum final 4L

Solució de bloqueig

5 grs llet en pols
100 mL PBS1X-tween 0,1%

3.2.4.5 Co-Immunoprecipitació

Partim de dos flascons petits de cèl·lules S2 que centrifuguem cinc minuts a 300 g i rentem amb PBS 3 vegades. Resuspenem el pèlet amb buffer de lisi, vortegem, es lisa amb el dounce B i es deixa 30 minuts a la roda a 4°C per extreure les proteïnes. Centrifuguem deu minuts a 13200 rpm a 4°C. A la quantitat d'extracte desitjada per immunoprecipitació li afegim 2 uL d'anticòs i es deixa incubant a 4°C 3 hores a la roda. Després s'hi afegeix 50 uL de proteïna

A-sefarosa equilibrada amb buffer de lisi i es deixa a la roda a 4°C durant dues hores. Després es renten els beads de proteïna A-sefarosa amb buffer de lisi tres vegades centrifugant entremig 1 minut a 3000 rpm. Finalment s'elueix afegint PLB1X-beta-mercapto 2%, bullint 5 minuts a 95°C i vortejant.

Amb les elucions de cada IP es corren en un gel SDS-PAGE i se'n fa un Western Blot utilitzant com a anticòs primari la proteïna d'interès que volem saber si es troba en la IP de la proteïna immunoprecipitada. Utilitzem com a input una quantitat d'extracte que correspon al 10% del que hem utilitzat per immunoprecipitar, que també correm al SDS-PAGE i en fem Western Blot.

3.2.5 Immunotincions

3.2.5.1 Cromosomes politènics

Disseccionem les glàndules salivals de larva III en buffer Cohen, en el que hi han de romandre 8-10 minuts. Fem aproximadament larves per cada preparació. Treiem el buffer i fixem primer amb la solució de fixació 1 (2 minuts) i després amb la solució de fixació 2 (3 minuts). Llavors transferim les glàndules a un cobreobjectes siliconitzat i hi posem un portaobjectes al damunt, i piquem diverses vegades amb un llapis, després girem el porta i apretem fort amb el dit. Llavors amb unes pinces ho suquem en Nitrogen líquid, li treiem el cobreobjectes i ho posem en un recipient amb PBS-tween 0.05%.

Fem tres rentats amb el PBS-tween 0.05% i bloquegem dos vegades 20 minuts amb PBS-tween 0.05%-BSA 2%. Es posa l'anticòs primari diluït en la mateixa solució de bloqueig (cada anticòs requereix la seva pròpia dilució) amb un cobreobjectes al damunt i s'incuba tota la nit a 4°C. Es fan tres rentats de cinc minuts amb PBS-tween 0.05%, i es posa l'anticòs secundari a dilució 1:400 en PBS-tween 0.05% durant una hora a temperatura ambient. Finalment rentem tres cops amb PBS-tween 0.05%, un cop amb PBS i muntem la preparació amb DAPI-Mowiol. Els anticossos de Ibf-1 i Ibf-2 els diluïm 1:250 i de CP190 1:500.

Buffer Cohen

MgCl₂ 10 mM
Glicerol de sodi 3P 25 mM
Ca Cl₂ 3 mM
KH₂PO₄ 10 mM
NP40 0.5%
Kcl 30 mM
Sacarosa 160 mM

Solució de Fixació 1

PBS 10X 250 uL
Formaldehid 37% 50 uL
Aigua 2,2 mL

Solució de Fixació 2

Àcid acètic 1,125 mL
Formaldehid 37% 50 uL
Aigua 1,325 mL

3.2.5.2 Squash de cervells

Disseccionem els cervells de larva III en medi salí 0.07% NaCl. Necessitem uns tres cervells per preparació. Fem un xoc hipotònic amb citrat de sodi al 0.05% durant deu minuts. Fixem amb formaldehid al 3,7% durant 30 minuts. Transferim els cervells a un cobreobjectes siliconitzat que conté àcid acètic al 60% i fem un squash com el descrit als politènics.

Bloquegem amb PBS-tritó 0.01%-llet 1% durant una hora. Després incubem amb l'anticòs primari (dilució variable segons l'anticòs) tota la nit a 4°C. Rentem dos vegades amb PBS-tritó-llet i incubem amb l'anticòs secundari 45 minuts a dilució 1:400. Rentem dos cops amb PBS-tritoo-llet, després dos cops amb PBS i muntem amb DAPI-Mowiol. Els anticossos de Ibf-1 i Ibf-2 els diluïm 1:500 i de CP190 1:1000.

3.2.5.3 Cèl·lules S2

Posem concanavalina a 0.5 mg/mL sobre els cobreobjectes sobre els que farem la immunotinció. Als cobreobjectes, en pouets, els hi afegim al damunt les cèl·lules encara en medi de creixement (uns 100 uL per pou) i

esperem 2-3 hores a que s'hagin enganxat als cobreobjectes. Després de rentar amb PBS fixem amb paraformaldehid 4% en PBS durant 15 minuts. Rentem dos cops 10 minuts amb PBS-tritó 0,1%-BSA 1g/L. Incubem amb l'anticòs primari diluït en aquesta mateixa solució durant una hora a temperatura ambient i després tota la nit a 4°C. Fem tres rentats de deu minuts amb PBS-tritó-BSA i incubem amb l'anticòs secundari a 1:400 en PBS-tritó-BSA durant una hora a temperatura ambient. Finalment rentem dos cops deu minuts amb PBS-tritó i dos més amb PBS i llavors muntem amb DAPI-Mowiol.

Si fem la immunotinció impactant les cèl·lules hem de fer algunes coses diferent. D'entrada podem les cèl·lules, uns 300 uL del cultiu, en solució hipotònica MAC i ho deixem cinc minuts. Llavors les impactem a la citocentrífuga a 500 rpm durant 10 minuts i deixem assecar al porta durant una hora. Després fixem amb paraformaldehid 3.7% durant deu minuts. Rentem amb PBS 15 minuts i bloquegem amb PBS-tween 0,2%-BSA 0,1% dos cops durant deu minuts. A partir d'aquí fem la immunotinció com l'explicada prèviament. Els anticossos de Ibf-1 i Ibf-2 els diluïm 1:200 i de CP190 1:400.

Buffer Mac

Glicerol 50 mM
Kcl 5 mM
NaCl 10 mM
CaCl₂ 0,8 mM
Sacarosa 10 mM

3.2.5.4 Embrions

Primer s'han de decoronar els embrions amb lleixiu directament sobre la placa de posta durant dos o tres minuts. Fixem amb 2 mL de PEM, 2 mL d'heptà i 250 uL de formaldehid durant 20 minuts. Extreiem el PEM i afegim 2 mL de metanol i vortegem 20 segons. Extreiem els embrions que queden al fons del tub i els passem a un eppendorf i rentem amb metanol tres cops.

Es fan tres rentats ràpids amb PBS-tritó 0,1% i bloquegem dos cops 30 minuts amb PBS-tritó 0,1%-BSA 1% i s'incuba amb l'anticòs primari diluït en la solució anterior en agitació tota la nit a 4°C. Al matí es fan tres rentats de 15 minuts amb PBS-tritó 0,1%-BSA 1% i s'incuba amb el secundari a 1:400 en PBS-tritó 0,1% durant dues hores a temperatura ambient. Llavors es renta tres cops 15 minuts amb PBS-tritó 0,1%, després només amb PBS i finalment es munta amb DAPI-Mowiol. Els anticossos de Ibf-1 i Ibf-2 els diluïm 1:500 i de CP190 1:1000.

PEM

Pipes pH 6,9 0,1M

MgSO₄ 2 mM

EGTA 1 mM

3.2.6 Purificació d'un complex proteic

Partim de 2 L de cultiu cel·lular de cèl·lules CP190-TAP induïdes amb CuSO₄ 0.07 mM durant tota la nit. Les diluïm amb buffer A utilitzant tres volums el volum del pèlet. Lisem les cèl·lules amb el dounce L i centrifuguem 15 minuts a 4500 rpm. Guardem el sobrenadant i tornem resuspendre el pèlet amb buffer A i tornem a centrifugar. Llavors resuspenem amb buffer B, utilitzant dues vegades el volum del pèlet. Aquest cop lisem amb el dounce S, alliberant l'extracte nuclear. Afegim NP40 0.1% i ho deixem a la roda a 4°C 15 minuts. Centrifuguem a 15000 rpm 30 minuts a 4°C. El sobrenadant és l'extracte nuclear soluble.

Posar junts l'extracte nuclear i els beads magnètics (Dynabeads, Invitrogen) recoberts amb BSA equilibrats amb buffer B. S'incuba una hora a 4°C a la roda. Es separa el sobrenadant del beads amb un iman i s'hi afegeixen beads magnètics recoberts amb IgGs i equilibrats amb buffer B. S'incuba 4 hores a 4°C a la roda. Es renten els beads 13 vegades amb buffer de rentat i finalment es traslladen a un eppendorf des d'on s'elueix amb Glicina 50 mM pH3. Se

n'afegeixen 30 uL, es vorteja, s'incuba 2 minuts i es recolecta l'elució, repetint-ho 4 vegades fins a tenir 120 uL.

Primer de tot fem un gel on hi carreguem un 10% de cada elució que es tenyeix amb plata per observar les bandes que hi apareixen. Després acabem de carregar totes les mostres en un altre gel tenyit amb Comassie que només deixarem córrer fins la frontera entre apilador i separador, de manera que totes les proteïnes estaran concentrades en una sola banda que tallarem i s'analitzarà per espectrometria de masses. Preparem un gel d'acrilamida de gradient 9-13%. A les elucions hi afegim PLB1X-beta-mercaptoetanol 2%. Es corre el gel a 60 mA. L'anàlisi de les mostres va ser per LC/MS a l'Unitat de Proteòmica de l'Institut de Recerca Vall d'Hebron (Barcelona) a càrrec del Dr. Francesc Canals.

Buffer A (2X)

Sacarosa 0.46 M
KCl 120 mM
NaCl 30 mM
MgCl₂ 0.5 mM
Spermina 1mM
Spermidina 0.3 mM
Tris pH 7.4 30 mM
Beta-mercaptoetanol 28 mM
Inhibidors de proteases 250 mL

Buffer B (2X)

Hepes-KOH pH 7.9 40 mM
Glicerol 40%
NaCl 0.6 M
MgCl₂ 3 mM
EGTA 1 mM
DTT 1 mM
Inhibidors de proteases 250 uL

Buffer de Rentat

Hepes pH7.9 20 mM
Glicerol 20%
NaCl 0.3M
NP40 0.1%
EGTA 0.5mM
DTT 0.5 mM
Inhibidors de Proteases 250 uL

3.2.7 CHIP-Seq

3.2.6.1 CHIP

Es parteix de cultius crescuts de cèl·lules S2 en flascons de 25 cm². S'hi afegeix formaldehid a una concentració final de 1.8% per crosslinkar proteïnes i DNA, durant 10 minuts a temperatura ambient balancejant-se. S'atura la reacció afegint glicins a una concentració final de 0.125 M. Es centrifuga 2 minuts a 1500 g, es fa un rentat amb PBS i el pèlet es resuspèn finalment amb 10 mL de buffer de rentat A. S'incuba 10 minuts a la roda, després es torna a centrifugar i es resuspèn el pèlet en 10 mL de buffer de rentat B, s'incuba 10 minuts a 4°C a la roda. Es torna a centrifugar i es renta amb TE. Després, en solució de TE, PMSF 1mM i SDS 0.1% es sonica. Soniquem al sonicador Bioruptor, en cicles de 30 segons d'encès i apagat, fent un total de 25 minuts. Al final es centrifuga a 14000 rpm cinc minuts i es guarda el sobrenadant. S'hi afegeix tritó, deoxicolat de sodi i NaCl a concentracions corresponents al buffer RIPA.

Llavors hi afegim 2uL de l'anticòs corresponent i ho deixem incubant a 4°C a la roda. Al matí afegim 40 uL de beads de proteïna A-sefarosa 3 hores a 4°C a la roda. Després es fan rentats dels beads de cinc minuts centrifugant 2 minuts a 3000 rpm entre cadascun. Cinc rentats amb RIPA, un rentat amb buffer LiCl i finalment dos rentats amb TE.

Afegim 40 uL de TE amb 0.5 uL de RNAsa i es posa a 37°C durant 30 minuts. S'elueix amb 50 uL de NaHCO₃ 0.2M i 10 uL de SDS 10%, vortejant i centrifugant a 14000 rpm durant mig minut, i això es repeteix tres vegades fins aconseguir una elució de 300 uL. S'incuba a 65°C durant tota la nit per descrosslinkar DNA i proteïnes. Pel matí es digereixen les proteïnes amb 3 uL de proteïnasa K 20mg/mL i s'incuba 3 hores a 55°C. Finalment es purifica el DNA pel mètode del fenol-cloroform i es precipita amb etanol i acetat de sodi 3M. Finalment el pèlet es resuspèn amb 25 uL d'aigua mili-Q.

Buffer de rentat A

Hepes pH 7.9 10 mM
EDTA 10 mM
EGTA 0.5 mM
Tritó X100 0.25%

Buffer de rentat B

Hepes pH 7.9 10 mM
EDTA 1 mM
EGTA 0.5 mM
Tritó X100 0.01%
NaCl 100 mM

RIPA

NaCl 140 mM
Tris Hcl pH8 10 mM
EDTA 1 mM
SDS 0.1%
Deoxicolat de sodi 0.1%

Buffer LiCl

LiCl 250 mM
TrisHCl pH8 10 mM
EDTA 1mM
NP40 0.5%
Deoxicolat de Sodi 0.5%

3.2.6.2 Seqüenciació i anàlisi

El DNA immunoprecipitat es va processar i seqüenciar a la Functional Genomics Core de l'Institut de Recerca Biomèdica de Barcelona per David Fernández i Herbert Auer. La llibreria es va preparar a partir de 10 ng de DNA, quantificats amb Qubit dsDNA HS Assay kit (Invitrogen). Reparació d'extremes, adenilació, lligació d'adaptadors i enriquiment per PCR de 18 cicles es van realitzar amb TruSeqRNA Sample Prep Kit (Illumina). Les llibreries purificades es van quantificar amb Qubit dsDNA HS Assay kit (Invitrogen) i la distribució per mides es va evluar amb el BioAnalyzer DNA 1000 (Agilent). La seqüenciació single-end de 50 nucleòtids es va fer al Genome Analyzer lix. El control de qualitat de les dades es va fer amb el Software Casava 1.8.2.

L'anàlisi de la seqüenciació, control de qualitat, anotació, solapaments i tests estadístics han estat realitzats per Òscar Reina de la unitat de Bioinformàtica de l'Institut de Recerca Biomèdica de Barcelona. Els anàlisis s'han dut a terme amb R i Bioconductor (Gentleman et al., 2004; Team, 2011). Les dades de seqüenciació Solexa/Illumina de CP190, CG8436 (Ibf-1) i CG9740 (Ibf-2) i input a mosques wild-type es van pre-processar amb Illumina pipeline versió 1.9 i les seqüències alineades al genoma de *Drosophila melanogaster* (UCSC dm3 version) amb el software Illumina Eland v2. Es van mantenir les seqüències que mapaven a una sola regió del genoma, deixant lloc per a fins dos mismatch a les primeres 32 bases i sense alineaments amb gaps.

Com que els artefactes de sobre-amplificació de la PCR típicament resulten en una mateixa seqüència repetida moltes vegades, el filtre DuplicateReads de htSeqTools package (Planet et al., 2012) va ser utilitzat per detectar-los i extreure'ls, amb un FDR de l'1%. També es van extreure biaixos específics de cadena utilitzant alignPeaks del mateix Package. Per determinar els llocs d'unió es van fer dos passos. Primer, es va fer servir enrichedRegions a htSeqTools per trobar regions genòmiques cobrint més de 10 reads i mostrant gran acumulació de seqüències a la IP comparat amb el control. Es seleccionen les regions amb coverage de més de 10 i es compara la proporció de reads dins/fora de cada regió entre la mostra IPP i el seu control amb una regressió logística de test de versemblança. Es van definir les regions enriquides com aquelles amb un p-valor Benjamini-Yekutieli menor que 0.05. En un segon pas, s'utilitza enrichedPeaks de htSeqTools per definir els pics com a localitzacions dins de les regions enriquides amb una diferència de coverage entre mostres de 100.

Els pics es van anotar amb annotatePeakInBatch del package ChIPpeakAnno (Zhu et al., 2012), que informa dels gens solapats amb pics identificats (amb distància màxima de 100 pb) i també el gen més proper a cada pic. Els solapaments entre proteïnes i els diagrames de Venn es van generar amb Vennable (Swinton, 2011).

El package rGadem es va utilitzar per fer les cerques de motifs entre les seqüències genòmiques, amb les opcions per defecte del software. Els logos es van generar amb el package SeqLogo.

Els gens amb promotors alternatius es van identificar com aquells que tenen més d'un TSS diferent a més d'1Kb de distància en el UCSC dm3. Un test de chi-quadrat va ser usat per estimar la significativitat estadística de l'enriquiment de gens amb binding sites entre gens amb promotor alternatiu. Un test amb gens seleccionats a l'atzar (1000 permutacions) es va realitzar per refusar significància d'enriquiment de gens random en promotors alternatius.

Les regions amb els promotors head-to-head (HH) o tail-to-tail (TT) es van classificar amb el genoma dm3 de UCSC. Per estimar la significància estadística de l'enriquiment HH o TT contra el que s'esperaria per atzar, es van simular binding sites (1000 permutacions) de manera que poguessin seguir la mateixa proporció per cromosoma com els experiments reals i la mateixa distribució d'amplada de pics. Per cada permutació es va computar la proporció de solapaments dels llocs a l'atzar amb els HH i TT i se'n va treure un p-valor.

3.2.7 Anàlisi de microarray d'expressió

3.2.7.1 RNAi

En plaques de sis pous, posem 2.5 milions de cèl·lules en 2.5 mL per cada pou. Afegim 15 ug de dsRNA a les 0 hores, 48 hores i a les 72 hores i recollim a les 96 hores. A les 72 hores, abans d'afegir-lo, diluïm les cèl·lules a la meitat. Extraïem l'RNA del cultiu i n'utilitzem 1 ug pel microarray.

3.2.7.2 Microarrays

La preparació de la llibreria de cDNA i la hibridació amb el microarray es van fer al Functional Genomics Core de l'Institut de Recerca Biomèdica de Barcelona (IRB). La preparació i amplificació de la llibreria de cDNA es van dur a terme partint de 25 ng d'RNA total amb WTA2 (Sigma-Aldrich) i 19 cicles

d'amplificació. 8 µg de cDNA van ser subseqüentment fragmentats amb DNAsal i biotinilitzats amb transferasa terminal obtinguda de GeneChip Mapping 10Kv2 Assay Kit (Affymetrix). La mescla d'hibridació es va preparar d'acord amb el protocol d'Affymetrix i es va hibridar la mostra a un GeneChip Drosophila Genome 2.0 Array d'Affymetrix. Els arrays van ser escanejats amb un GeneChip Scanner 3000 (Affymetrix). Els arxius CEL es van generar d'arxius DAT amb el software GCOA (Affymetrix). Per generar les estimacions d'expressió log₂, la intensitat global de l'array es va normalitzar entre arrays així com la intensitat de les sondes utilitzant l'algoritme Robust Multichip Average, RMA (Irizarry et al., 2003).

3.2.9 Anàlisi de retard d'electromobilitat (EMSA)

En primer lloc cal hibridar els oligonucleòtids complementaris per fer les sondes de DNA doble cadena. Posem 5 pmols de cada oligo en 200 uL de biffer d'hibridació. Ho posem al bany a 100°C i deixem que es refredi durant tota la nit.

Després hem de marcar les sondes radioactivament. Posem 1.5 pmol d'oligos hibridats amb 3.3 pmols de dATP amb fòsfor marcat (α^{32} P), una barreja de dCTP, dGTP i dTTP a 1mM i 1 uL de polimerasa Klenow. Ho deixem 15 minuts a 30°C. Al final afegim una barreja de dNTPs 1mM i 15 minuts a temperatura ambient. Ho purifiquem amb una columna Micro-Spin i finalment diluïm afegint 70 uL de buffer d'hibridació als 34 uL.

Llavors incubem la proteïna amb el DNA. La proteïna l'hem obtingut d'expressar en bacteris un plàsmid induïble amb IPTG que conté el CDS de cada proteïna i que després purifiquem utilitzant un tag d'histidines i una columna de níquel (Sambrook et al., 2001). Posem 5 uL de l'extracte, 1 uL de DNA d'esperma de salmó a 0.1 uG/uL, 1 uL de BSA 10ug/uL, 1uL de la sonda radioactiva i buffer de binding fins a 10uL. Ho incubem 15 minuts a 4°C.

El gel és del 8% d'acrilamida en TBE 0.25X, de 0.75 mm de gruixudesa. Abans de carregar les mostres fem un pre-run a 200 V i 4°C fins que l'amperatge baixa a la meitat de l'inicial. Llavors carreguem les mostres incubades, i corre a 200V. Un cop corregut s'asseca el gel a l'assecador 30 minuts i es revela a la pantalla radioactiva Storage Phospho Screen (Molecular Dynamics). Finalment s'obté la imatge revelant la pantalla al Phosphoimager Typhoon 8600 (Molecular Dynamics).

4.- RESULTATS

4.1 Anàlisi funcional dels *insulators* del Complex

Bitòrax

4.1.1 Elements reguladors entre l'iab-5 i l'iab-6

Els assaigs d'hipersensibilitat de la cromatina a les nucleases (DNAsaI o nucleasa micrococcal) són una manera de detectar zones de baixa densitat de nucleosomes, o directament de regions lliures de nucleosomes, regions que estan directament relacionades amb el mecanisme que permet que diferents factors d'unió al DNA, es tracti de maquinària transcripcional, de silenciadors com Polycomb, d'activadors com Trithorax, o de proteïnes d'*insulator* puguin accedir al DNA ja que la cromatina es troba en un estat "obert" i per tant es genera l'espai físic on aquestes proteïnes s'emplacen per dur a terme la seva funció.

Així, aquesta hipersensibilitat a la digestió amb nucleases és una de les característiques que ha permès identificar els diferents elements reguladors -ja siguin *enhancers*, *insulators* o PREs- presents a la regió no codificant que s'extén entre l'extrem 3' del gen Abdominal-B i el 5' d'Abdominal-A al cromosoma 3 de *Drosophila melanogaster*, dins del Complex Bitòrax.

En aquest context l'element més vastament estudiat és el Fab-7, situat entre els dominis iab-6 i iab-7, i que conté seqüències adjacents corresponents a activitat *insulator* i a PRE. En efecte al Fab-7 s'hi han identificat quatre llocs hipersensibles (Karch et al., 1994) dos dels quals (HS1 i HS2) han estat descrits com a *insulator* degut a la seva activitat *enhancer-blocking* (Hagstrom et al., 1996) i un tercer, al costat 5' d'aquests dos, com a PRE (Mishra et al., 2001). A 3' de l'*insulator* s'hi troba el quart lloc hipersensible, minoritari i no sempre detectable, que ha estat descrit com a *Promoter Targeting Sequence* o PTS (Chen et al., 2005). Aquest tipus de seqüències serien capaces de sobrepassar un *insulator* i per tant d'aproximar un *enhancer* distal al promotor d'un gen encara que hi hagués un *insulator* entremig. Es tracta d'una hipòtesi que explicaria com

els *enhancers* de les regions iab-6, iab-5 o iab-4 són capaces d'establir contacte amb el promotor d'Abdominal-B tenint en compte que enmig hi ha el Fab-7 i el Fab-8.

No només al Fab-7 estan ben estudiades aquestes regions hipersensibles, sinó també altres elements reguladors de l'expressió d'Abdominal-B com Mcp (Karch et al., 1994) o Fab-8 (Barges et al., 2000). De la mateixa manera altres *insulators* estudiats al genoma de *Drosophila melanogaster* són hipersensibles a nucleases, com l'element scs (Vazquez and Schedl, 1994) o elements *insulator* de vertebrats com al locus de la beta-globina de pollastre (Chung et al., 1993).

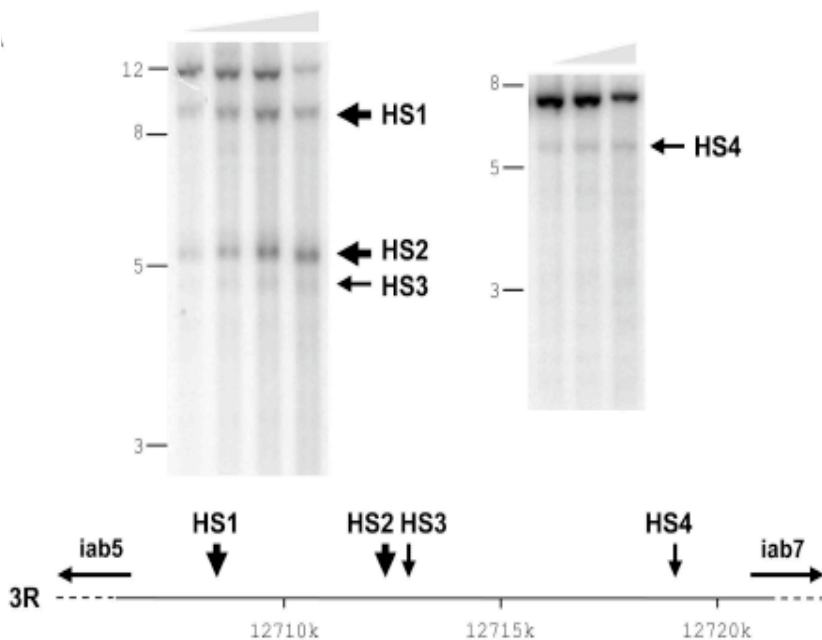


Figura 17.- Els llocs hipersensibles a DNAsal a l'iab-6. Es va digerir DNA de nuclis d'embrions de 0 a 16 hores amb la DNAsal i posteriorment amb KpnI (esquerra) o HindIII (dreta). A l'esquerra es mostren els pesos moleculars en kb i a dalt el triangle indica concentracions creixents de DNAsal. Amb fletxes cadascun dels quatre llocs identificats, i a sota un esquema de la seva localització al llarg de l'iab-6.

És per tot això, i pel desconeixement existent fins aquell moment sobre els elements reguladors presents entre l'iab-5 i l'iab-6 que, en el marc dels estudis duts a terme per la doctora Sílvia Pérez-Lluch, es va realitzar una cerca de dominis hipersensibles a la DNAsal en la regió (Pérez-Lluch et al., 2008). Utilitzant com a material embrions de 0 a 16 hores, l'assaig va revelar l'existència de dos llocs hipersensibles principals (HS1 i HS2) i dos de menors (HS3 i HS4), disposats en l'ordre indicat en la figura 17.

L'únic element regulador que havia estat identificat en aquesta zona amb anterioritat és l'*enhancer* IAB-6 que dirigeix l'expressió d'Abdominal-B al para-segment 11 de l'embrió (Mihaly et al., 2006). La localització descrita d'aquest *enhancer* IAB-6 sembla correspondre clarament amb la regió identificada en l'assaig d'hipersensibilitat com a HS4. Per tant queda per determinar quina funció tenen les altres tres regions hipersensibles.

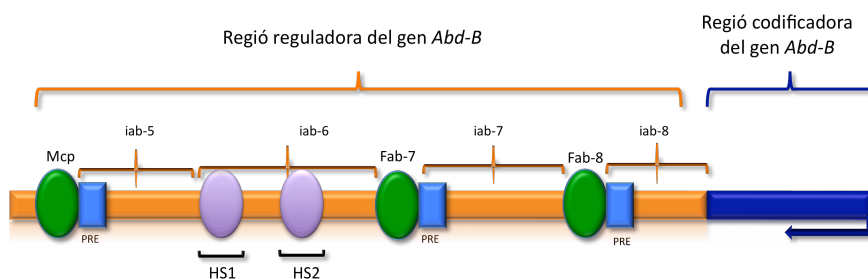
4.1.1.1 Presència de Polycomb i GAGA a les regions hipersensibles

Per tal d'elucidar la funció de les noves regions reguladores identificades vam realitzar experiments d'immunoprecipitació de cromatina. Ja era conegut que algunes proteïnes del grup Polycomb s'uneixen als elements Mcp i Fab-7 (Schwartz et al., 2006). A part del fet que els elements front-abdominals que es coneixien fins al moment contenien tots un PRE, les regions hipersensibles HS1 i HS2 tenen seqüències consens d'unió de proteïnes com Pleiohomeotic (una de les proteïnes del Grup Polycomb) i GAGA. Aquest és un factor que s'uneix al DNA i que ha estat relacionat tant amb processos d'activació (es va definir en un principi com a membre del Grup Tritòrax) com de silenciament, ja que s'ha vist que GAGA col-localitza amb Polycomb a diferents PREs (Strutt et al., 1997) i que els seus llocs d'unió són necessaris per la seva activitat de silenciament (Busturia et al., 2001).

Utilitzant embrions com a material de partida vam realitzar immuno-precipitacions de cromatina seguit de PCRs amplificant regions reguladores d'interès (figura 2). Així vam comprovar que en efecte hi havia presència de la proteïna Polycomb (Pc) i de GAGA tant a l'HS1 com a l'HS2. També vam confirmar els resultats ja coneguts, i com a control positiu dels nostres experiments, amb la presència d'ambdues proteïnes als altres elements reguladors d'Abdominal-B: Mcp, Fab-7 i Fab-8. Com a control negatiu es va utilitzar una regió del gen de la proteïna SAP on no hi hauria de ser cap de les dues proteïnes. Paral·lelament es van verificar aquests resultats per qPCR (Pérez-Lluch et al., 2008).

A la figura 18-b es mostra el gel d'agarosa amb els productes de les PCRs de cada regió. Es veu que hi ha immunoprecipitació a totes les regions reguladores amplificades. En canvi no n'hi ha al control negatiu ni al carril de la mostra sense anticòs. Per tant concloem que hi ha presència de Polycomb i GAGA a HS1 i HS2 de la mateixa manera que n'hi ha a Mcp, Fab-7 i Fab-8.

A)



B)

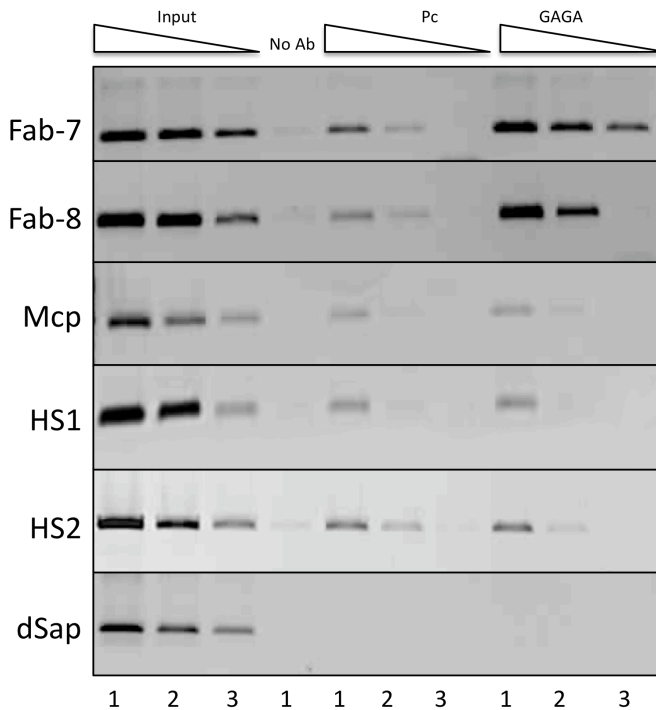


Figura 18.- Immunoprecipitacions de cromatina a la regió reguladora d'AbdB. A) Esquema de la localització dels diferents elements reguladors d'AbdB que testem per ChIP; B) la cromatina crosslinkada es va immunoprecipitar amb anticòs de Polycomb (Pc), GAGA i sense anticòs (No Ab) i analitzada per PCR estandard utilitzant primers específics per a cada regió. Les PCRs es van realitzar amb mostra sense diluir, diluïda 1/5 i diluïda 1/25. Els carrils de l'input corresponen als productes de PCR obtinguts abans de la immunoprecipitació a partir del 10% del material a immunoprecipitar.

4.1.1.2 Activitat *enhancer-blocking* de les regions hipersensibles

Amb l'objectiu d'aprofundir en la funció d'aquestes noves regions reguladores, vam generar línies de *Drosophila white⁻* (sense color d'ulls endogen) que incorporaven al seu genoma construccions ectòpiques que contenen un enhancer i un gen reporter mini-white (Figura 19). En una construcció on no hi ha insulator entre enhancer i promotor els ulls són de color vermell. Si hi

insertem un element regulador entremig aquest pot afectar a l'acció que té l'enhancer sobre el promotor del gen. És doncs el nivell d'expressió del gen mini-white ectòpic el que indicarà l'efecte que té sobre ell l'element regulador: des de blanc –repressió total- fins a vermell si no hi ha cap tipus de repressió.

Si l'element regulador té activitat PRE s'espera que hi hagi com a mínim un cert grau de repressió, és a dir que l'ull no serà vermell sinó que tindrà com a molt fosc un color ataronjat. A més, una característica particular dels PREs és que presenten pairing-sensitive silencing (PSS), el que significa que quan hi ha dues còpies en homozigosi de la construcció ectòpica en lloc de sumar-se les expressions dels dos mini-white i incrementar proporcionalment la pigmentació, passa just el contrari. És a dir que el color d'ulls dels homozigots serà més clar que el dels heterozigots.

D'aquesta manera vam generar tres construccions: una en la que l'element regulador era el Fab-7, una altra amb l'HS1 i l'altra amb la suma de l'HS2 i del minoritari HS3, que en el genoma estan l'un al costat de l'altre. Totes tres presentaven silenciament. En el cas de Fab-7 un 50% de línies estaven reprimides i d'aquestes un 60% presentaven PSS, resultats similars als ja publicats (Hagstrom et al., 1997). En el cas de l'HS1 un 40% estaven reprimides i un 50% d'aquestes feien PSS, i en H2+3 un 40% reprimides i un 40% PSS (Pérez-Lluch et al., 2008). Per tant de la mateixa manera que passa amb el Fab-7, les regions hipersensibles aquí descrites tenen activitat PRE.

La següent pregunta era si, a més de PRE, alguna d'aquestes seqüències contenen un *insulator*, tal i com es preduria del fet que entre tots els altres iab se n'ha trobat un excepte en aquesta regió entre l'iab-5 i l'iab-6. Els *insulators* tenen activitat *enhancer-blocking*, el que vol dir que són capaços de bloquejar l'acció d'un *enhancer* sobre un promotor i així anul·lar o si més no reduir l'activitat transcripcional d'un gen. Per tant en el cas de les tres construccions amb les que treballem això significaria que el color d'ulls seria d'un color com a molt fosc ataronjat. El problema és que aquest efecte es confon amb el del PRE,

que tot i que el mecanisme és diferent (el PRE no bloqueja l'enhancer sinó que silencia la regió a través del reclutament del complex Polycomb) el resultat a nivell observable és el mateix. I el cert és que aquestes regions hipersensibles poden tenir les dues activitats, PRE i *insulator* com a Mcp, Fab-7 i Fab-8.

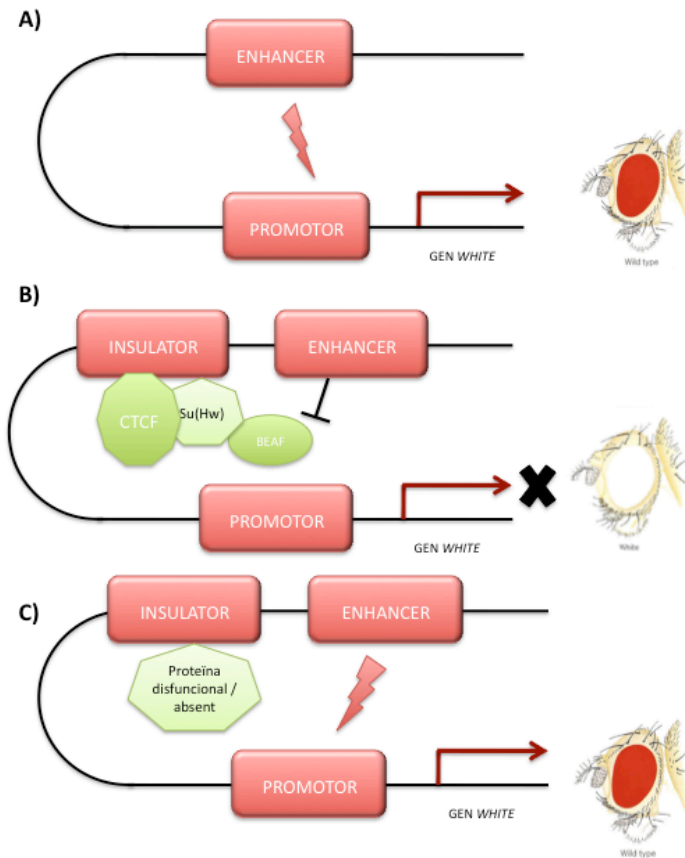


Figura 19.- Assaig d'enhancer-blocking en constructes transgènics amb el gen reporter mini-white.

A) Si no hi ha un insulator, l'enhancer activa el promotor del mini-white i la mosca presenta ulls vermells; B) si hi ha un insulator entre mig, les proteïnes d'insulator s'hi uneixen i bloquegen l'acció de l'enhancer i això es reflecteix en els ulls de la mosca que perden pigment; C) si alguna de les proteïnes necessàries per l'activitat insulator està mutada, aquest no bloquejarà (o bloquejarà pitjor) l'enhancer i la mosca recuperarà el color d'ulls.

Les seqüències *insulator* funcionen reclutant alguna o algunes proteïnes com CP190 o CTCF, que són les que d'alguna manera els confereixen l'activitat

enhancer-blocking. És el cas per exemple del Fab-7 que necessita reclutar ambdues proteïnes (Moon et al., 2005; Mohan et al., 2007). Per tant és d'esperar que si alguna de les construccions contenia un *insulator* s'incrementaria la pigmentació dels ulls en cas de portar una mutació d'una proteïna d'*insulator*. Vam creuar mosques que tenien les construccions Fab-7, HS1 i HS2+3 amb mosques mutants de CP190 (CP190^{H31.2}).

Amb una sola còpia de la mutació CP190^{H31.2} ja es pot observar un increment de la pigmentació dels ulls en la construcció Fab-7 (figura 20), confirmant que es tracta d'un *insulator* a més d'un PRE i demostrant que necessita CP190 pel seu correcte funcionament. El mateix resultat s'observa en la línia que porta la construcció HS1, mentre que no s'observa cap canvi en la construcció HS2+3. Aquests resultats indiquen que la regió HS1 conté un PRE i un *insulator*, mentre que la HS2+3 només tindria activitat PRE.

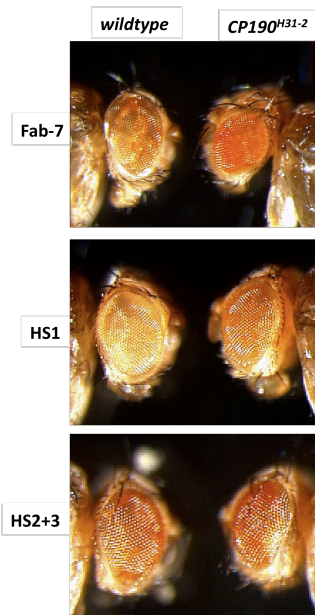


Figura 20.- La mutació de CP190 afecta a l'HS1 però no a l'HS2. Els ulls de les mosques de la dreta (mutants CP190^{H31.2}) augmenten la pigmentació en el cas del Fab7 i en l'HS1, però no a HS2.

Així vam realitzar el mateix experiment creuant la construcció HS1 amb un mutant de dCTCF ($CTCF^{y+6}$). La descendència, amb només una còpia mutant de dCTCF, presentava un lleuger increment en la pigmentació ocular (figura 21-a). Això confirma la presència d'un *insulator* que és dependent tant de dCTCF com de CP190.

Posteriorment vam introduir elements FRT a banda i banda dels elements reguladors en les construccions. Aquests elements són reconeguts per l'enzim flipasa i s'excideix tota la regió que hi ha entre els dos. A la figura 21-b observem que en creuar la línia HS1 amb línies portadores de flipasa la descendència augmenta del tot la pigmentació dels ulls, confirmant que efectivament el color ataronjat es devia a la presència de l'element PRE/*insulator* HS1. Aquest doncs correspon a l'insulator Fab-6, que separa els dominis iab-5 i iab-6 del Complex Bitòrax.

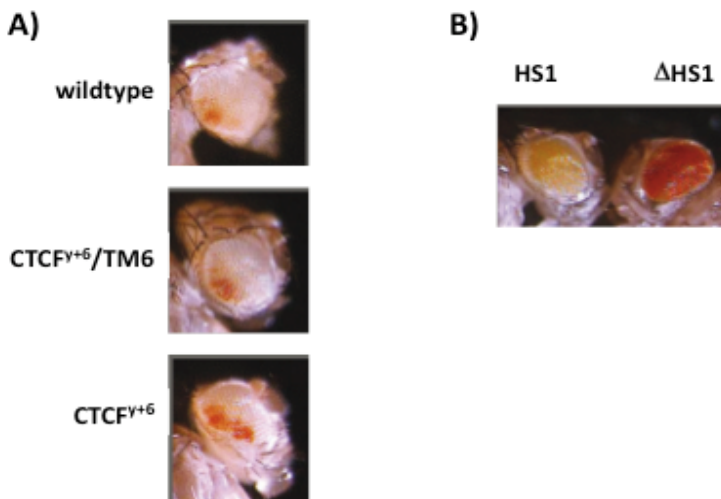


Figura 21.- La mutació de CTCF afecta a l'HS1. A) L'ull wild-type té una mica de pigmentació a l'ull, que augmenta amb un mutant heterozigot de CTCF ($CTCF^{y+6}/TM6$) i encara més amb un homozigot $CTCF^{y+6}$. B) Quan excidim l'HS1 de la construcció per acció de la flipasa, la pigmentació de l'ull passa de groc pàlid a vermell, demostrant que l'HS1 estava bloquejant la plena expressió del mini-white.

4.1.2 Implicació de JIL-1 en la regulació d'Abdominal-B

JIL-1 és una quinasa nuclear de *Drosophila melanogaster* que s'associa a la cromatina en tots els estadis del desenvolupament. És responsable de la major part de la fosforilació de la serina 10 de la histona H3 (H3S10p) a interfase (Wang et al., 2001). En cèl·lules de mamífer la marca H3S10p s'ha vist implicada en activació transcripcional. En el cas de *Drosophila*, la unió de JIL-1 a politènics es produeix majoritàriament en interbandes, relacionades amb cromatina descondensada transcripcionalment activa (Jin et al., 1999). A més, el reclutament ectòpic de JIL-1 fusionada a LacI a repeticions LacO porten a una descondensació local de la cromatina en politènics (Bao et al., 2008). JIL-1 té un rol essencial en organització cromosomal encarregant-se de mantenir el balanç entre eucromatina i heterocromatina (Deng et al., 2008).

El fet que JIL-1 pugui reforçar l'estat descondensat i actiu de cromatina i que a més participi en l'organització cromosomal ens va fer pensar que potser podria ser important per les regions reguladores del Complex Bitòrax. En efecte, aquestes regions –que comprenen com hem demostrat tant *insulators* com PREs- no només presenten un estat descondensat de la cromatina, com s'observa a partir dels assaigs d'hipersensibilitat a nucleases, sinó que a més es pensa que actuen a través de contactes intra-cromosomals, reorganitzant la cromatina a nivell local a través de contactes amb *enhancers* i amb el promotor –en el cas del nostre model d'estudi- d'Abdominal-B (Lanzuolo et al., 2007).

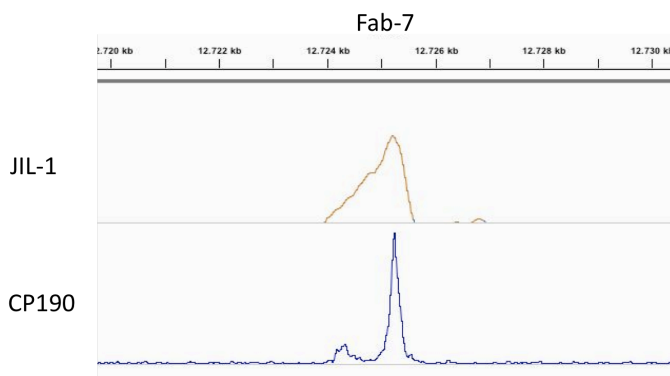
4.1.2.1 JIL-1 és necessària per l'activitat *insulator*

Per analitzar la possible funció de JIL-1 en els *insulators* hem fet servir una línia mutant (Z60) on el gen de JIL-1 té una deleció a la seva regió 5', li falta el principi del gen, el promotor i també el gen adjacent CG6279. Aquesta deleció provoca que la viabilitat del homozigots es redueixi a un 1.7% i que els nivells de proteïna siguin del 0.3% respecte el WT (Zhang et al., 2003).

A la base de dades ModEncode es pot comprovar com JIL-1 es troba unit a la cromatina en la regió Fab-7 en cèl·lules S2 (Figura 22-a). A escala genòmica s'observa, també a través dels ChIPs publicats a ModEncode, com hi ha un fort grau de col·localització entre JIL-1 i els llocs d'unió de proteïnes d'*insulator* (Celniker et al., 2009). De 6017 llocs d'unió de JIL-1 a la cromatina, 2309 coincideixen amb llocs d'unió de proteïnes d'*insulator*, és a dir un 38% dels llocs on JIL-1 s'uneix a la cromatina són llocs on les proteïnes d'*insulator* també ho fan.

Vam crear la línia mutant JIL-1^{Z60} amb les línies que portaven les construccions Fab-7, HS1 o Fab-6 i HS2+3 o PRE-6. El que vam observar (Figura 22-b) en la descendència és que hi havia un augment de la pigmentació dels ulls en aquelles mosques que portaven una còpia de la mutació Z60 respecte les que expressaven JIL-1 de manera wild-type només en el cas de les Fab-7 i Fab-6, que contenen un *insulator*, però no en el cas de la línia HS2+3, que com hem vist només té activat PRE. Això significa que d'alguna manera JIL-1 és necessari pel correcte funcionament dels *insulators* Fab-6 i Fab-7. En global aquests resultats suggereixen un rol de JIL-1 en el correcte funcionament dels *insulators* a *Drosophila*.

A)



B)

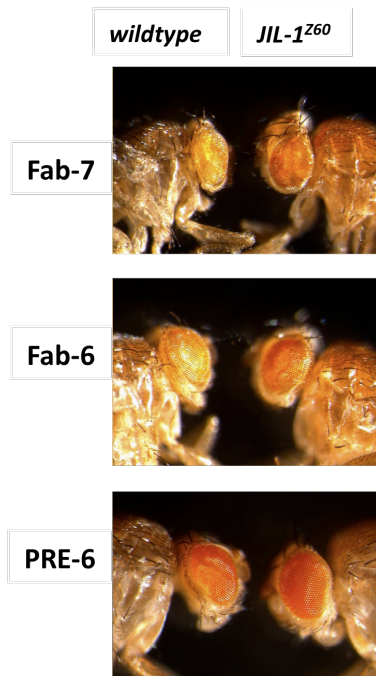


Figura 22.- La mutació de JIL-1 afecta el funcionament de Fab-7 i HS1. A)Amb les dades del Modencode veiem que hi ha enriquiment de JIL-1 i de CP190 a la zona on hi ha el Fab-7; B) La pigmentació dels ulls de les línies amb construccions Fab-7 i HS1 augmenta en mutants de JIL-1, mentre que no ho fa en les línies amb construccions HS2.

4.1.2.2 Fosforilació de la histona H3 als *insulators*

JIL-1 és, com s'ha dit, una quinasa que fosforila la serina 10 de la histona H3. Un possible mecanisme doncs que podria fer que JIL-1 afectés l'activitat *insulator* és a través d'aquesta marca, que posteriorment podria reclutar factors que mantindrien un estat obert de la cromatina, característica dels *insulators*.

Donat que no hi ha dades publicades a escala genòmica de la presència d'H3S10p a la cromatina de *Drosophila*, vam decidir realitzar

immunoprecipitacions de cromatina utilitzant un anticòs específic d'aquesta marca a partir d'embrions. A la figura 23 es mostren els resultats de les PCRs fetes sobre el material immunoprecipitat al Fab-6 i al Fab-7 i com a control negatiu el gen dSap. Encara que aquestes immunoprecipitacions van ser força variables i que l'anticòs molt sovint immunoprecipitava DNA inespecífic, s'observa presència d'H3S10p als dos *insulators*, suggerint que efectivament aquesta és la via d'actuació de JIL-1 en aquestes regions reguladores.

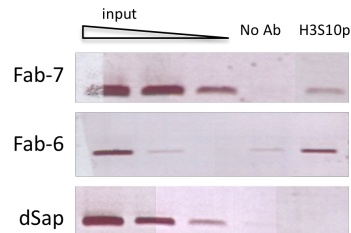


Figura 23.- Immunoprecipitacions de cromatina d'H3S10p. La cromatina crosslinkada es va immunoprecipitar amb anticòs de la histona H3 fosforilada a la serina 10 (H3S10p) i sense anticòs (No Ab) i analitzada per PCR estandard utilitzant primers específics per a cada regió. Les PCRs de l'input es van realitzar amb mostra sense diluir, diluïda 1/5 i diluïda 1/25. Els carrils de l'input corresponen als productes de PCR obtinguts abans de la immunoprecipitació a partir del 10% del material a immunoprecipitar.

4.1.3 Trànscripts no codificants al Fab-6 i Fab-7

Els primers estudis que es van fer sobre transcripció no codificant al Complex Bitòrax ja descobrien extensa transcripció al llarg de regions intergèniques als inicis del desenvolupament (Sánchez-Herrero and Akam, 1989). La transcripció a través d'aquestes regions intergèniques està restringida a nivell espacial al llarg de l'eix antero-posterior i s'inicia abans de l'activació dels gens Hox (Bae et al., 2002). Com que aquestes unitats transcripcionals es troben presents només en aquells para-segments on l'expressió del gen homeòtic és controlat pel seu domini, es creu que la generació de trànscripts a través d'aquestes regions es correspon amb l'obertura seqüencial de la cromatina a cada iab per tal d'activar la transcripció del gen homeòtic. Aquesta transcripció intergènica no s'observa només als iab sinó també a través dels elements reguladors que els separen (Bae et al., 2002). En efecte, hibridacions *in situ* mostren com trànscripts de l'Mcp o del Fab-7 s'expressen en aquells para-segments que corresponen al domini iab que controlen (Rank et al., 2002).

Nosaltres hem volgut verificar que aquesta transcripció detectada a nivell d'elements reguladors també es dona al nou element Fab-6. Vam extreure RNA de larva III i el vam retrotranscriure amb primers específics que ens permetien discriminar l'orientació del trànscrit. Després l'amplificavem per PCR quantitativa i comparavem els valors agafant com a referència un gen house-keeping com l'*rp49*, que no té uns valors d'expressió molt alts i són aproximadament del mateix ordre de magnitud que els dels trànscripts que nosaltres mirem. El que observem (Figura 24) és que detectem transcripció *sense* (que significa en la mateixa orientació de transcripció que Abdominal-B) tant al Fab-7 com al Fab-6. En tots dos casos la transcripció *sense* és molt major que l'*anti-sense*, en coherència amb tots els treballs publicats anteriorment.

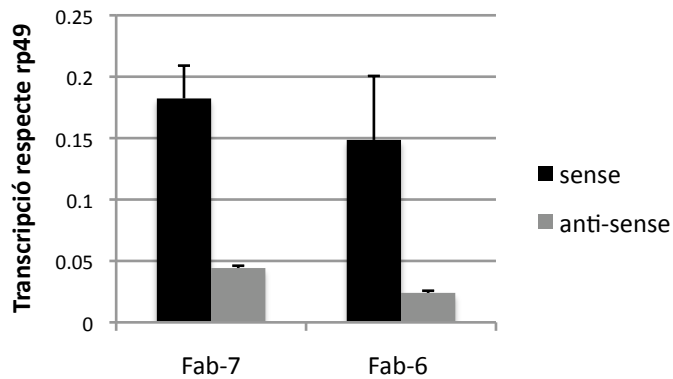


Figura 24.- Transcripció al Fab-6 i Fab-7 a larva. Les columnes indiquen el valor relatiu de cada trànscrip respecte el nivell de transcripció del gen rp49, calculat amb PCR real-time. Les columnes en negre és el trànscrip sense, en el mateix sentit de transcripció que Abd-B. En gris anti-sense, és a dir transcrits en direcció oposada a com ho fa Abd-B.

Hem estudiat quin és l'estat transcripcional de Fab-6 i Fab-7 durant el desenvolupament embrionari (Figura 25). El resultat és similar a l'observat en larves: hi ha transcripció sense en tots dos elements reguladors en el mateix sentit de transcripció que Abd-B, tot i que molt més forta en el Fab-7 que en el Fab-6.

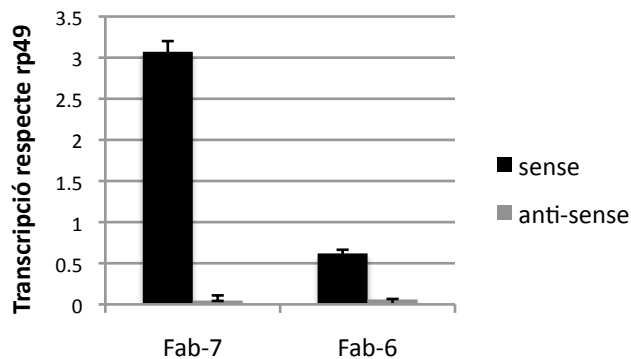


Figura 25.- Transcripció al Fab-6 i Fab-7 a embrions. Les columnes indiquen el valor relatiu de cada trànscrip respecte el nivell de transcripció del gen rp49, calculat amb PCR real-time. Les columnes en negre és el trànscrip sense, en el mateix sentit de transcripció que Abd-B. En gris anti-sense, és a dir transcrits en direcció oposada a com ho fa Abd-B.

Tant a embrions com a larva ens trobem amb la limitació que la transcripció que observem és de fet una barreja de tots els estats transcripcionals de tots els teixits de l'organisme, tant aquells que expressen Abdominal-B com els que no. Decidim utilitzar cèl·lules en cultiu per la seva homogeneïtat transcripcional respecte un embrió o una larva sencera. En el cas de les cèl·lules de *Drosophila*, sabíem per observacions prèvies que la línia S2 no expressa el gen Abdominal-B mentre que la línia SG4 sí que ho fa. A part de confirmar-ho per RT-qPCR hem mirat què passava al Fab-7. El que s'observa és també transcripció només en les SG4 i no a les S2 (Figura 26), per tant concloem que existeix una correlació entre transcripció a l'element regulador i l'expressió del gen.

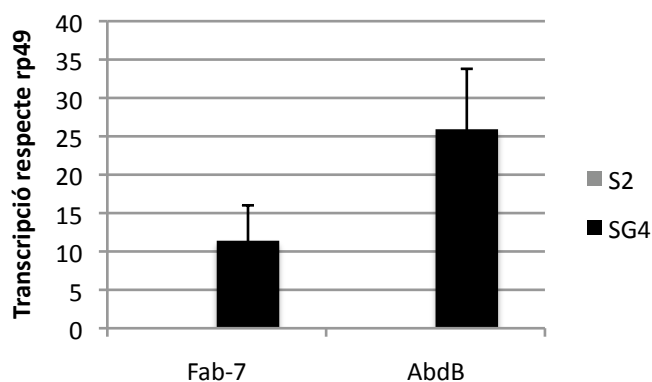


Figura 26.- Transcripció d'Abd-B i Fab-7 a cèl·lules en cultiu S2 i SG4. Les columnes indiquen el valor relatiu de cada trànscrip respecte el nivell de transcripció del gen rp49, calculat amb PCR real-time. Les columnes en negre és el trànscrip a S2, en gris SG4. Els nivells de trànscrip detectats a S2 —tant d'Abd-B com Fab-7— són molt a prop de zero en relació a l'expressió de l'rp49. Per contra, a SG4 són molt més alts, 10 cops més de Fab-7 (sense) i 25 cops més d'Abd-B.

Hem volgut investigar també quina contribució fan les proteïnes d'*insulator* com CTCF a aquest procés transcripcional, ja que no hi ha res estudiat al respecte. De la mateixa manera ens hem fet la mateixa pregunta pel que fa a JIL-1. Sabem pels nostres experiments que JIL-1 contribueix al correcte funcionament de Fab-6 i Fab-7, i sabem per la literatura que JIL-1 es

correlaciona, a través de la fosforilació que deixa a la histona H3, amb activitat transcripcional. Per tant la nostra hipòtesi és que per tal que hi hagi transcripció a través dels elements reguladors cal que JIL-1 hi sigui present per mantenir un estat descondensat de la cromatina.

A la figura 27 es mostren els resultats de RT-qPCR utilitzant larves III mutants de dCTCF i de JIL-1. En aquest cas només s'ensenya el transcrit sense relatiu al gen house-keeping actina i normalitzat a la transcripció del WT. S'observa que hi ha una decaiguda en els nivells de transcrit quan manca dCTCF i que aquesta decaiguda és encara més intensa quan manca JIL-1, d'acord amb la nostra hipòtesi.

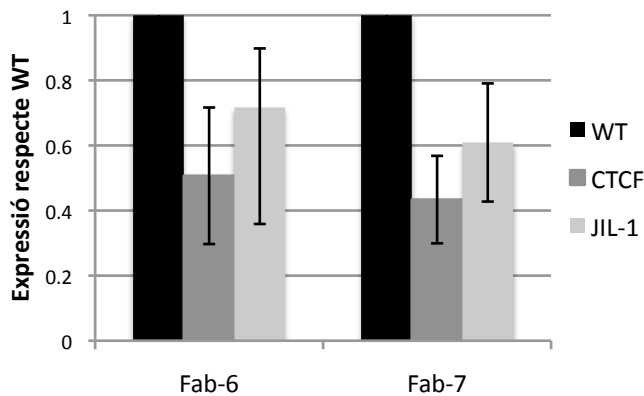


Figura 27.- Transcripció a larves wild-type i larves mutants de JIL-1 i CTCF. Es mostren els nivells de transcrit detectats en relació a l'rp49 normalitzats pels valors obtinguts a larves wild-type.

4.2 Identificació i anàlisi de noves proteïnes involucrades en l'activitat *insulator*

4.2.1 El complex proteic de CP190

CP190 és una proteïna que té un rol central en el funcionament i organització dels *insulators* a *Drosophila*. Els seus dominis BTB i Zinc finger li confereixen la capacitat d'unir-se tant al DNA com a altres proteïnes. Tanmateix l'hipòtesi regnant fins al moment és que no s'uneix al DNA directament sinó que ho faria a través de les demás proteïnes d'*insulator*, que són les que tindrien especificitat de seqüència. Les dades existents demostren que les proteïnes BEAF, Su(Hw) i dCTCF s'uneixen a diferents llocs del genoma amb una colocalització parcial entre elles, essent comú a totes elles una forta colocalització amb CP190 (Bushey et al., 2009). És a dir que els estudis fets fins al moment suggereixen un model en el que CP190 seria una proteïna comú a tots els tipus d'*insulators*.

Amb l'objectiu d'identificar noves proteïnes que puguin ser necessàries pel correcte funcionament dels *insulators*, decidim purificar el complex del que forma part CP190 i analitzar amb quines proteïnes interacciona de manera estable. Per a això expressem CP190 fusionada a un tag en cèl·lules S2, en purifiquem primer l'extracte nuclear i posteriorment la proteïna fusionada al tag, i finalment identifiquem per espectrometria de masses les proteïnes que han estat purificades amb CP190.

4.2.1.1 Expressió induïda de la proteïna en cultiu cel·lular

En primer lloc vam generar el constructe que havia de servir per expressar CP190 a cèl·lules. Vam escollir utilitzar el tag TAP, que conté la proteïna A i el *Calmodulin binding peptide* (CBP) separats per un lloc de restricció per la proteasa TEV (Figura 28). Aquest tag va ser dissenyat en un origen per a dur a terme una purificació en dos passos: en un primer pas es purifica l'extracte

fent servir l'afinitat de la proteïna A per les IgG, després s'elueix amb la proteasa TEV que talla el tag fent que només li quedi el CBP (Puig et al., 2001). Llavors es passa per una columna de calmodulina i s'elueix afegint EGTA, que quela les molècules de calci que permetien l'unió a la calmodulina. L'objectiu de fer la purificació en dos passos és reduir la quantitat de contaminants i per tant augmentar l'especificitat de la purificació. Per contra, això redueix significativament la quantitat de proteïna obtinguda al final del procés. Per les proves que vam fer, vam decidir no fer servir els dos passos sinó només la purificació amb les IgG. El motiu és que nosaltres necessitàvem recuperar la major quantitat possible de proteïna per a poder identificar per espectrometria de masses totes les proteïnes del complex, que podien ser molt menys abundants que el *bait*, CP190. Per evitar tenir excessius contaminants, que també poden posar en perill l'identificació de proteïnes específiques, vam utilitzar de partida l'extracte nuclear i no pas l'extracte cel·lular total, ja que la fracció citoplasmàtica hagués pogut arrossegar contaminants i en canvi, per la naturalesa del complex que volíem estudiar, no ens aportava res.

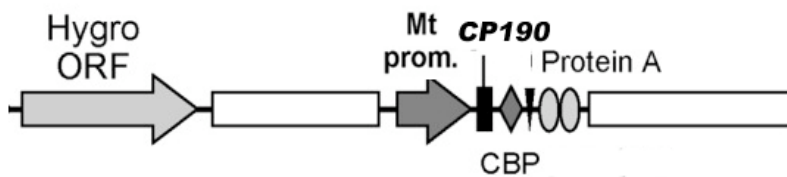


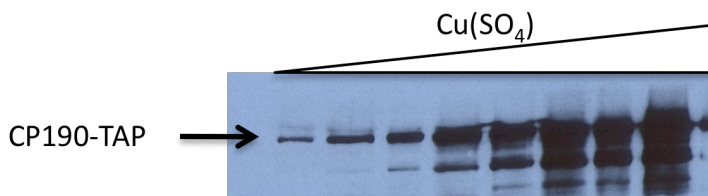
Figura 28.- Constructe per expressar CP190-TAP a cèl·lules. Hem insertat el CDS de CP190 a N-terminal del TAP, que està format per un CBP i la proteïna A. La proteïna de fusió està sota el control d'un promotor de la metal·lotioneïna. La selecció de les cèl·lules que integren el constructe es fa gràcies a la presència d'un gen de resistència a la higromicina.

Així, vam clonar al plàsmid pMK33-CTAP el CDS de CP190, de manera que el TAP quedava a l'extrem C-terminal de la proteïna. Al cantó 5' del gen hi ha un promotor de la metal·lotioneïna que s'indueix en presència de coure. Això evita els possibles efectes adversos sobre la viabilitat cel·lular que pugui tenir

l'expressió constitutiva de la proteïna de fusió i a més ens permet decidir el grau d'expressió que desitgem en funció de la quantitat de coure que hi afegim. A 5' del promotor hi ha un ORF de resistència a l'higromicina que ens serveix per seleccionar les cèl·lules que han integrat el plàsmid.

Vam transfectar el plàsmid a cèl·lules S2 i per selecció amb higromicina vam generar una línia estable. A la figura 29-a es mostra com l'addició de coure (CuSO_4) induueix l'expressió de CP190-TAP proporcionalment a la quantitat de coure afegit. A la figura 29-b es mostra una immunotinció amb anticòs contra el TAP d'aquesta línia cel·lular induïda amb coure on es comprova que efectivament la proteïna de fusió localitza al nucli, de la mateixa manera que ho fa CP190 endògena.

A)



B)

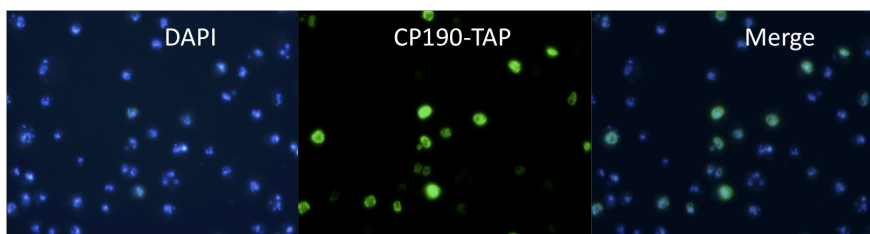


Figura 29.- Expressió de la proteïna de fusió CP190-TAP a cèl·lules. A) Western Blot amb anticòs antiTAP que demostra la presència de la proteïna de fusió a la línia cel·lular i la seva expressió dependent de la quantitat de Coure afegida; B) En una immunotinció de la línia cel·lular amb anticòs antiTAP s'observa la presència de la proteïna de fusió al nucli cel·lular.

4.2.1.2 Purificació del complex

Com s'ha dit, es va utilitzar només la fracció nuclear de les cèl·lules induïdes i es van purificar els complexos amb beads-IgG. A la figura 30 es mostra l'elució en un gel de gradient del 9 al 13% amb tinció de plata a on es pot observar CP190-TAP a dalt de tot i també tot un seguit de bandes que corresponen a proteïnes que han co-purificat. Per tal d'identificar-les totes, inclús aquelles que per abundància insuficient no apareixen al gel de plata, fem córrer l'elució en un gel d'acrilamida el temps just perquè s'apilin però no es separin. Així, tallem la banda amb les proteïnes apilades i l'analitzem per LC/MS.

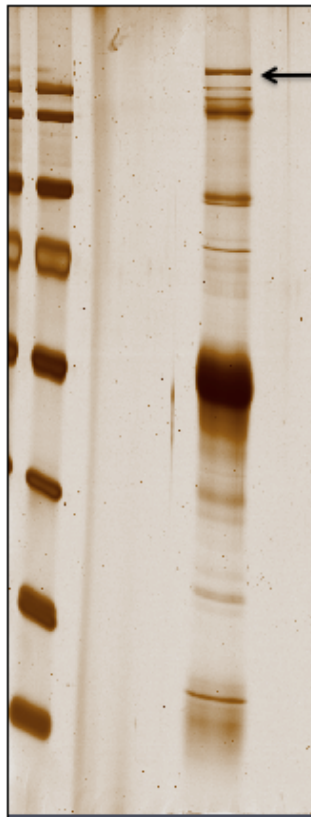


Figura 30.- Purificació del complex de CP190. Gel d'acrilamida de gradient 9-13% amb tinció de plata. S'ha corregut l'elució de la purificació de CP190. S'observen diverses bandes. Amb la fletxa s'assenyala la que correspon a CP190-TAP.

4.2.1.3 Identificació dels components del complex

L'experiment va ser fet tres vegades, i en els tres casos les principals proteïnes identificades van ser les mateixes (Taula 4). En primer lloc, i amb grans Mascot scores i molts pèptids sempre s'identificava la pròpia CP190, que ens serveix de control positiu que l'experiment i la inducció han estat correctament fetes.

D'entrada cal remarcar que hi ha algunes proteïnes la interacció de les quals amb CP190 ja havia estat demostrada anteriorment per altres laboratoris i que la seva identificació al nostre experiment corroboren o al menys reforcen la idea que la purificació del complex ha estat específica i aporta versemblança als resultats. És el cas de MAP60, que apareix en les tres rèpliques i que es va descobrir fa més de quinze anys precisament per la seva interacció amb CP190 (Kellogg et al., 1995). MAP60 és una proteïna el rol principal de la qual té lloc als centrosomes durant la mitosi (a on interacciona amb CP190) però que algunes evidències apunten a que no té un rol conjunt amb CP190 en la regulació de la cromatina a interfase (Oegema et al., 1997). Encara que això no sigui del tot descartable, nosaltres decidim no investigar més aquesta proteïna i centrar-nos en aquelles que o bé ja hi ha evidències que regulen la cromatina i per tant podrien tenir un rol als *insulators* o bé aquelles la interacció de les quals sigui novedosa.

Entre les proteïnes identificades hi ha aquelles que ja se sap que formen complex amb CP190 i que tenen activitat *insulator*: dCTCF, Mod(mdg4) i Su(Hw). N'hi ha d'altres que per contra no s'havia mostrat mai que estiguessin al mateix complex que CP190: NURF-38, Lola, Pita i Z4. Finalment, i també en totes tres rèpliques hi ha les proteïnes CG8436 i CG9740, no descrites amb anterioritat. De CG8436 se'n troben 6, 9 i 4 pèptids en cada rèplica, que significa un coverage de 31.4%, 49.2% i 29.8% respectivament, amb Mascot scores de 498.7, 610.2 i 360.1. En el cas de CG9740 se'n troben 4, 5 i 2 pèptids en cada rèplica, que significa un coverage de 37.9%, 51.8% i 16.9% respectivament, amb

Taula 4.- Identificació per espectrometria de masses de les proteïnes presents a les elucions de la purificació. Per cada proteïna identificada a cada rèplica de l'experiment se n'indiquen el nombre de pèptids identificats, quant representen del total de la proteïna (coverage) i el Mascot Score.

Proteïna	Rèplica 1			Rèplica 2			Rèplica 3		
	Pèptids	Coverage	Mascot score	Pèptids	Coverage	Mascot score	Pèptids	Coverage	Mascot score
CP190	38	56.7	2554.2	39	47.4	2676.4	14	20.4	880.1
MAP60	15	40.2	983.8	17	48	1134.9	2	9.1	108.7
CTCF	4	8.2	266.3	2	6	198.5	-	-	-
Mod(mdg4)	7	24.4	542.3	7	17.7	502.2	8	18	521.3
Su(Hw)	-	-	-	10	13	547.1	9	15	506.9
CG8436	6	31.4	498.7	9	49.2	619.2	4	29.8	360.1
CG9740	4	37.9	242.9	5	51.8	340.9	2	16.9	151.8
NURF-38	5	28.6	263.1	-	-	-	-	-	-
Z4	3	6.8	202.3	-	-	-	-	-	-
lola I	2	3.0	181.1	2	3.2	106.2	-	-	-
pita	3	6.7	162	4	10	193.3	2	5.1	99

Mascot scores de 242.9, 340.9 i 151.8. És a dir, que en ambdós casos es tracta d'una identificació prou robusta com per investigar el seu rol en el complex de CP190.

Cal dir també que s'han identificat altres proteïnes (dades no mostrades) amb molt poca robustesa i que hem considerat que molt probablement eren contaminants donat el fet que també han estat identificades en altres purificacions fetes al laboratori de proteïnes no relacionades amb CP190.

4.2.2 Anàlisi de les proteïnes identificades

4.2.2.1 Proteïnes d'*insulator*

Com s'ha explicat a la introducció, Su(Hw) és una proteïna que va ser descrita pel seu rol en l'*insulator* del retrotransposó Gypsy (Spana et al., 1988). S'uneix al DNA a través d'una seqüència específica i també s'uneix a CP190, tal com s'ha vist per assaig de dos híbrids i per co-immunoprecipitació (Pai et al., 2004; Gerasimova et al., 2007). A part de la colocalització a politènics, s'ha vist que Su(Hw) és necessària per la unió de CP190 a l'*insulator* de Gypsy (Pai et al., 2004).

Una altra proteïna detectada entre els resultats de LC/MS és Mod(mdg4), que també participa en el funcionament de l'*insulator* de Gypsy a través de la seva unió a Su(Hw), amb qui sembla interaccionar principalment, i a més és essencial per la formació d'*insulator* bodies (Pai et al., 2004). De totes maneres no és necessària per la unió de CP190 a Gypsy. Sempre que participa en activitat *insulator* ho fa interaccionant amb Su(Hw), i la interacció amb CP190 ja ha estat demostrada per co-immunoprecipitació (Pai et al., 2004; Gerasimova et al., 2007).

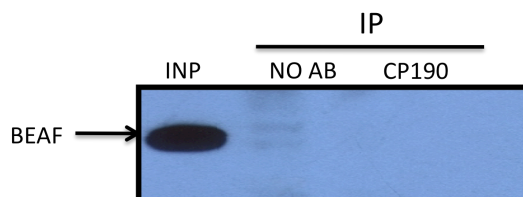
dCTCF és l'altra proteïna identificada formant complex amb CP190 i que està implicada en activitat *insulator*. En aquest cas, la seva interacció amb CP190 també ha estat demostrada per co-immunoprecipitació (Mohan et al., 2007).

Després ens preguntem si el fet que BEAF no aparegui a cap de les tres rèpliques de la purificació del complex de CP190 es deu a que realment no interaccionen o ha estat una limitació de la tècnica i la interacció se'ns ha passat per alt. BEAF és l'única proteïna d'*insulator* que no hem detectat, tot i que colocalitza extensament amb CP190 tal com s'ha vist per immunoprecipitació de cromatina (Bushey et al., 2009). BEAF s'uneix al DNA on reconeix una seqüència consens i per tant s'ha especulat que CP190 podria unir-se a ella de la mateixa manera que passa amb Su(Hw). No obstant, quan nosaltres hem realitzat una

immunoprecipitació amb extracte cel·lular amb l'anticòs de CP190 i després mirem per Western Blot si hi ha enriquiment respecte un control de falsa immunoprecipitació sense anticòs, el resultat és que no hi ha una banda de BEAF detectable al carril de la immunoprecipitació, malgrat l'anticòs de BEAF ser molt específic tal com es demostra a l'input (Figura 31-a). Per tant la interacció de BEAF amb CP190 continua sense ser demostrable, en coherència amb el que hem vist per LC/MS.

Per validar-ho comprovem que amb el mètode per immunoprecipitar que seguim alguna de les proteïnes d'insulator que trobem al complex sí que apareix en la co-IP. Vam transfectar de manera transitòria a les cèl·lules un plàsmid que expressava la proteïna Su(Hw) fusionada a un tag HA i en vam immunoprecipitar l'extracte amb anticòs específic de CP190, per després realitzar un Western Blot amb anticòs específic d'HA (Figura 31-b). Al carril de la immunoprecipitació amb anti-CP190 es detecta presència de Su(Hw)-HA, i en canvi no se'n detecta en el carril de la falsa immunoprecipitació sense anticòs, confirmant que per co-IP detectem aquelles proteïnes que hem trobat al complex.

A)



B)

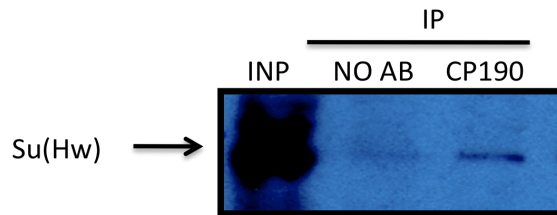


Figura 31.- Co-immunoprecipitacions de Su(Hw) i BEAF. A) Western Blot amb l'anticòs anti-BEAF. Primer carril input (10% del material immunoprecipitat), segon carril IP sense anticòs, tercer carril IP amb anti-CP190. Tot i que l'anticòs de BEAF és específic com es mostra al carril de l'input, no hi ha cap banda al carril de la IP de CP190. B) . Western Blot amb anticòs contra Su(Hw). Primer carril input (10% del material immunoprecipitat), segon carril IP sense anticòs, tercer carril IP amb anti-CP190. En aquest últim s'hi observa una banda qu corre igual que la de l'input.

4.2.2.2 NURF-38

Hem identificat NURF-38 com a membre del complex format per CP190. NURF-38 és un dels components del complex NURF, *Nucleosome Remodeling Factor* que forma part de la família ISWI. Es tracta d'un complex remodelador de la cromatina ATP-dependent que és dirigit directament a la cromatina a través d'interaccions amb factors de transcripció i histones modificades. De fet va ser identificat per primera vegada a *Drosophila* per la seva capacitat de potenciar l'accessibilitat del factor GAGA a cromatina reconstituïda. En particular, NURF-38 és una pirofosfatasa inorgànica que catalitza la incorporació de nucleòtids a cadenes d'àcid nucleic en processos de transcripció, replicació o reparació de l'ADN (Gdula et al., 1998).

Nosaltres hem volgut verificar la interacció amb CP190 per co-immunoprecipitació. Vam utilitzar extracte de cèl·lules S2 per immunoprecipitar amb l'anticòs de CP190 i analitzar la presència de NURF-38 per Western Blot. Com es veu a la figura 32 hi ha una clara banda al carril de la IP de CP190 que no hi és al carril del control negatiu sense anticòs, per tant confirmant la interacció. Malgrat que ja s'havia demostrat la necessitat del complex NURF en la funció

insulator (Li et al., 2010) aquesta és la primera vegada que es demostra una interacció de CP190 amb algun dels components de NURF.

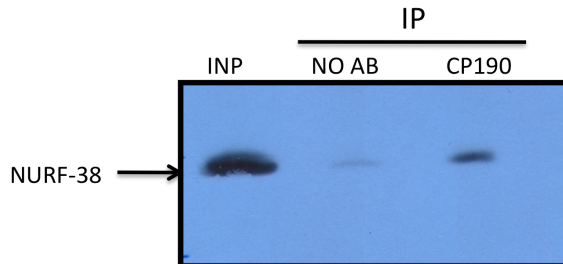


Figura 32.- Co-immunoprecipitació de NURF-38 amb CP190. A) Western Blot amb anticòs contra NURF-38. Primer carril input (10% del material immunoprecipitat), segon carril IP sense anticòs, tercer carril IP amb anti-CP190. S'observa una banda de la mateixa mida que la de l'input al carril de la IP de CP190.

4.2.2.3 Z4

Z4 és una proteïna amb dominis zinc finger que s'uneix a les interbandes dels cromosomes politènics, que està relacionada tant amb activació transcripcional com amb organització cromosomal (Eggert et al., 2004). S'ha vist que forma part d'un complex amb DREF essent essencial per l'activació d'alguns gens requerits per la replicació i també regulant positivament la via de senyalització Notch (Hochheimer et al., 2002). A més s'ha constatat que per la seva regulació de la via Notch fa complex amb tots els components de NURF (Kugler and Nagel, 2010).

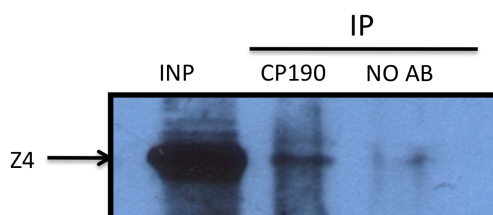


Figura 33.- Co-immunoprecipitació de Z4 amb CP190. A) Western Blot amb anticòs contra Z4. Primer carril input (10% del material immunoprecipitat), segon carril IP anti-CP190, tercer carril IP sense anticòs. S'observa una banda de la mateixa mida que la de l'input al carril de la IP de CP190.

Nosaltres verifiquem la interacció amb CP190 per co-immunoprecipitació (Figura 33). Immunoprecipitem l'extracte proteic de cèl·lules S2 amb anticòs de CP190 i fem un Western Blot amb anticòs de Z4. Tot i que feble, detectem una banda que correspon al tamany de Z4 (observat a l'input) al carril de la IP de CP190, que és més intensa que el que el que s'observa al carril del control negatiu sense anticòs.

4.2.2.3.1 Z4 colocalitza parcialment amb CP190 a politènics

Hem volgut comprovar fins a quin punt aquesta interacció entre Z4 i CP190 es dona a nivell d'unió a la cromatina. Hem preparat immunotincions de cromosomes politènics amb anticòs de CP190 i anticòs de Z4. El que observem és una colocalització parcial, moltes bandes no coincideixen però hi ha un subset de loci a on s'uneixen ambdues proteïnes. Això faria pensar que en cas d'actuar juntes ho farien només en determinats llocs del genoma, tal i com passa per exemple entre CP190 i Su(Hw) o CP190 i dCTCF.

Volem saber si Z4 és necessari per la unió de CP190 a la cromatina ja que Z4 s'uneix al DNA. Com que sabíem que la manca de la proteïna Chromator provoca que Z4 deixi d'estar unit a la cromatina (Gan et al., 2011), hem utilitzat larves que expressen RNAi de Chromator sota el control d'Actina-Gal4 per fer immunotinció de politènics (Figura 34). En primer lloc observem que aquests cromosomes deixen de tenir Z4 unit respecte els wild-type, però que a grans trets no perden unió de CP190, com si la unió de Z4 a la cromatina no fos necessària per CP190. Almenys això és el que s'observa a gran escala, a nivell de cromosomes sencers. Si bé és cert que, havent vist que aquestes dues proteïnes només colocalitzen a certs llocs, el que s'hauria de fer és analitzar detalladament el binding de CP190 només en aquelles bandes que colocalitza amb Z4.

A)



B)

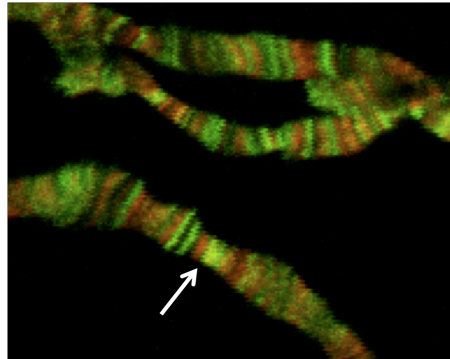


Figura 34.- Immunotincions de cromosomes politènics de larva wild-type i larva mutant de Chromator. A) La fletxa assenyala el cromosoma politènic que correspon al mutant de Chromator i que no té tinció amb l'anticòs de Z4 (vermell) però sí de CP190 (verd); B) En alguns loci hi ha colocalització de CP190 i Z4 (fletxa) en un cromosoma wild-type, però només és una colocalització parcial ja que en molts altres no colocalitzen.

4.2.2.3.2 Z4 afecta a l'activitat *insulator*

Hem volgut verificar si Z4, ara que sabem que interacciona amb CP190 i que s'uneix al DNA a llocs on també s'hi uneix CP190, és capaç d'afectar a la capacitat *enhancer-blocking* en línies transgèniques que porten construccions on l'*insulator* està situat entre l'*enhancer* del white i el gen reporter mini-white.

Hem observat que en dues construccions que porten l'insulator del Fab-7 i en una del Fab-6 hi ha un lleuger increment de la pigmentació en mutants de

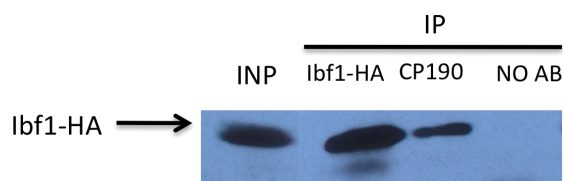
Z4. Encara que el nombre de línies en les que s'observa l'efecte és baix, aquests resultats indiquen que Z4 afecta a l'activitat enhancer-blocking al Fab-6 i Fab-7.

4.2.3 Caracterització i anàlisi funcional de Ibf-1 i Ibf-2

4.2.3.1 Interacció amb CP190

Entre les proteïnes identificades al complex n'hi ha dues que apareixen en totes tres rèpliques i que no han estat descrites fins al moment: CG8436 i CG9740, d'ara endavant anomenades Ibf-1 i Ibf-2 per *Insulator Binding Factor 1* i 2. En primer lloc verifiquem la interacció per co-immunoprecipitació. Ja que al moment de fer aquest experiment no disposàvem d'anticossos específics per aquestes dues proteïnes, vam generar plàsmids que les expressaven fusionades a un tag diferent per a cada una, Ibf-1 taguejada amb V5 i Ibf-2 amb HA. Vam transfectar-los de manera transitòria a cèl·lules S2 i vam realitzar immunoprecipitacions amb els anticossos contra els tags i contra CP190, del que sí teníem anticòs específic. A la figura 35 es mostren els Western Blot contra un tag i contra l'altre. D'entrada el que s'observa és la detecció d'una banda de cadascuna d'ambdues proteïnes al carril de la immunoprecipitació de CP190. Però a més també es constata un banda de Ibf-2 a la IP de Ibf-1. És a dir que no només es verifica la interacció d'ambdues proteïnes amb CP190 sinó també entre elles, suggerint un complex format per les tres proteïnes.

A)



B)

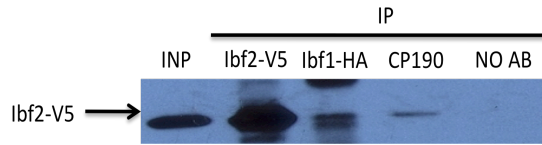


Figura 35.- Interaccions entre Ibf-1, Ibf-2 i CP190. A) Western Blot amb anticòs de V5 revelant presència d'Ibf-2 a les IP de Ibf2-V5, Ibf1-HA i CP190. B) Western Blot contra HA revelant presència d'Ibf1-HA a les IP d'Ibf1-HA i CP190.

4.2.3.2 Ibf-1 i Ibf-2 són gens adjacents i proteïnes similars

Ibf-1 i Ibf-2 són proteïnes codificades per dos gens situats al braç dret del cromosoma 3 (cromosoma 3R) entre les coordenades 5.083.200 i 5.086.200. Es troben doncs l'un al costat de l'altre, es transcriuen en sentits oposats i els seus TSS (*transcription start site*) estan separats per 470 bp (Figura 36).

El gen Ibf-1 (3R:5.084.691-5.086.122 [+]) fa 958 bp i genera un sol transcrit que conté quatre introns. Al traduir-se aquest transcrit codifica una proteïna de 242 aminoàcids, amb un pes predit de 27.7 KDa i un pI teòric de 6.38. Segons les dades d'expressió publicades a ModEncode té un màxim d'expressió a embrions d'entre 2 i 10 hores, mentre que l'expressió a adults és baixa. S'expressa a cèl·lules S2 i a la majoria de línies cel·lulars utilitzades de *Drosophila*. A adults, els majors nivells d'expressió es troben al Sistema Nerviós Central (CNS) i a ovaris.

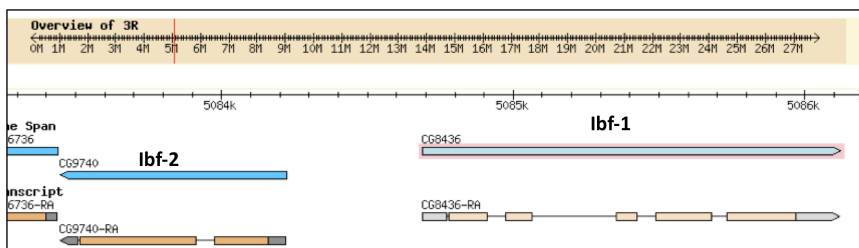
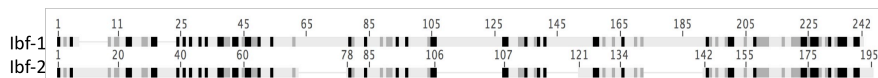


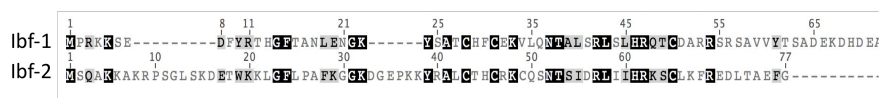
Figura 36.- Els gens CG8436 (Ibf-1) i CG9740 (Ibf-2) al genoma. Es troben l'un al costat de l'altre al cromosoma 3R però es transcriuen en direccions oposades.

El gen *lbf-2* (3R:5.083.449-5.084.223 [-]) de 708 bp genera un sol transcrit amb un sol intró. Aquest codifica una proteïna de 195 aminoàcids, amb un pes predit de 21.9 kDa i un pI teòric de 8.19. Té els majors nivells d'expressió a embrions d'entre 0 i 12 hores, s'expressa a cèl·lules S2 i a la majoria de línies cel·lulars utilitzades de *Drosophila*. A adults, els majors nivells d'expressió es troben al CNS, a ovaris i a testicles.

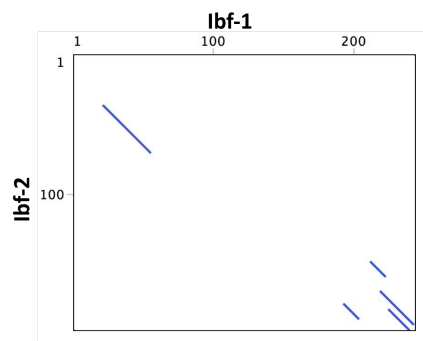
A)



B)



C)



D)

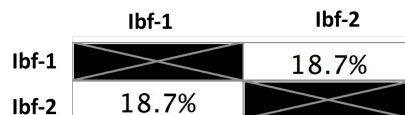


Figura 37.- Similitud entre *lbf-1* i *lbf-2*. A) Alineament Clustal de les dues proteïnes senceres, amb el grau de conservació indicat pel color, de més clar (no conservació) fins a més fosc (conservació). S'observa que tant a la regió N-terminal com a la C-terminal hi ha més posicions en negre (similars) que al mig de la proteïna. B) Alineament Clustal de la regió N-terminal de les dues proteïnes on veuen els aminoàcids compartits per ambdues. C) Plot d'una proteïna contra l'altra, en el que es ressalten només les zones més similars, que són al principi de cada proteïna i al final. D) Les proteïnes tenen en global un 18.7% de similitat segons l'alineament Clustal.

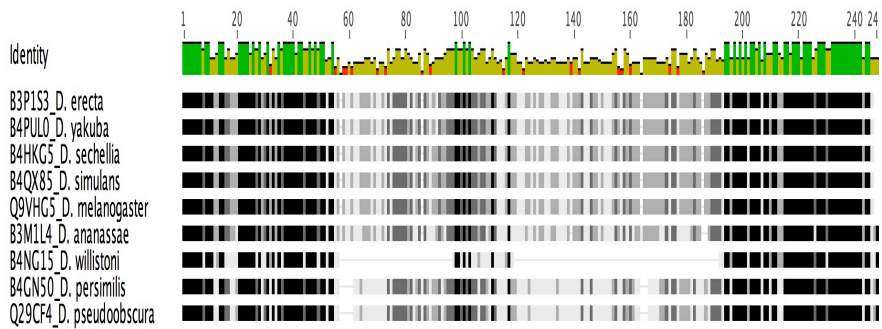
Al realitzar una cerca de proteïnes similars a lbf-1 dins de *Drosophila melanogaster* amb BLAST (Altschul, 1997) i amb les condicions per defecte el primer i únic hit és precisament lbf-2. El mateix passa al realitzar la cerca amb lbf-2, apareix lbf-1 com a proteïna més similar. Això suggereix que es tracta de dues proteïnes que, al menys en alguna part de la seva seqüència, s'assemblen. Per això decidim fer un alineament ClustalW (Larkin et al., 2007): les dues proteïnes tenen un 18.7% de similaritat (Figura 37). De l'alineament però sobretot del Dot Plot n'observem que hi ha una regió particularment ben conservada a la meitat N-terminal de les proteïnes, entre els aminoàcids 20 i 80 aproximadament. De la mateixa manera hi ha una altra regió ben conservada als extrems C-terminal, als últims 30 aminoàcids de cada proteïna.

4.2.3.3 lbf-1 i lbf-2 estan conservades dins del gènere *Drosophila*

La següent pregunta és si aquestes dues proteïnes estan conservades en altres espècies. Per a això utilitzem altra vegada l'eina BLAST i la base de dades OrthoDB (Waterhouse et al., 2011). En detectem homòlegs a totes les espècies del subgènere *Sophophora*, és a dir, totes les espècies del gènere *Drosophila* tret de *Drosophila virilis*, *D. mojavensis* i *D. grimshawi*, les tres més allunyades evolutivament de la resta.

Hem alineat tots els homòlegs d'una i altra proteïna amb ClustalW (Figura 38) i es torna a observar una forta conservació de les regions N-terminal (aminoàcids 1-60 a lbf-1 i 20-80 a lbf-2) i C-terminal (aminoàcids 200-248 a lbf-1 i 180-223 a lbf-2) a totes les espècies del subgènere. La matriu de distàncies que se'n deriva reflecteix amb algun matís l'arbre filogenètic d'aquestes espècies. A notar que l'homòleg de lbf-2 a *Drosophila ananassae* sembla haver patit una duplicació (GF18423 i GF24686) en la que un dels dos gens derivats ha perdut la meitat N-terminal.

A)



B)

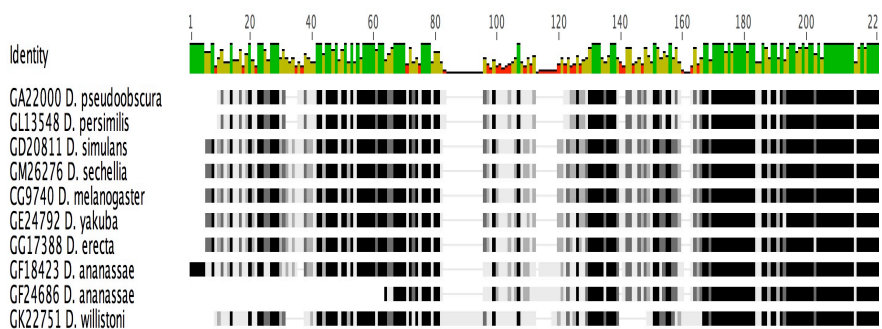


Figura 38.- Alineament de Ibf-1 i Ibf-2 amb els seus homòlegs a Drosophila. Hem alineat Ibf-1 (A) i Ibf-2 (B) cadascun amb els seus corresponents homòlegs a les espècies de Drosophila que en tenen. A sobre de cada alineament es mostra un gràfic d'identitat essent verd màxima identitat i vermell mínima.

4.2.3.4 Ibf-1 i Ibf-2 tenen un domini d'unió al DNA

L'observació que hi hagi dues regions altament conservades no només entre cada proteïna i els seus homòlegs sinó també d'una proteïna amb l'altra ens fa preguntar-nos si alguna d'aquestes regions o totes dues contenen algun motiu proteic conegut. Si bé per la regió C-terminal la cerca no dóna cap resultat clar, per la regió N-terminal trobem que la seqüència d'aminoàcids segueix un patró similar al descrit pel domini zinc finger BED (cerca Pfam E-value 0.12).

El zinc finger BED (Aravind, 2000) és domini proteic del que s'ha suggerit que la seva principal funció és la unió a DNA. Va ser anomenat així perquè va ser descobert inicialment en les proteïnes de *Drosophila* BEAF i DREF (regulador transcripcional de gens involucrats en la fase S del cicle cel·lular). Es troba en factors reguladors i en transposases de plantes, animals i fongs. Té de 50 a 60 aminoàcids que contenen un motiu característic $C_{x_2}C_{x_n}H_{x_{3-5}}[H/C]$, essent x_n un espaiador variable que forma un zinc finger, i a N-terminal d'aquesta signatura s'hi troben conservades dues posicions aromàtiques.

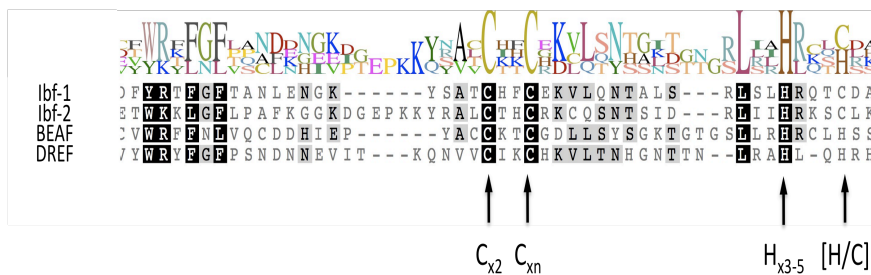
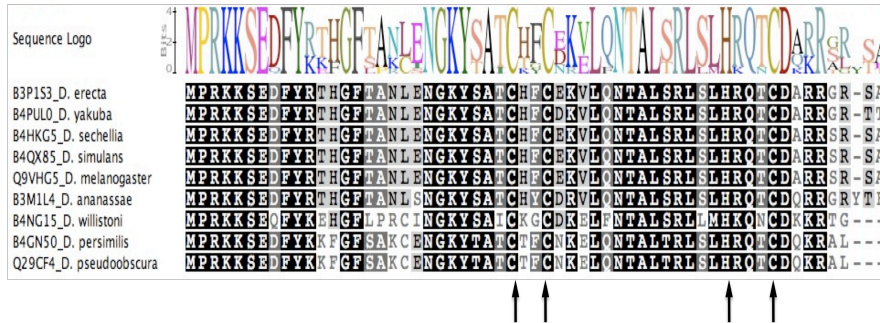


Figura 39.- BED finger. Alineament de les regions N-terminal de Ibf-1 i Ibf-2 alineades amb les regions N-terminal de DREF i BEAF on tenen el motiu BED. En negreta són les posicions conservades típiques d'aquest domini. Les fletxes indiquen les quatre posicions que formen el zinc finger: $C_{x_2}C_{x_n}H_{x_{3-5}}[H/C]$.

Analitzant-les al detall, tant Ibf-1 com Ibf-2 mostren aquest patró característic. A la figura 39 es mostra l'alineament de les regions N-terminal d'ambdues proteïnes amb BEAF i DREF i s'assenyalen els aminoàcids que conformen el motiu. Això explicaria el per què Ibf-1 i Ibf-2 tenen conservada la regió N-terminal: les dues tenen el mateix motiu d'unio al DNA. Ara bé, si això és correcte es podria predir que els homòlegs trobats a les demés espècies de *Drosophila*, que també tenen conservat l'N-terminal, haurien de contenir també la signatura del zinc finger. En efecte aquest és el cas tal i com es mostra a la figura 40, on es mostra l'alineament de tots els homòlegs i amb una fletxa s'assenyalen els aminoàcids característics del motiu.

A)



B)

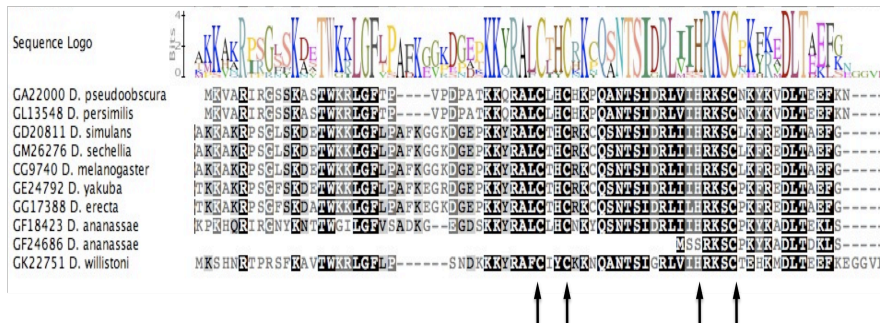


Figura 40.- Conservació del BED finger. Alineament de les regions N-terminal de Ibf-1 (A) i Ibf-2 (B). A sobre es mostra el logo generat per l’alineament, que és una matriu de conservació. Les fletxes de sota indiquen les quatre posicions més conservades del zinc finger.

4.2.3.5 Generació d’un mutant de Ibf-1

La línia PBac{RB}CG8436^{e03576} conté un transposó transgènic PBac{RB} situat al lloc 3R:5.084.968, just al límit entre el primer intró i el segon exó de Ibf-1. Aquesta línia expressa de manera normal la proteïna Ibf-1 (no mostrat). Per obtenir el mutant fem saltar el transposó. Aquest transposó conté també un gen mini-white que serveix de reporter: si el transposó ha saltat els ulls passen a ser blancs. Per a fer-lo saltar es va crear aquesta línia amb una línia que expressa una transposasa que excideix el transposó. De les 40 descendents que havien

passat a tenir ulls blancs, una no havia produït un salt net: si el salt és net no s'excideix cap part del gen i no es produeix un mutant.

A la línia que havia produït un salt no-net la vam anomenar Δ lbf-1. Un anàlisi genètic per PCRs d'aquest salt va revelar que l'extrem 3' del transposó havia romangut intacte (és l'extrem que queda més pròxim a l'extrem 5' del gen) i que el salt s'havia endut tota la meitat 5' del transposó, tota la resta del gen lbf-1 fins al seu final 3' i la major part del gen adjacent, *VhaM8.9*, del que només queda l'extrem 3'. Aquest gen està implicat en la via de senyalització Wnt i la via Planar Cell Polarity, i la manca total de la seva proteïna causa letalitat embrionària. Això suposa el problema que no podem comprovar quins fenotips s'obtenen per manca d'lbf-1.

Quan fem extracció proteica de larves heterozigotes i en fem un Western Blot contra l'anticòs específic de lbf-1 observem que el nivell de proteïna baixa significativament tot i tenir encara una còpia wild-type del gen (Figura 41). Per tot això considerem que Δ lbf-1 és un mutant nul de lbf-1.

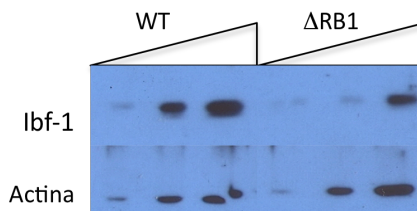


Figura 41.- Nivell d'lbf-1 al mutant Δ lbf-1 en heterozigosi. Western Blot contra lbf-1 amb extracte proteic de cervell de larva III de larva wild-type (esquerra) i de larva Δ RB1/TM6 (dreta). A sota es mostra la banda d'actina com a control de càrrega.

4.2.3.6 Generació d'un mutant de lbf-2

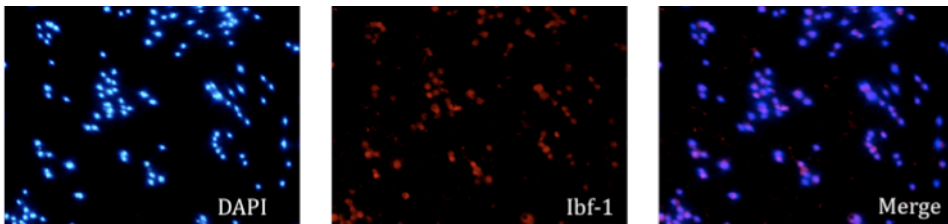
De la mateixa manera vam voler generar una línia mutant de lbf-2. En aquest cas disposàvem d'una línia, P{GSV6}GS16482, que té un transposó P{GSV6} insertat al lloc 3R:5.084.094 que cau al primer exó de lbf-2. Aquesta línia és de per si mutant de lbf-2: els homozigots no l'expressen en absolut (no

mostrat). Així, ens calia fer saltar la part del transposó que tenia el mini-white perquè aquesta línia no es pot utilitzar en assaigs d'enhancer-blocking degut a la presència al transposó d'un mini-white. Vam doncs creuar aquesta línia amb una línia que expressava la transposasa i de cinquanta descendents amb ulls blancs, deu no havien fet un salt net. Vam comprovar per PCR i seqüenciació que al menys dues d'aquestes línies, $\Delta\text{lbf-2}(17)$ i $\Delta\text{lbf-2}(20)$, se'ls havia produït un salt imperfecte del transposó que feia que hagués saltat el white que hi ha dins del transposó però no la resta, de manera que continuaven sent mutants d'lbf-2. Els GSV17 presenten un 90% de supervivència a larva III i un 70% a adults. Els mascles són estèrils.

4.2.3.7 Localització a cèl·lules S2

Hem generat anticossos que reconeixen específicament Ibf-1 i Ibf-2. Volem saber si tenien localització nuclear, tal i com esperàvem tenint en compte que CP190 localitza al nucli a interfase. Per tant primer de tot vam fer una immunotinció de cèl·lules S2 on es veu que la seva tinció corresponia amb la del DAPI que marca el nucli de les cèl·lules interfàsiques (Figura 42). Malgrat tot, l'anticòs de Ibf-1 sembla que dóna un cert soroll de fons. També vam fer la immunotinció de cèl·lules S2 però aquesta vegada havent-les impactat, i altra vegada es va confirmar que Ibf-1 i Ibf-2 localitzen al nucli a cèl·lules S2 interfàsiques de la mateixa manera que CP190 (Figura 43).

A)



B)

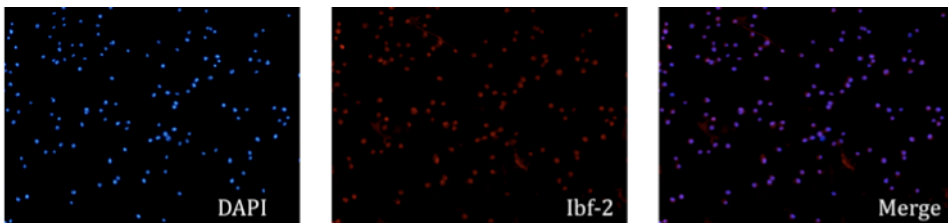
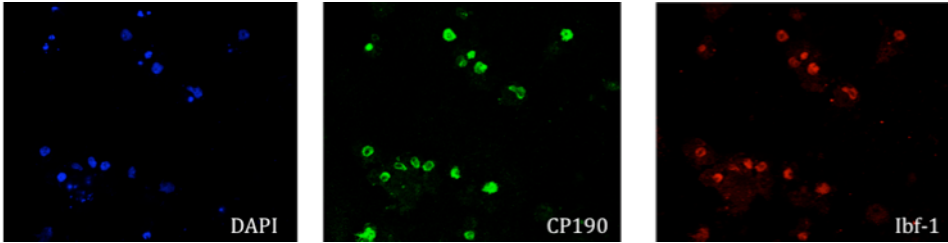


Figura 42.- Immunolocalització de Ibf-1 i Ibf-2 cèl·lules S2. Tant Ibf-1 (A) com Ibf-2 (B) localitzen al nucli en una immunotinció en cèl·lules interfàsiques.

A)



B)

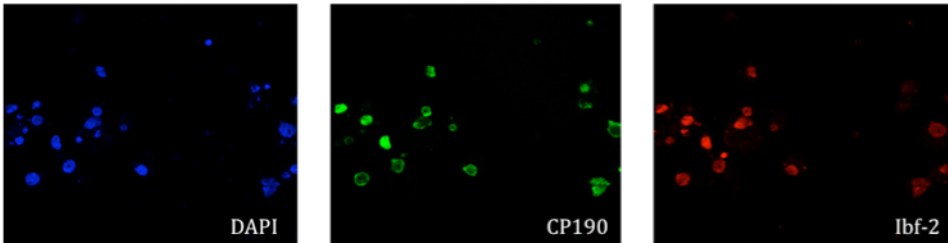
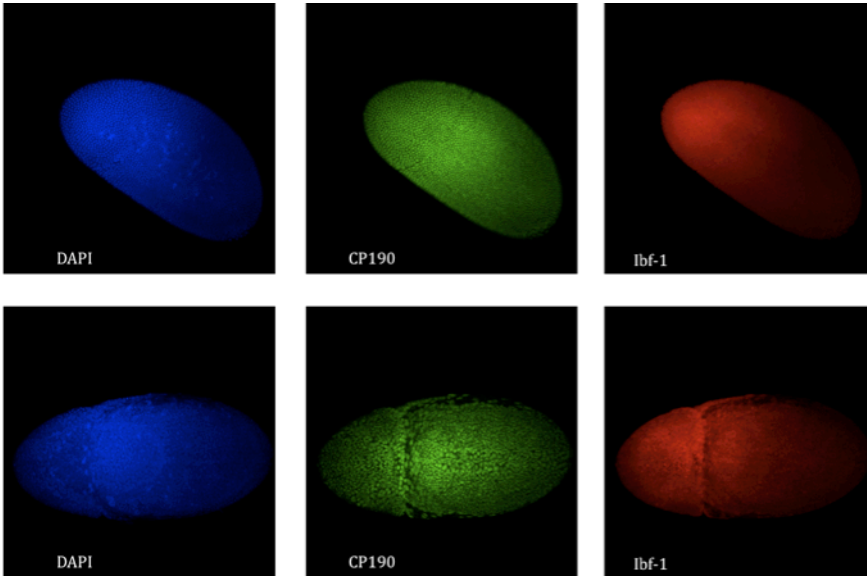


Figura 43.- Immunolocalització de Ibf-1, Ibf-2 i CP190 a cèl·lules impactades. A cèl·lules impactades es verifica que Ibf-1 (A) i Ibf-2 (B) són nuclears i colocalitzen al nucli amb CP190.

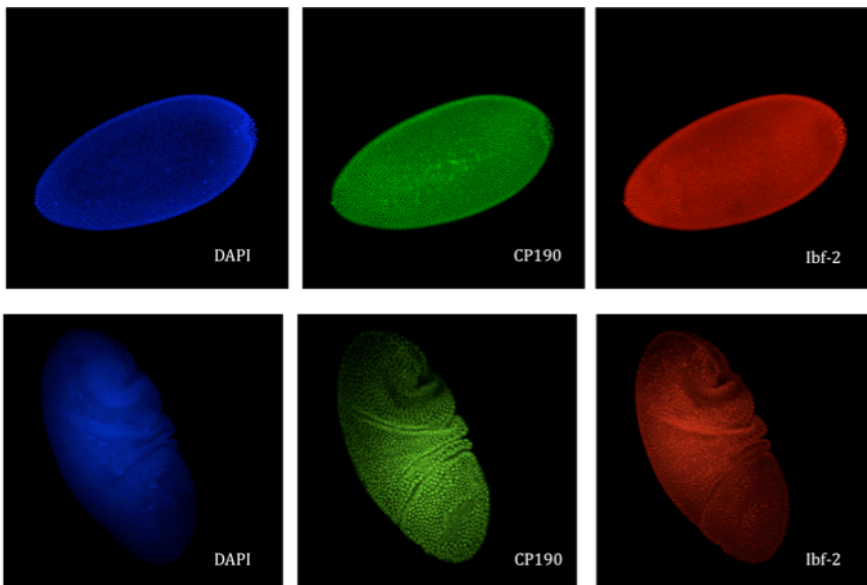
4.2.3.8 Localització a embrions

Hem volgut mirar la localització de Ibf-1 i Ibf-2 a embrions de diferents estadis del desenvolupament per tal de saber on localitzen aquestes dues proteïnes i si ho fan de manera diferent segons el moment del desenvolupament embrionari (Figura 44). Ibf-2 queda clarament delimitada al nucli cel·lular en tots els estadis, coincidint amb la senyal de CP190, també nuclear. Per contra, la senyal d'Ibf-1 no queda delimitada al nucli sinó que es veu per tot l'embrió. Hem mirat la localització d'Ibf-2 als dominis mitòtics (zones de l'embrió en que totes les cèl·lules estan dividint-se) i és difícil dir on localitza dins del nucli, ja que la senyal és molt difosa i no és clar ni que estigui unida als cromosomes mitòtics ni que —com CP190— localitzi als centrosomes.

A)



B)



C)

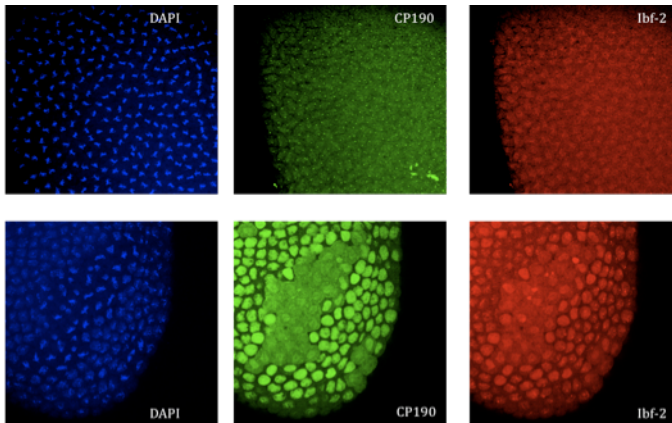


Figura 44.- Immunolocalització de Ibf-1, Ibf-2 i CP190 als embrions. A) CP190 localitza clarament a nuclis tant a embrions primerencs com tardans, mentre que la senyal d'Ibf-1 no sembla restringida al nucli; B) La senyal d'Ibf-2 sí que queda delimitada als nuclis de les cèl·lules embrionàries tant a embrions primerencs com a embrions tardans, coincidint amb la senyal de CP190; C) A dominis mitòtics, Ibf-2 continua localitzada al nucli.

4.2.3.9 Unió a cromosomes politènics

Ara que ja sabem que es tracta de dues proteïnes nuclears, vam voler mirar si s'unien a la cromatina. Per a fer-ho vam fer immunotincions de cromosomes politènics de glàndules salivals de larva III, tenyides amb els anticossos específics de Ibf-1 i Ibf-2, de CP190 i DAPI per marcar el DNA. Com s'observa a la figura 45, tant Ibf-1 com Ibf-2 s'uneixen de forma clara als politènics. Formen prop d'un centenar de bandes que, comparant-los amb la tinció del DAPI, mostren preferència per les interbandes. Les interbandes són les zones dels cromosomes politènics que queden poc o gens tenyides pel DAPI (en oposició a les bandes, fortament marcades) i que s'associen a regions riques en gens i transcripcionalment actives. La tinció amb l'anticòs de CP190 mostra un nombre molt major de bandes, també preferencialment coincidint amb les interbandes del DAPI. Quan comparem les tincions de Ibf-1 o de Ibf-2 amb la de CP190 s'observa que hi ha una gran part de llocs on hi coincideixen. Naturalment, a l'unir-se

CP190 a molts més llocs hi ha molts llocs de CP190 sense co-localització. Però mirat a la inversa la majoria de llocs de Ibf-1 o de Ibf-2 cauen en llocs també ocupats per CP190.

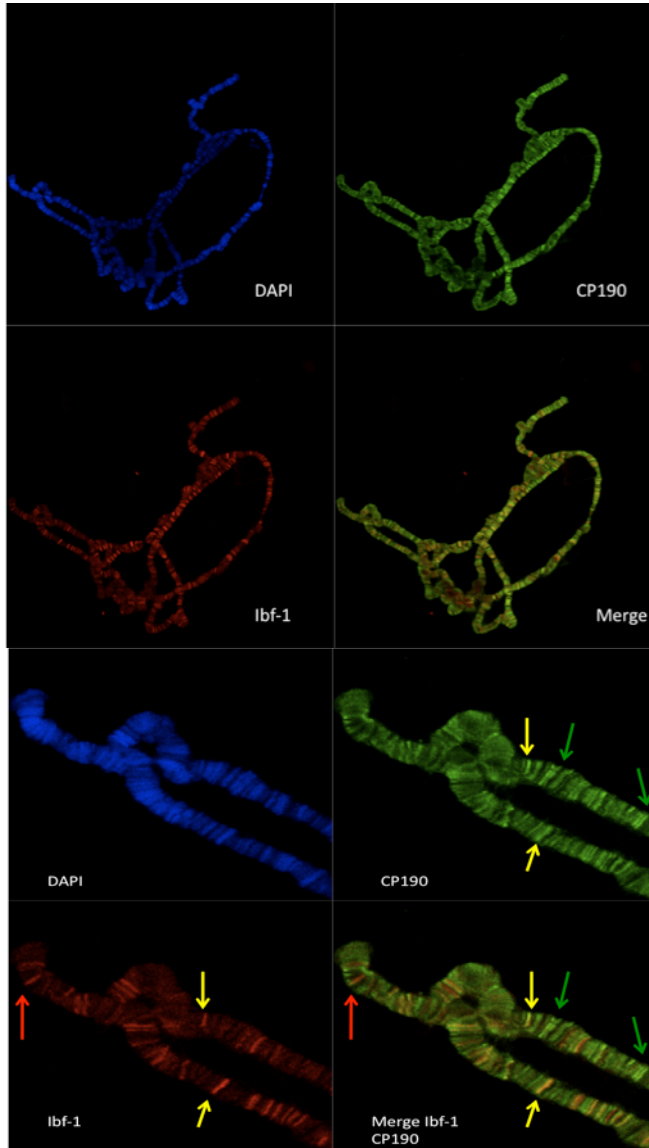


Figura 45.- Immunotinció d'Ibf-1 i CP190 de cromosomes politènics. A) CP190 i Ibf-1 coincideixen a molts llocs dels genoma. Les fletxes verdes indiquen bandes on només hi ha CP190; les fletxes vermelles bandes on només hi ha Ibf-1; en fletxes grogues bandes on hi són totes dues.

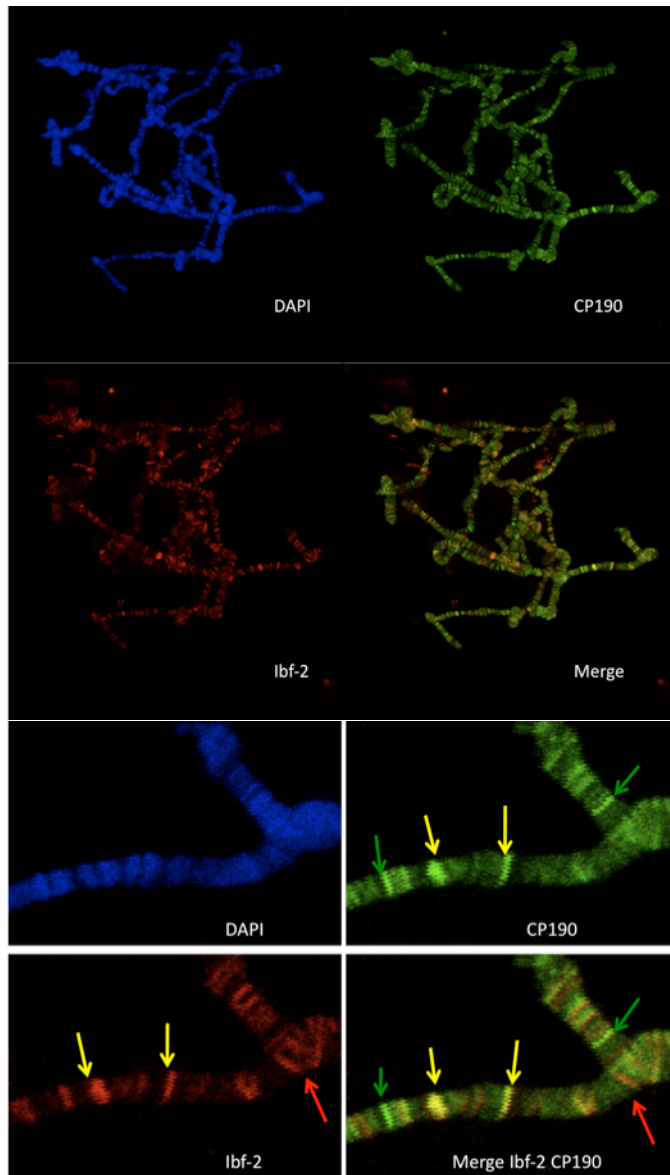
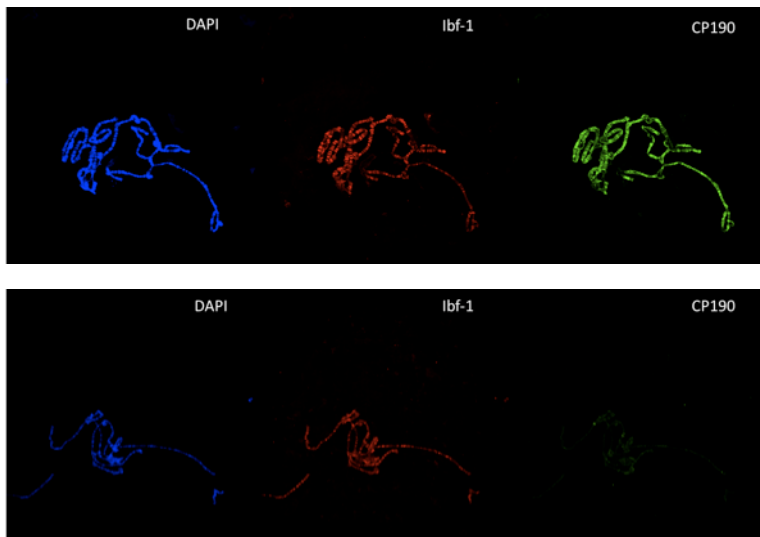


Figura 46.- Immunotinció d'lbf-2 i CP190 a cromosomes politènics. A) CP190 i lbf-2 coincideixen a molts llocs dels genoma; B) les fletxes verdes indiquen bandes on només hi ha CP190; les fletxes vermelles bandes on només hi ha lbf-2; en fletxes grogues bandes on hi són totes dues.

La següent pregunta és si CP190 és necessària per la unió de Ibf-1 o de Ibf-2. Fem la immunotinció aquesta vegada utilitzant larves mutants de CP190. Són larves que combinen dues mutacions diferents de CP190, P1 i P11, ja que cadascuna d'aquestes no és viable en homozigosi a larva III. Al combinar-les hi ha letalitat a partir de larva III, de manera que podem escollir aquelles larves que tinguin les dues mutacions a la vegada a través del seguiment del cromosoma balancejat TM6 que porta el marcador tubby. Així, les larves que no tinguin fenotip tubby (més curtes del normal) seran mutants nul·les de CP190. Com es veu a la figura 46, no s'observen canvis significatius en la unió de Ibf-1 o de Ibf-2 als politènics quan falta CP190 respecte d'un cromosoma de larva wild-type.

A)



B)

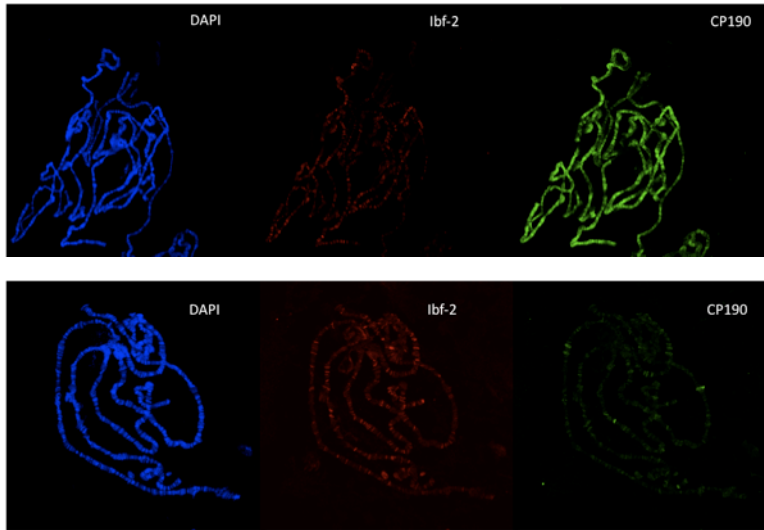


Figura 47.- Immunotinció d'Ibf-1, Ibf-2 i CP190 a cromosomes politènics de larves mutants de CP190. Cromosomes de larva *wild-type* i cromosomes de larva mutant de CP190 (CP190^{P1}/CP190^{P11}) on s'observa que CP190 deixa d'unir-se a la cromatina mentre que Ibf-1 (A) i Ibf-2 (B) hi continuen units.

Aquest resultat mostra que Ibf-1 i Ibf-2 s'uneixen a la cromatina independentment de CP190. Per saber què passa en sentit invers, és a dir, si Ibf-1 o Ibf-2 són necessàries per la unió de CP190 a la cromatina utilitzem larves mutants nul·les de Ibf-2 i larves que combinen la mutació de Ibf-1 amb un RNAi de Ibf-1 sota el control de LioGal4 (driver a glàndules salivals), ja que les homozigotes mutants de Ibf-1 no arriben a larva III.

D'entrada el primer que es constata és que la unió de Ibf-2 és necessària per la unió de Ibf-1 a la cromatina. Als cromosomes de larves mutants de Ibf-2 no hi ha cap banda de Ibf-1 (Figura 48). En canvi quan mirem la tinció de CP190 a aquests cromosomes no observem canvis significatius a escala cromosòmica. A les larves mutants de Ibf-1 també observem una dràstica reducció del nivell de Ibf-2 unit a la cromatina (Figura 49).

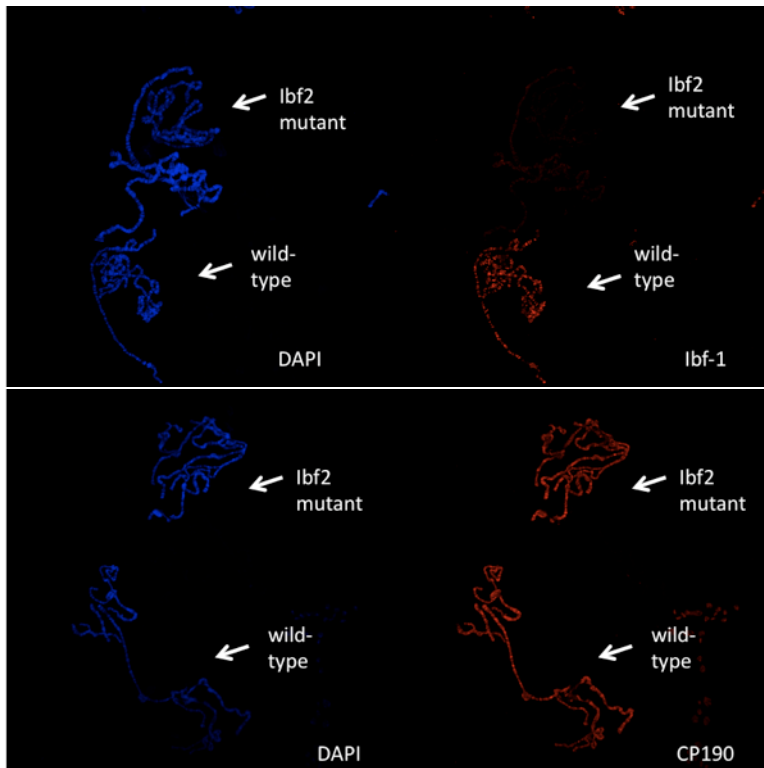


Figura 48.- Immunotinció d'Ibf-1 i CP190 a cromosomes politènics de larva mutant d'Ibf-2. Ibf-1 deixa d'unir-se a la cromatina en cromosomes de larva mutant d'Ibf-2 mentre que la unió de CP190 és independent de la unió d'Ibf-2.

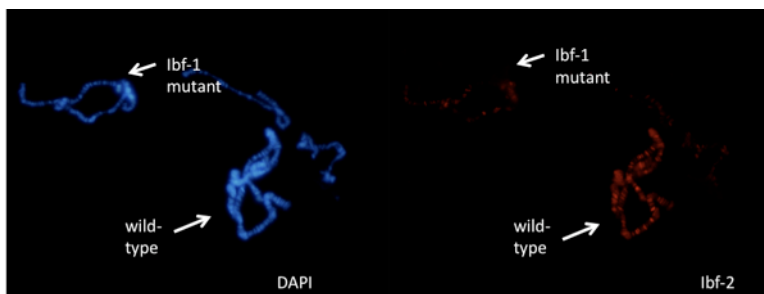
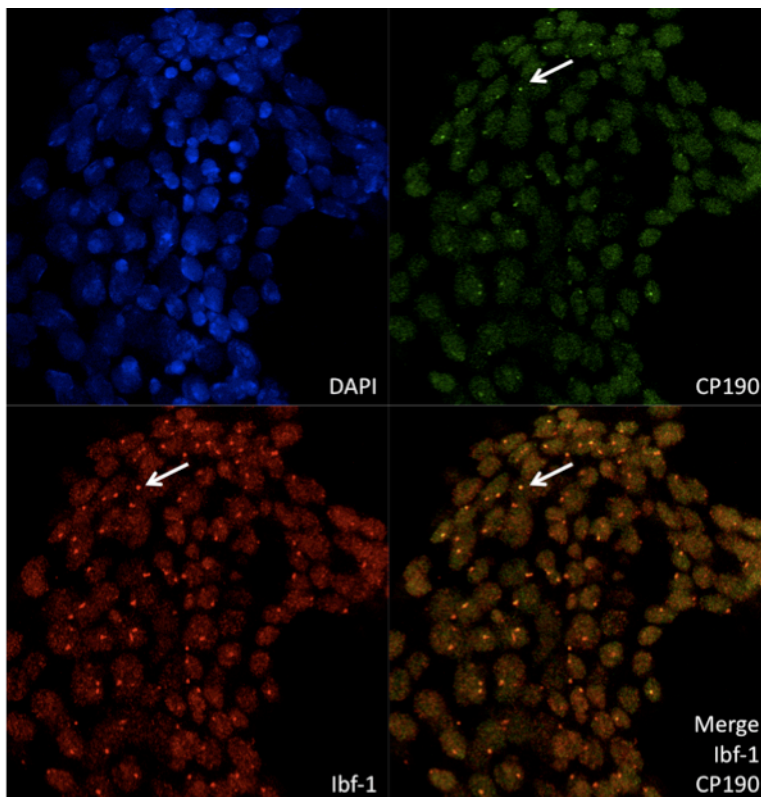


Figura 49.- Immunotinció d'Ibf-2 a cromosomes politènics de larva mutant d'Ibf-1. Ibf-2 deixa d'unir-se a la cromatina en cromosomes de larva mutant d'Ibf-1. Per tal d'obtenir larva III mutant d'Ibf-1 hem combinat en un al·lel la deleció Δ lbf-1 amb un RNAi contra Ibf-1 sota el control de LioGal4 que s'expressa a glàndules salivars.

4.2.3.10 Presència d'*insulator bodies*

Ha estat demostrat en diversos estudis que l'activitat *insulator* es correlaciona estretament amb la formació d'*insulator bodies* (Capelson and Corces, 2004, 2005), que són llocs dins del nucli on hi ha varis *insulators* junts causant que la fibra de cromatina formi loops que se suposa que coincideixen amb dominis independents d'expressió gènica (Gerasimova and Corces, 2001). L'acumulació en aquests corpuscles de proteïnes d'*insulator* és visible per immunotinció, i se sap que les diferents proteïnes d'*insulator* hi coincideixen (Gerasimova et al., 2007).

A)



B)

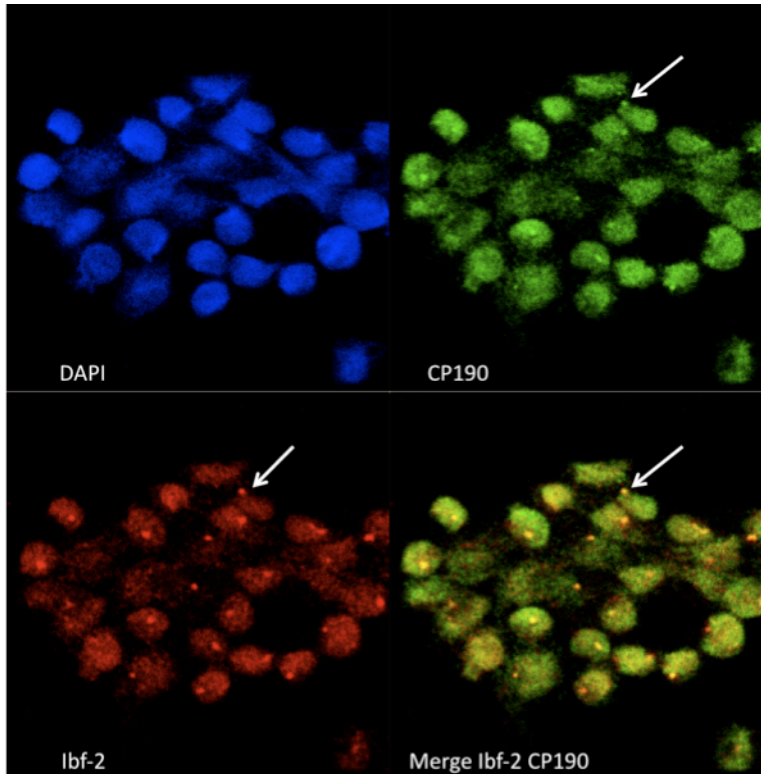


Figura 50.- Insulator bodies a nuclis de cèl·lules de cervell. Les immunotincions a cèl·lules de cervell presentaven insulator bodies quan els teníem amb Ibf-1 (A) i Ibf-2 (B), que colocalitzaven amb els insulator bodies que s'observen quan es teneix amb l'anticòs de CP190, tot i que en aquest cas a l'haver-hi més senyal de la proteïna al nucli són més difícils de distingir.

Per comprovar si Ibf-1 i Ibf-2 són presents als *insulator* bodies amb una distribució similar a la de CP190, vam realitzar immunotincions de cèl·lules de cervell de larva III. A la figura 50 s'observa com en totes dues proteïnes hi ha una tinció particular d'uns corpuscles dins del nucli, que a més també hi són en la tinció amb CP190, suggerint que en efecte es tracta d'*insulator* bodies. No obstant a les tincions de CP190 eren més difícils de distingir perquè la forta presència d'aquesta proteïna al nucli dóna senyal molt forta.

4.2.3.11 Localització en cromosomes mitòtics

CP190 és una proteïna que, si bé a interfase es troba dins del nucli unida a la cromatina, quan la cèl·lula entra en mitosi localitza als centrosomes (Oegema et al., 1995; Whitfield et al., 1995).

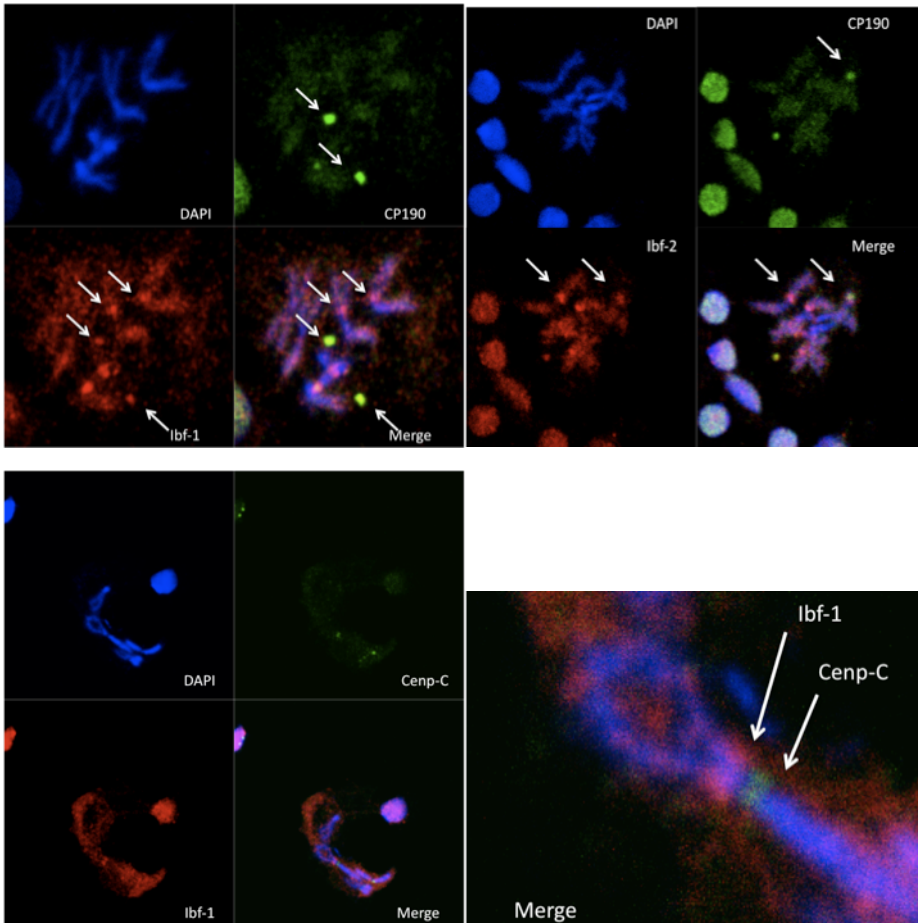


Figura 51.- Localització d'Ibf-1 i Ibf-2 a cromosomes mitòtics. CP190 localitza a dos estructures puntuades que corresponen als centrosomes. Ibf-1 està unit tot al llarg dels cromosomes mitòtics, amb particular intensitat al centròmer, i també al centrosoma; Ibf-2 també es troba unida al llarg del cromosoma mitòtic, amb especial intensitat al centròmer i amb una marca als centrosomes on colocalitza amb CP190; la senyal d'Ibf-1 al centròmer està just al costat que la de CENP-C, però no al mateix lloc.

Nosaltres ens vam preguntar si aquest també era el cas de Ibf-1 i Ibf-2, i si no, a on localitzen durant la mitosi. Utilitzant cèl·lules de cervell de larva III vam buscar cromosomes mitòtics, que són molt reconeixibles amb la tinció amb DAPI perquè estan molt condensats i tenen una forma característica (Figura 51).

D'entrada la primera observació és que Ibf-1 i Ibf-2 a mitosi mostren un patró difós i no sempre clar, observació que ja havíem remarcat a embrions. No sembla clar per les nostres imatges que vagin als centrosomes, tot i que en alguns casos sí que hi observem senyal. Més aviat mostren un patró d'unió als cromosomes mitòtics, amb una possible marca particular als centròmers, la regió que separa els dos braços dels cromosomes. Per això vam fer una co-tinció amb Cenp-C, una proteïna que localitza als centròmers. El que observem no és una co-localització perfecta sinó que Ibf-1 i Ibf-2 se situen just al costat de Cenp-C (Figura 51-d).

4.2.3.12 Alteracions en l'expressió gènica

Les proteïnes d'*insulator* actuen sobre un gran nombre de llocs al genoma controlant de maneres diverses les regions reguladores de molts gens, essent essencial el correcte funcionament dels *insulators* per mantenir l'estat transcripcional d'un tipus cel·lular. Alteracions o defectes en les proteïnes que duen a terme aquesta tasca porten directa o indirectament a canvis en la regulació transcripcional. Per tant és d'esperar que una conseqüència de mutar una proteïna d'*insulator* serà que una sèrie de gens augmentarà la seva transcripció i uns altres la disminuiran, ja que els *insulators* no semblen estar lligats a activació o repressió sinó més aviat a separació d'unitats transcripcionals i de regions reguladores, estiguin aquestes en estat actiu o silenciades. Exemples d'aquesta desregulació generalitzada han estat trobats a CP190 i a dCTCF a *Drosophila* (Bartkuhn et al., 2009).

Nosaltres hem volgut saber si la falta de Ibf-1 i Ibf-2 també té conseqüències a nivell transcripcional. Volem veure si actuen conjuntament amb CP190 pel que fa a la regulació de l'expressió gènica. Utilitzant cultius de cèl·lules S2 eliminarem per separat

les tres proteïnes del complex (Ibf-1, Ibf-2, CP190) amb RNAi i n'analitzarem els possibles canvis en transcripció amb un microarray.

4.2.3.12.1 RNAi a S2

El mètode que utilitzarem és el *bathing*, que consisteix en produir RNA de doble cadena que correspongui a un fragment d'uns 150-500 bp del transcrit que volem eliminar, i afegir-lo al medi de cultiu a on creixen les cèl·lules. Aquestes integraran les molècules de dsRNA que es trobaran repartides pel medi i un cop dins la cèl·lula s'activa la màquinaria d'RNAi, que tallarà el dsRNA en petites seqüències que dirigiran la maquinària d'eliminació directament fins al transcrit d'interès.

Per tant en primer lloc el que fem és escollir una regió exònica que sigui comú a tots els transcrits del gen en cas que n'hi hagi més d'un i dissenyar primers que l'amplifiquin per PCR. Aquests primers porten a cada extrem un promotor T7 que permetran la transcripció *in vitro* de l'amplicó. El dsRNA generat l'afegim al medi de cultiu en tres dies diferents. Al cap de quatre dies, recuperem les cèl·lules i fem una extracció d'RNA, que purifiquem per ser analitzat per microarray.

Repetim aquest procés tres vegades i ho fem per quatre mostres diferents cada vegada: Ibf-1, Ibf-2, CP190 i un RNAi control. Aquest RNAi control és una molècula de dsRNA que no té per què afectar a cap gen en concret i que serveix per tenir una mostra de l'RNA wild-type que tingui en compte els eventuais canvis transcripcionals que pugui tenir l'activació de la màquinaria RNAi. En el nostre cas escollim una regió del gen bacterià LacZ. A la figura 52 es mostra un Western Blot d'extracte proteic de cèl·lules que han rebut el tractament d'RNAi revelat contra la proteïna que hem intentat eliminar. Com s'observa ni Ibf-1 ni Ibf-2 han aconseguit ser eliminades del tot, tot i que sí que hi ha una baixada significativa dels nivells de proteïna. En el cas de CP190 la baixada és menor.

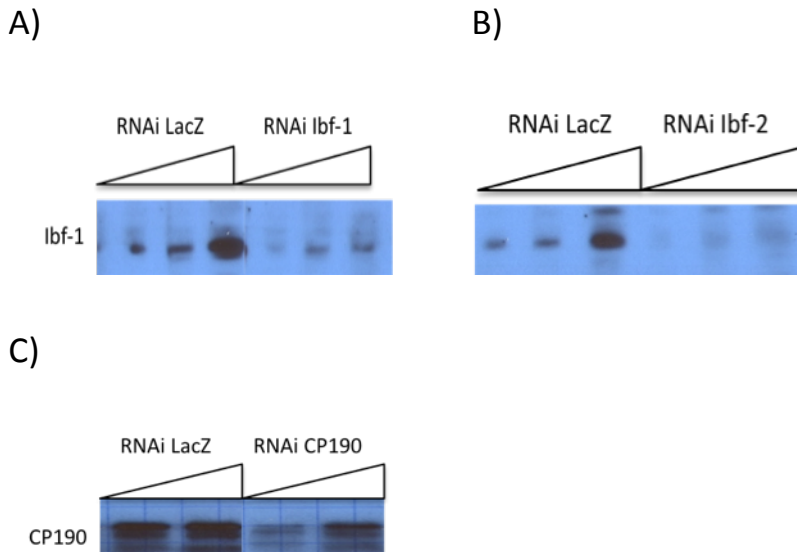


Figura 52.- Baixada del nivell de proteïna en resposta a RNAi. Western Blots de Ibf-1 (A), Ibf-2 (B) i CP190 (C) mostrant els nivells després de tractament RNAi amb LacZ i RNAi de la pròpia proteïna.

4.2.3.12.2 Canvis d'expressió gènica

L'RNA extret de cada mostra es retrotranscriu i s'amplifica, i el cDNA resultant s'hibrida amb el Microarray amb el qual obtenim un valor de variació per cada gen. A l'analitzar cada rèplica per PCA (Anàlisi de Components Principals) s'observa que la tercera rèplica està molt allunyada de les dues primeres, per totes les mostres. Tant allunyada que la variabilitat entre rèpliques és més gran que la que hi ha entre mostres, i això podria esbiaixar l'anàlisi estadístic. D'aquesta manera decidim tirar endavant l'anàlisi només amb les rèpliques 1 i 2, que és just el mínim per poder treure conclusions sobre gens diferencialment expressats. El principal problema d'utilitzar només dues rèpliques és que la robustesa de l'anàlisi serà menor, és a dir, que possiblement estiguem fent una sub-estimació i molts gens que realment estiguin diferencialment expressats no siguin detectats com a tals perquè no han passat el test de probabilitat.

A la taula 5 s'indica el fold-change de cadascuna de les tres proteïnes (en aquest cas, els seus trànscrips) en cadascun dels tres tractaments. Podem veure doncs com tant *Ibf-2* com CP190 han baixat més de 4 vegades la seva quantitat de trànscrit respecte el control en els seus propis RNAi, mentre que l'RNAi de *Ibf-1* no ha estat tan efectiu i ha baixat 2.8 vegades. També veiem que la manca de CP190 no ha causat variacions en l'expressió de *Ibf-1* ni *Ibf-2*, ni cap d'elles dues fa variar significativament CP190. Ara bé, sí que es constata que la manca de *Ibf-2* fa augmentar la transcripció de *Ibf-1* (fc 1.875), i viceversa (fc 1.776).

Taula 5.- Canvis d'expressió d'*Ibf-1*, *Ibf-2* i CP190 en cadascun dels tres knock-downs.

<i>gen afectat</i>	<i>fold change</i>		
	<i>RNAi Ibf-1</i>	<i>RNAi Ibf-2</i>	<i>RNAi CP190</i>
CP190	-1.038	-1.143	-4.424
<i>Ibf-1</i>	-2.837	1.875	-1.047
<i>Ibf-2</i>	1.776	-4.904	-1.031

En quant al quadre de gens diferencialment expressats (Taula 6), el primer que s'observa és totes tres mostres (*Ibf-1*, *Ibf-2* i CP190) mostren variacions en l'expressió gènica respecte el control. En segon lloc s'observa que totes tres presenten gens *up-regulats* (gens l'expressió dels quals ha augmentat) i gens *down-regulats* (gens l'expressió dels quals ha disminuït), tot i que en tots tres casos hi ha més gens *up-regulats* que no pas a la inversa. Tant és així que en el cas de la mostra CP190 n'hi ha més del doble respecte els *down-regulats*. En el cas de l'RNAi de *Ibf-1*, a on hi ha molt pocs gens diferencialment expressats, no podem descartar que sigui degut a que és la proteïna l'RNAi de la qual ha funcionat pitjor de les tres, per tant com que el nivell de la proteïna ha baixat menys també veiem menys efectes en el microarray. Si a això li sumem que estem treballant amb només dues rèpliques i per tant com ja hem explicat estem sub-estimant els canvis, podríem suposar que aquest baix nombre de gens està significativament sub-estimat.

Taula 6.- Gens significativament upregulats i downregulats. Nombre de gens que han augmentat o disminuït la seva expressió en cadascun dels tres knockdowns. Es consideren gens diferencialment regulats aquells que que fold-change és major que 1.25 o menor que -1.25 amb una probabilitat major que 0.75.

lbf-1		lbf-2		CP190	
<i>Up</i>	<i>Down</i>	<i>Up</i>	<i>Down</i>	<i>Up</i>	<i>Down</i>
14	7	43	39	114	52

Ens hem preguntat si hi havia enriquiment d'un cert tipus de gens en particular, però l'anàlisi de Gene Ontology (no mostrat) no dona cap resultat conclouent: la diversitat de grups afectats és molt alta i no n'hi ha cap de significativament molt més enriquit que els demés. Aquest resultat era esperable donat que no creiem trobar-nos davant de proteïnes que segueixin un programa de regulació en concret o formin part d'una via de senyalització, sinó que més aviat les proteïnes d'*insulator* tenen un rol mecànic.

En el mateix sentit ens hem preguntat si afectaven a grups de gens que formin clusters al genoma, o simplement si afecten a regions determinades del genoma de manera particularment enriquida. Aquest anàlisi de clusters no ha donat cap resultat significatiu: no hi ha clusters de gens que estiguin diferencialment expressats en cap dels tres experiments.

4.2.3.12.3 Co-regulació

Atès que totes tres proteïnes formen un complex que podria estar treballant regulant la cromatina de manera conjunta, ens preguntem fins a quin punt la variació global de l'expressió gènica varia en la mateixa direcció per les tres mostres. És a dir, si en un knockdown un gen es downregula, també ho hauria de fer als altres dos knockdowns, i si s'upregula també ho farà a als altres dos. D'això es prediu que si posem en un gràfic cada gen sent l'eix X un knockdown i l'Y un altre, la majoria estaran

als quadrants inferior esquerre i superior dret. A la figura 53 cada gen és un punt i cada eix és el resultat obtingut a cada microarray per tots els gens.

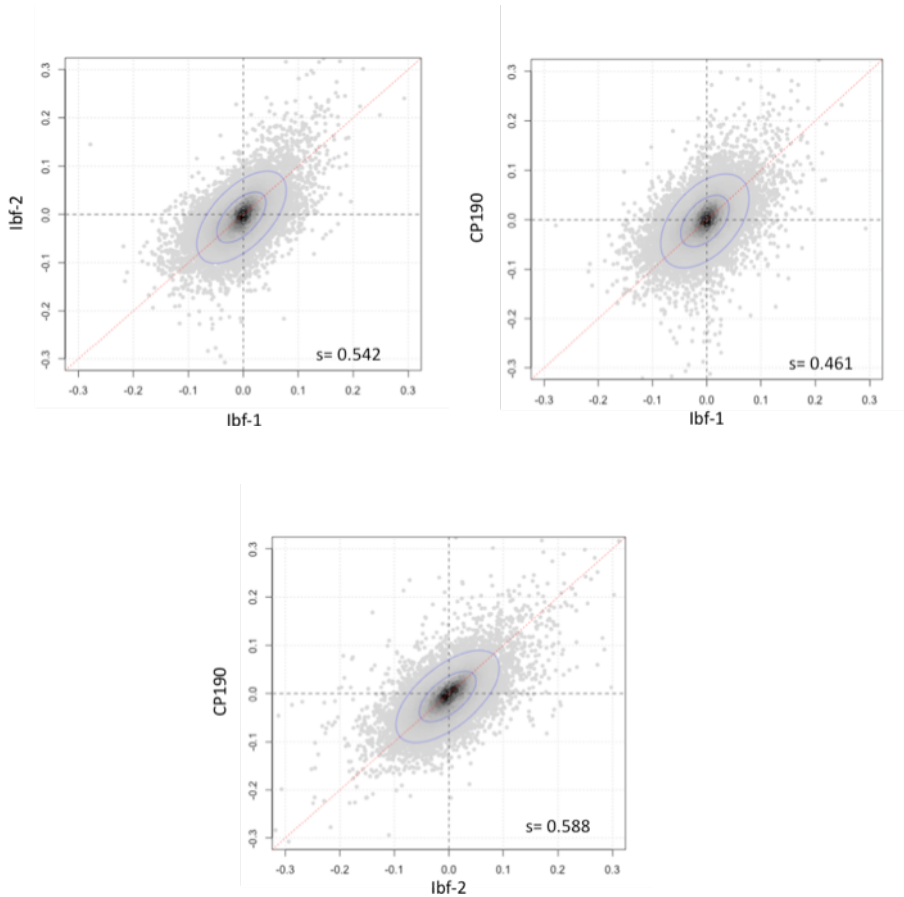


Figura 53.- Correlacions d'Spearman entre els tres knockdowns. Cada punt representa un gen, situat al gràfic en funció del log₂ fold-change respecte LacZ de dues condicions a l'eix x i y. El color dels punts és representatiu de la densitat d'aquests. Com més fosc més acumulació de punts. A sota a la dreta s'indica amb una s el valor de la correlació d'Spearman. El primer cercle blau indica el 50% dels punts, el segon cercle el 90%, i a línia vermella correspon a una correlació de Spearman de 1. Tots tres valors tenen un p-value de <0.0001. A) La correlació entre el knockdown d'lbf-1 i lbf-2 és de 0.542, B) entre lbf-1 i CP190 de 0.461 i C) entre lbf-2 i CP190 de 0.588.

N'hem quantificat la correlació calculant el coeficient de correlació d'Spearman. En aquest test, si X augmenta i Y també o si X disminueix i Y també, obtindrem un valor positiu entre 0 i 1. En canvi si quan una variable augmenta l'altra disminueix s'obté un valor negatiu, entre 0 i -1. Si no hi ha cap tendència que fagi que Y augmenti i disminueixi quan X augmenta, llavors el valor d'Spearman és de 0.

La correlació entre les variacions d'expressió de gens en el knockdown de *Ibf-1* i el knockdown de *Ibf-2* dona un coeficient d'Spearman de 0.542, és per tant positiu indicant una correlació positiva. La correlació de les variacions d'expressió de gens de la mostra *Ibf-1* respecte CP190 dona un coeficient de 0.461, i de *Ibf-2* amb CP190 dona un coeficient de 0.588. Els tres valors són positius, per tant podem concloure que els canvis d'expressió gènica es produeixen majoritàriament en el mateix sentit en els experiments d'RNAi de *Ibf-1*, *Ibf-2* i CP190, suggerint que les tres proteïnes actuen en conjunció sobre la regulació de l'expressió gènica.

El valor de correlació que hem calculat (coeficient d'Spearman) no és no obstant el valor real de co-regulació. Això es deu a que la variabilitat experimental de la tècnica i la variabilitat de les dades baixa la correlació màxima teòrica de 1 a un valor que desconexim. Volem saber quant lluny està el valor de correlació observat del valor màxim de correlació que els nostres experiments permeten, i això ens donarà una estimació del percentatge de co-regulació entre dos knockdowns.

Per a tal propòsit generem una estructura de dades que reflecteixi els valors estadístics observats per cada gen als microarrays (mitjana, varianza) i de cada gen n'agafem un valor a l'atzar i en calculem la correlació d'Spearman. Fem aquesta simulació 1000 vegades per cada gen, i la fem primer de tot sota la hipòtesi que no hi ha co-regulació: cada gen es comporta com si no tingués res a veure amb el mateix gen de l'altre microarray amb el qual l'estem comparant.

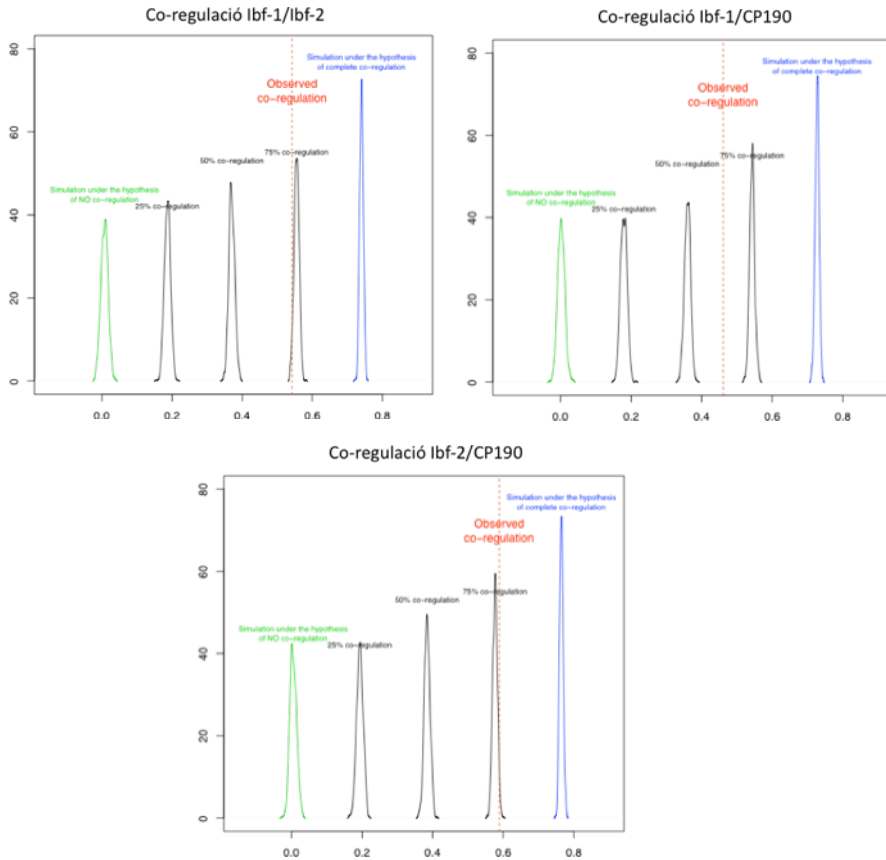


Figura 54.- Valors simulats i observats de co-regulació. Per cada parella de knockdowns, es mostren els intervals de valors de Spearman que ha donat cada simulació segons les hipòtesis de no co-regulació (verd), co-regulació del 25%, co-regulació del 50%, co-regulació del 75% i co-regulació total (blau). La línia de punts vermella indica el valor de Spearman observat. El valor de Spearman observat entre Ibf-1 i Ibf-2 correspon segons les simulacions a una co-regulació de 0.7-0.76 ($pvalue < 0.0001$); entre Ibf-1 i CP190 s'estima una co-regulació d'entre 0.6 i 0.67 ($pvalue < 0.0001$); entre Ibf-2 i CP190 la co-regulació estimada és entre 0.74 i 0.79 ($pvalue < 0.0001$).

És a dir, provenen de dues distribucions diferents, cadescuna amb la seva mitjana i desviació. Amb els nous valors d'expressió simulats calculem els log2 fold changes simulats. Això dona un interval de valors que representen la correlació mínima esperable, al voltant de zero. Ara repetim la simulació però aquesta vegada sota la

hipòtesi que hi ha co-regulació completa: cada gen es comporta com si formés part de la mateixa entitat que el gen de l'experiment amb el que comparem. És a dir, provenen de la mateixa distribució amb igual normal i desviació. Això dóna un interval de valors de Spearman que representen la màxima correlació que es podria obtenir teòricament amb les nostres dades. Seguint la mateixa lògica podem calcular els valors teòrics de 25%, 50% i 75% de co-regulació.

Així, quan comparem el valor de Spearman observat amb els valors obtinguts de les simulacions podem inferir que entre lbf-1 i lbf-2 hi ha entre un 0.7 i 0.76 de co-regulació, entre lbf-1 i CP190 entre un 0.6 i 0.67 i entre lbf-2 i CP190 hi ha entre un 0.74 i un 0.79 de co-regulació (Figura 54).

4.2.3.13 Anàlisi exhaustiu de la localització de Ibf-1, Ibf-2 i CP190 al genoma

Sabent que es tracta de proteïnes que s'uneixen directa o indirectament al DNA i presumiblement a seqüències reguladores, la següent pregunta era a on exactament del genoma s'unien, a quants llocs i quina era la naturalesa d'aquests llocs. Finalment també volíem saber si, ja que tenen un domini d'unió al DNA, Ibf-1 i Ibf-2 tenien afinitat per alguna seqüència en concret tal i com és el cas de Su(Hw), dCTCF i BEAF (Van Bortle et al., 2012; Yang et al., 2012).

Vam realitzar immunoprecipitació de cromatina seguida de seqüenciació de tot el DNA immunoprecipitat (ChIP-seq) partint de cèl·lules SL2 i utilitzant els anticossos específics de Ibf-1, Ibf-2 i CP190. Aquest experiment permet conèixer amb gran nivell de precisió tots els llocs en que una proteïna s'uneix al genoma. A més, disposem de les dades publicades a Modencode on hi ha els resultats dels ChIPs de les altres proteïnes d'*insulator* de *Drosophila*, fets també a partir de cèl·lules SL2, per poder comparar els resultats.

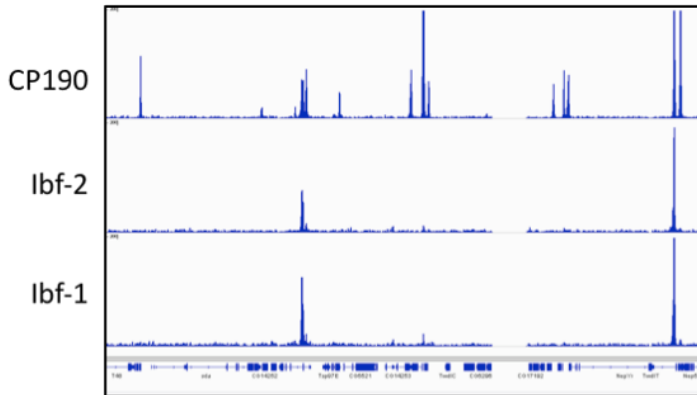
Taula 7.- Pics obtinguts al ChIP-Seq. Per cadascun dels ChIP-seq (Ibf-1, Ibf-2 i CP190) s'indiquen el nombre de pics i la longitud mitja dels pics en parells de bases.

	<i>nombre de pics</i>	<i>longitud mitja dels pics</i>
CP190	2060	273.8 bp
Ibf-1	616	229.7 bp
Ibf-2	783	274.4 bp

Com es mostra a la figura 55 i a la Taula 7, els pics obtinguts (intervals de DNA on hi ha unió de la proteïna) són força discrets: de mitjana estan entre els 229 bp (Ibf-1) i els 274 bp (Ibf-2). Per qui observem més pics és CP190 (2060), seguit de Ibf-2 (783) i Ibf-1 (616). Trobem casos on hi ha pics de les tres proteïnes al mateix lloc, casos on el

pic és més intens per lbf-1 i lbf-2 que per CP190 i viceversa, casos on no hi és CP190 i molts casos on hi ha CP190 però no lbf-1 o lbf-2. Ens trobem sovint també amb dobles pics: pics que en realitat són dos però molt junts.

A)



B)

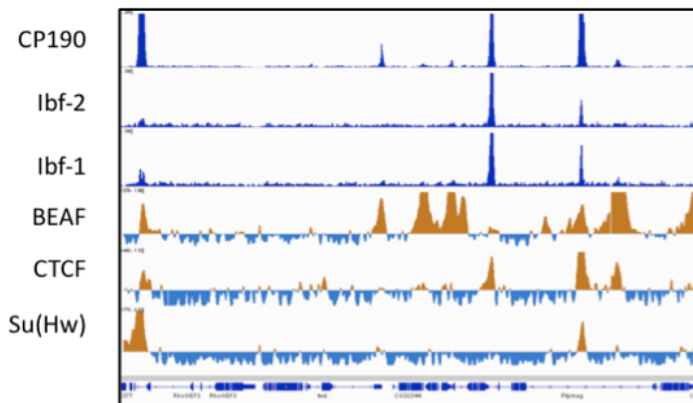
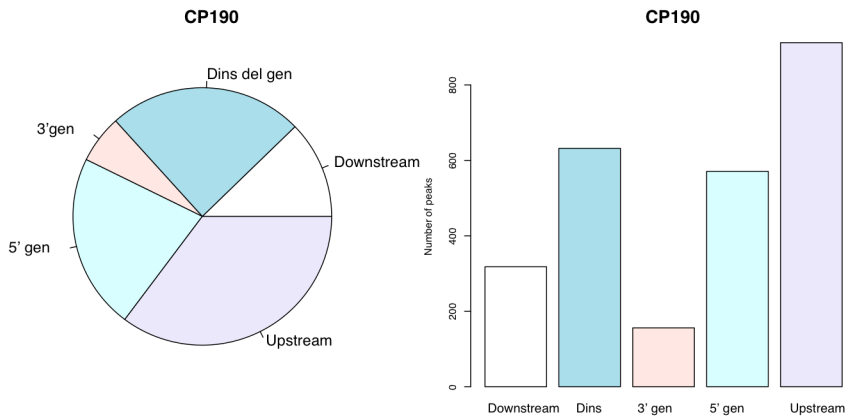


Figura 55.- Exemples de pics al genoma. A) Es mostren els carrils de CP190, lbf-1 i lbf-2 en una regió ampliada del cromosoma 3. Els pics d'lbf-1 i lbf-2 coincideixen entre ells i amb pics de CP190, mentre que CP190 té altres pics on no coincideix amb les altres dues; B) es mostren a més carrils de BEAF, CTCF i Su(Hw) —del Modencode— on s'observa que hi ha regions on hi són totes les proteïnes, i altres on es van combinant.

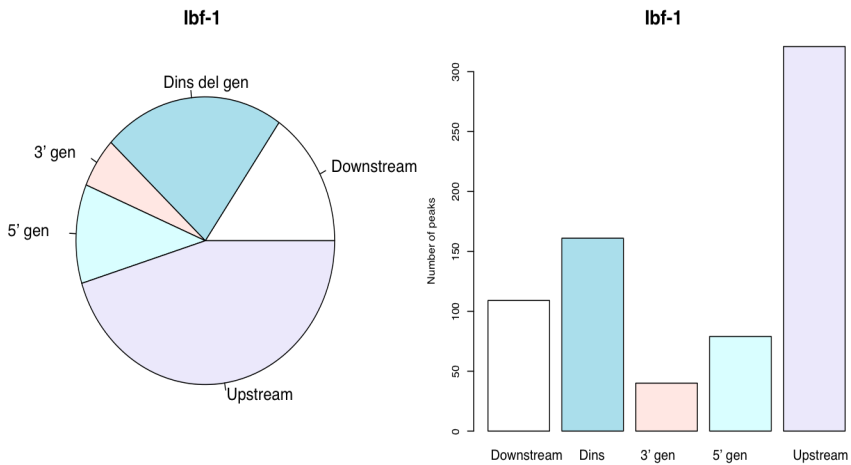
4.2.3.13.1 Localització al genoma segons el tipus de regió

Analitzem la localització de tots els pics en cada ChIP-Seq segons el lloc on es troben en relació al gen més pròxim: upstream; downstream; sobreposat a l'inici 5' del gen o al final 3'; dins del gen.

A)



B)



C)

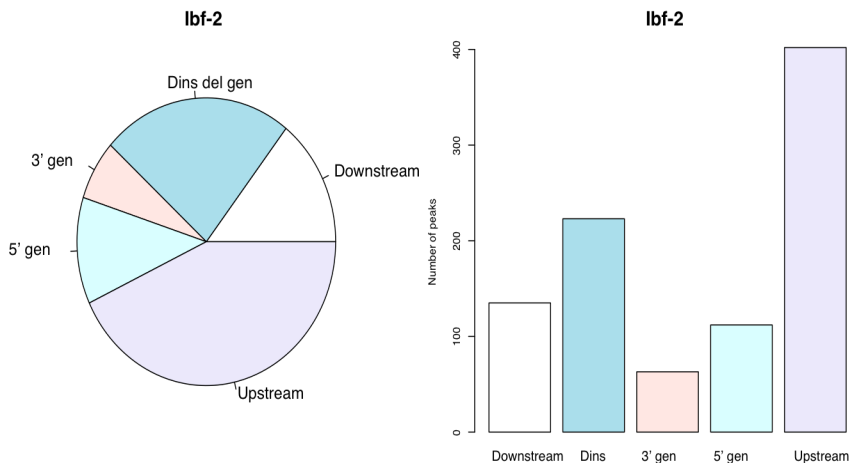


Figura 56.- Localització segons la regió genòmica. Es mostra per CP190 (A), lbf-1 (B) i lbf-2 (C) les distribucions dels seus pics en funció de classificar el genoma en regions segons la seva posició respecte els gens. 5' gen és al voltant del TSS; 3' gen és al final del gen; dins del gen és entre aquests dos punts siguin introns o siguin exons; Upstream és quan la part del gen més pròxima al pic és el seu 5', mentre que Downstream és quan la regió del gen més pròxima és el final del gen. Upstream i Downstream junts conformarien les zones intergèniques, i 3'gen, 5'gen i Dins del gen serien les regions gèniques.

En el cas de CP190 trobem que la majoria es troben en posició intergènica upstream del gen més pròxim (Figura 56-a). No obstant hi ha una gran quantitat també a la zona del TSS i a dins del gen. Una porció menor es troben downstream dels gens i encara n'hi ha menys que coincideixen amb el final dels gens.

Tant si observem lbf-1 com lbf-2 ens trobem amb algunes diferències respecte CP190 (Figura 56-b i c). La posició preferent dels pics torna a ser upstream dels gens, però aquest cop la fracció és encara més gran: pràcticament la meitat dels pics es troben en aquesta regió. En segon lloc, una cinquena part dels pics estan a dins dels gens. I en tercer lloc hi ha els pics situats downstream dels gens, una gran diferència

amb CP190. Finalment hi ha aquells que es troben a l'inici del gen i un últim i minoritari grup és el dels que es troben al final del gen.

Per tant el que veiem és que lbf-1 i lbf-2 es troben més sovint que CP190 a regions intergèniques ja siguin upstream o downstream del gen més proper. CP190 en canvi té una fracció més important de pics situada al voltant del TSS. En tots tres casos hi ha una part substancial situada dins del gen.

4.2.3.13.2 Posicionament sobre els gens

El següent que ens hem preguntat és quin perfil mostren aquestes tres proteïnes sobre els gens, vist que una gran part dels pics es troba al voltant del TSS, però que també una part important està dins dels gens -introns i exons- i fins i tot hi ha una part significativa al final dels gens.

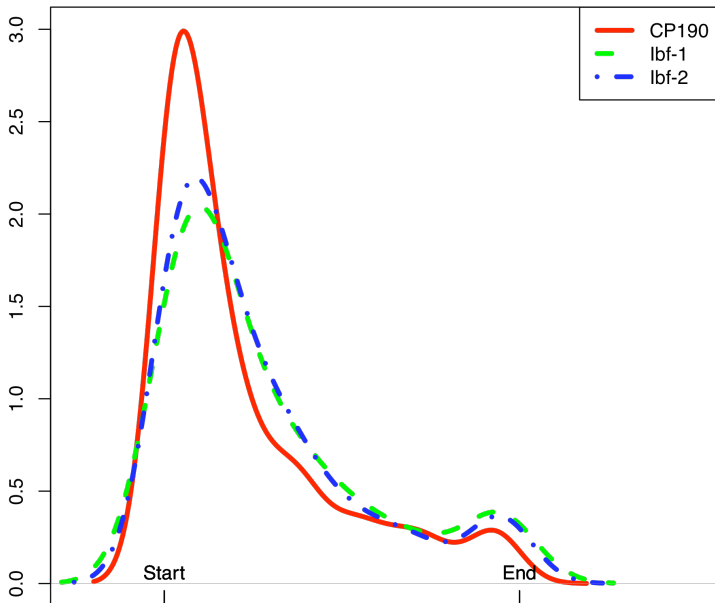


Figura 57.- Localització dels pics sobre els gens. Representació de la localització d'lbf-1, lbf-2 i CP190 al llarg dels gens. L'eix de les Y és una mesura de la densitat de pics a cada punt de l'eix X, que és una estandardització dels gens de genoma.

El que mostra aquest perfil sobre un gen estandarditzat és que tant per lbf-1 com lbf-2 com CP190 tenen una clara preferència per l'extrem 5' (Figura 57). A partir d'aquí la corba disminueix fins a l'extrem 3', on fa un últim pic, clarament més petit que el pic inicial. Observem no obstant una diferència entre el perfil de CP190 i el mostrat per lbf-1 i lbf-2: en aquest últim cas el pic a 5' no és tant marcat i el nivell mostrat per la corba dins del gen és més elevat que en CP190, així com el pic a 3'. És a dir, que lbf-1 i lbf-2 semblen estar menys preferentment localitzades ben a l'inici del gen, i per contra es localitzen més sovint al llarg del gen així com al seu final.

4.2.3.13.3 lbf-1 i lbf-2 colocalitzen quasi totalment entre elles i amb CP190

Hem volgut saber amb exactitud quantes vegades lbf-1 i lbf-2 s'uneixen als mateixos llocs al DNA. Hem dissenyat un diagrama de Venn (Figura 58) on es mostren els pics que es solapen en cada combinació entre les tres proteïnes estudiades.

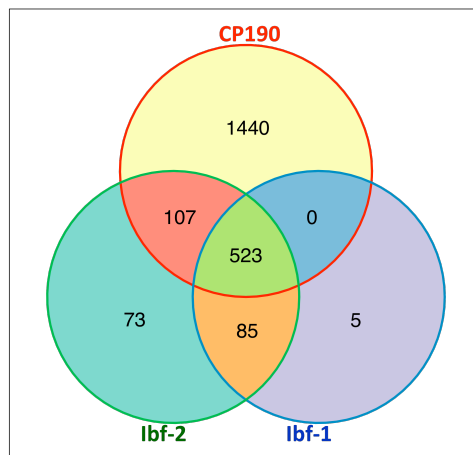


Figura 58.- Diagrama de Venn de les tres proteïnes del complex. La majoria de pics d'lbf-1 es solapen amb lbf-2 i CP190. La majoria de pics d'lbf-2 també, però hi ha un nombre important de pics no solapats (73) amb cap altra. N'hi ha 85 d'lbf-1 i lbf-2 que no coincideixen amb CP190, però el cas majoritari és que hi coincideixin (523 + 107). CP190 té molts més pics que lbf-1 i 2 i que per tant no hi coincideixen (1440).

El que s'observa concorda bastant amb el que s'observava als cromosomes politènics. CP190, que té molts més llocs d'unió, solapa en una porció d'aquests (30.4%) amb Ibf-1 i Ibf-2, i aquestes al seu torn coincideixen amb CP190 majoritàriament (85.3% Ibf-1 i 80% Ibf-2). Entre elles, Ibf-1 i Ibf-2 es solapen extensament. Cal tenir en compte que hi ha menys pics de Ibf-1 que de Ibf-2, potser degut a factors experimentals. Sigui com sigui, el 99.2% dels llocs Ibf-1 estan també ocupats per Ibf-2. Ibf-2, amb més pics, té un 22.8% de llocs on no hi ha pic de Ibf-1. Dels llocs que coincideixen Ibf-1 i Ibf-2 entre elles, un 86% es solapen amb CP190.

4.2.3.13.4 Ibf-1 i Ibf-2 colocalitzen parcialment amb les demés proteïnes d'*insulator*

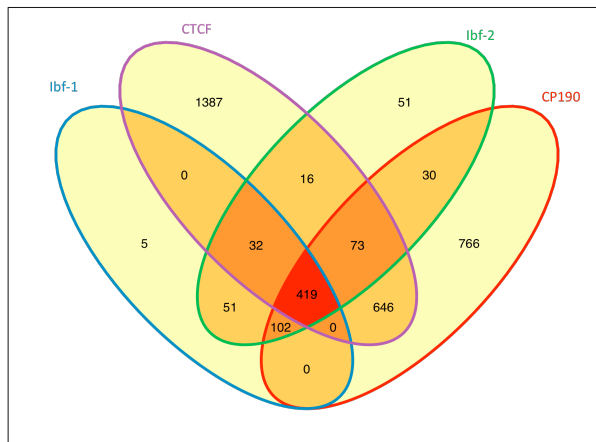
Com se sap, CP190 colocalitza a un bon nombre dels seus llocs d'unió amb les demés proteïnes d'*insulator* (Bushey et al., 2009; Schwartz et al., 2012) amb qui coopera per dur a terme la funció *insulator* (Pai et al., 2004; Gerasimova et al., 2007). De la mateixa manera, vam voler investigar si aquest era el cas amb el nostre CHIP-seq de CP190 i els de Ibf-1 i Ibf-2. Vam estudiar els solapaments entre aquestes tres proteïnes i CTCF, BEAF i Su(Hw) i ho vam representar en diagrames de Venn.

Al primer diagrama (Figura 59-a) s'observa com hi ha una forta colocalització de les quatre proteïnes CP190, Ibf-1, Ibf-2 i CTCF de 419 llocs, contra els 102 en que es troben CP190, Ibf-1 i Ibf-2 però no CTCF. D'altra banda, CTCF i CP190 coincideixen en molts altres llocs on no hi ha Ibf-1 i Ibf-2 (646). Com s'observa a la figura 59-d és molt més freqüent que CP190, Ibf-1 i Ibf-2 es trobin amb CTCF que no pas sense ella. Aquesta colocalització podria tenir algun sentit funcional o podria ser fruit simplement que CTCF té una distribució genòmica similar a la de Ibf-1 i Ibf-2.

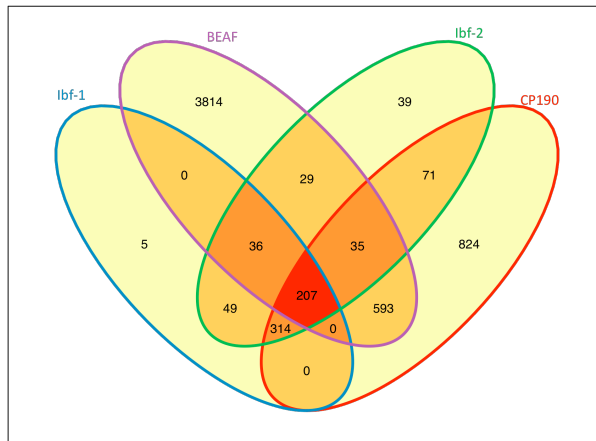
Al segon diagrama (Figura 59-b) es veu que la colocalització de les tres proteïnes del complex es dóna majoritàriament sense solapament amb BEAF, com també s'observa al gràfic 59-d. Al tercer gràfic (Figura 59-c), amb Su(Hw) es torna a repetir el mateix però encara més extrem, amb només 136 llocs comuns a les quatre

proteïnes. Per tant concloem que, tot i que CP190, Ibf-1 i Ibf-2 colocalitzen substancialment amb les altres tres proteïnes d'insulator, ho fan preferencialment amb CTCF, menys amb BEAF i encara menys amb Su(Hw), subratllant el fet que Ibf-1 i Ibf-2 es troben en un punt mig de la distribució fortament intergènica de Su(Hw) i la fortament gènica de BEAF, tal i com és el cas de la distribució genòmica de CTCF.

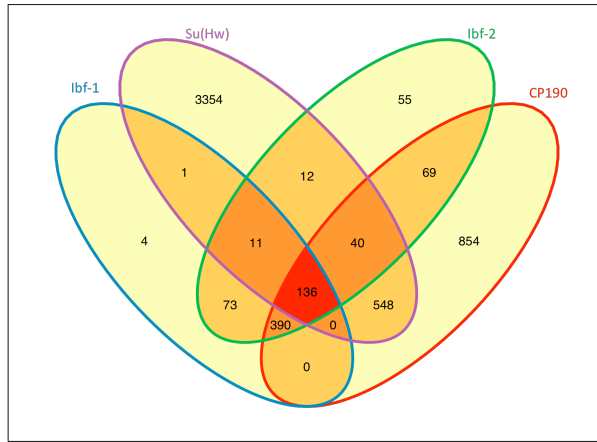
A)



B)



C)



D)

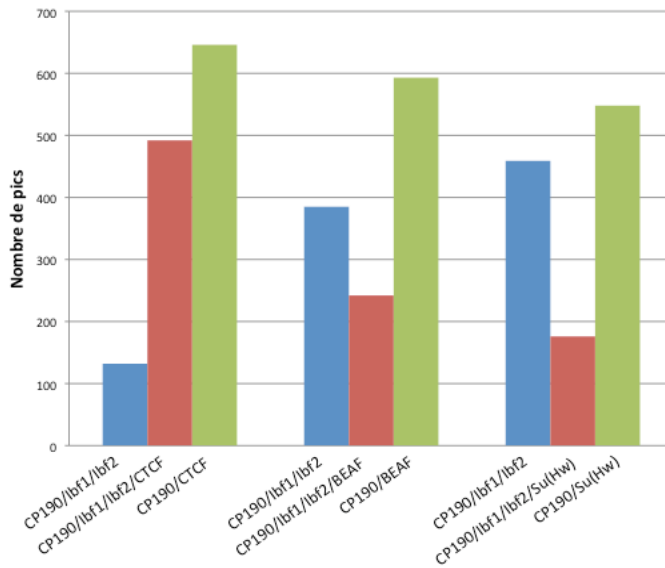


Figura 59.- Solapaments de les Ibf-1, Ibf-2 i CP190 amb les altres proteïnes d'insulator. Els diagrames de Venn mostren el solapament de Ibf-1, Ibf-2 i CP190 amb CTCF (A), BEAF (B) i Su(Hw) (C). En aquests diagrames les xifres indiquen el nombre de pics que es troben en cada combinació. Els resultats es sumarien en un gràfic (D) en el que es redueixen les combinacions a tres possibilitats en cada cas.

Schwartz i col·laboradors (Schwartz et al., 2012) van proposar l'existència de diferents classes de binding sites agrupats per combinacions específiques de proteïnes insulator. Nosaltres mirem quants dels pics dels nostres ChIP-seq es troben a cada combinació (Figura 60). Per a fer-ho agafem cadascun dels pics que corresponen a cada grup, segons els seus resultats, i mirem si són pics pertanyents a lbf-1 o lbf-2. El que veiem és que dels pics pertanyents al grup CP190 + CTCF, el 62% tenen presència d'lbf-1 i el 71% d'lbf-2, corroborant la observació que aquestes quatre proteïnes tenen molts llocs en comú. Molt presents també ho estan a aquells pics que tenen CP190 + CTCF + Su(Hw) + Mod(mdg4) (56% i 68% respectivament) mentre que la presència és molt menor al grup CP190 + Su(Hw) + Mod(mdg4) (4.5% i 5.3%). Per tant concloem que lbf-1 i lbf-2 estan a llocs que a més de coincidir amb CP190 amb qui formen complex, també són llocs on hi ha CTCF i a molts també s'hi sumen Su(Hw) i Mod(mdg4).

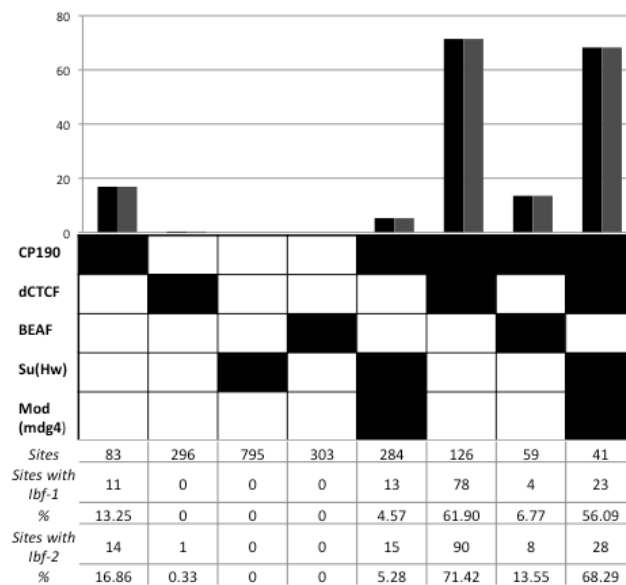


Figura 60.- Quadre de combinacions de proteïnes d'insulator. Hem adaptat les xifres publicades (Schwartz et al., 2012) de les diferents combinacions de proteïnes d'insulator. Prenent les dades dels intervals representats per cada pic a cada combinació hem identificat aquells intervals que es solapen amb intervals de lbf-1 o lbf-2.

Hem volgut saber, en el context de totes les proteïnes de la cromatina de les que hi ha publicades dades de CHIP al ModEncode, amb quines s'assemblen més Ibf-1 i Ibf-2 en quant a posicionament al genoma.

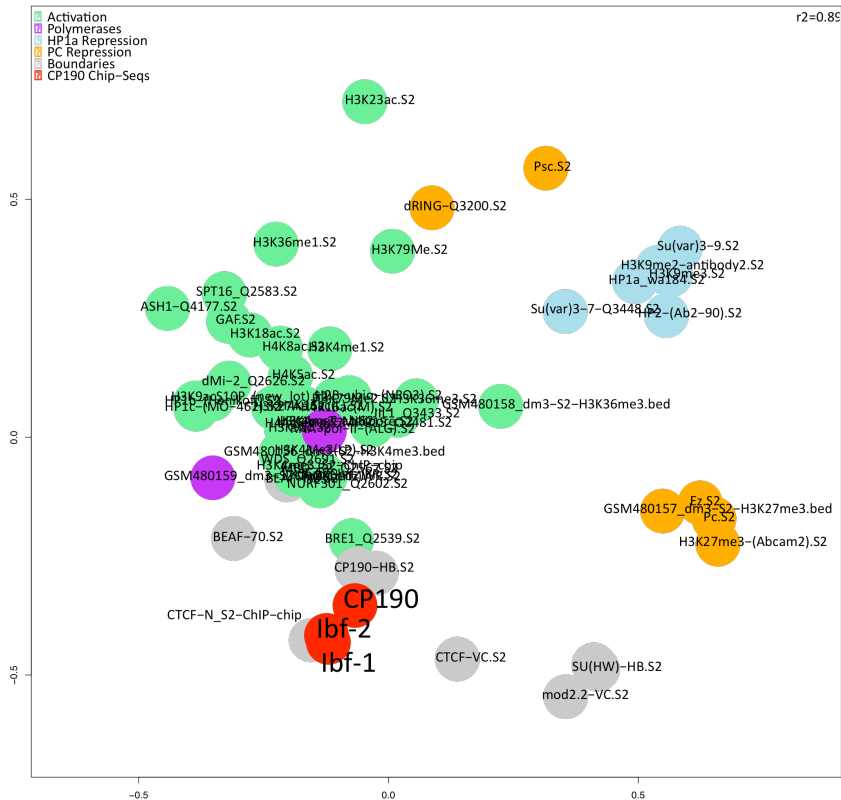


Figura 61.- Posicionament d'Ibf-1, Ibf-2 i CP190 respecte altres proteïnes de la cromatina. Utilitzant el software ChroGPS, que posiciona en un gràfic les proteïnes de la cromatina en funció de la seva localització al genoma i per tant agrupa en clusters aquelles que coincideixen més i allunya aquelles que coincideixen menys. Es mostren totes les proteïnes de les que es disposen les dades de CHIP al ModEncode, i els colors faciliten la visualització d'agrupaments posicionals-funcionals (en verd proteïnes d'activació; en groc de Polycomb; en blau de repressió per HP1a; en lila les polimerases; en gris les proteïnes d'insulator i en vermell les tres proteïnes que hem analitzat nosaltres). A sobre de cada cercle s'indica l'experiment de ModEncode a partir del que s'han tret les dades, indicant la proteïna i sovint l'anticòs i la línia cel·lular. La posició de les tres proteïnes que nosaltres hem analitzat, en vermell, correspon a la zona on hi ha les demés proteïnes d'insulator, en gris.

Hem utilitzat chroGPS (Reina and Rossell, 2012) per col·locar els nostres CHIP-seq en un mapa global de totes les proteïnes posicionades en un gràfic de dos dimensions segons els seus llocs d'unió a la cromatina (Figura 61). En aquest gràfic les proteïnes tenen tendència a ajuntar-se segons si coincideixen més o menys al genoma, i això acaba reflectint els grups funcionals als que pertanyen. Per exemple hi ha el grup de les proteïnes relacionades amb activació, les de silenciament Polycomb, les de silenciament HP1a, i hi ha també un grup separat de la resta on hi ha les proteïnes d'insulator. Hem observat que Ibf-1 i Ibf-2 cauen clarament en aquest grup, al costat de CP190 i de CTCF.

4.2.3.13.5 Localització respecte els promotors

Hem detectat que Ibf-1, Ibf-2 i també CP190 molt sovint es troben sobre gens que tenen més d'un promotor (Figura 62-a). Com s'ha explicat a la introducció, aquests s'anomenen promotors alternatius i donen lloc a trànscripats diferents tot i tractar-se del mateix gen. És un sistema de regulació gènica que ha d'estar finament controlat per a que cada trànscrit s'expressi al teixit i al moment del desenvolupament que li correspon. Els mecanismes moleculars que s'encarreguen de regular-ho són desconeguts.

Hem analitzat la presència de cada pic de cada CHIP-seq als gens de *Drosophila* que tenen promotors alternatius (gens amb més d'un promotor situats a més d'1 Kb de distància entre ells) i hem testat estadísticament si aquesta presència és superior a la que s'esperaria per atzar. En tots tres casos es verifica que les tres proteïnes tenen una presència més alta a gens amb promotors alternatius del que s'esperaria per atzar tenint en compte la proporció d'aquest gens al genoma, amb p-valors de $1.307214e-31$ per Ibf-1, de $1.606708e-38$ per Ibf-2 i $2.253535e-76$ per CP190 (tots tres força inferiors a 0.05, per sota del qual refutem l'hipòtesi nula que la distribució és a l'atzar).

Una altra observació a partir dels CHIP-seq era que, quan Ibf-1 i Ibf-2 es troben a promotors, molt sovint es tracta de promotors head-to-head, que són aquells en que

dos gens que es transcriuen en direccions oposades tenen els promotors propers (Figura 62-b).

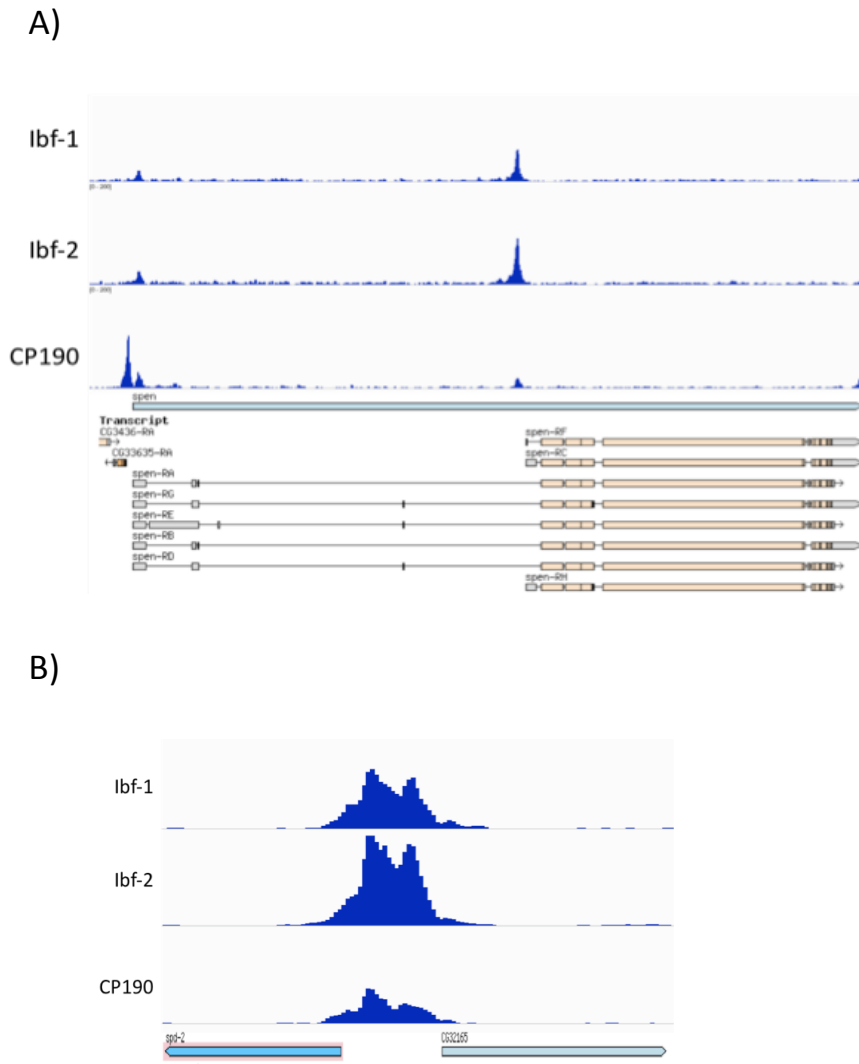


Figura 62.- Localització sobre promotors. A) El gen *spen* té dos promotors alternatius, el primer fa un transcrit llarg amb cinc isoformes, i el segon fa un transcrit més curt amb tres isoformes, i *lbf-1*, *lbf-2* i *CP190* s’hi uneixen; B) Els gens *spd-2* i *CG32165* tenen disposició head-to-head i tenen presència de *lbf-1*, *lbf-2* i *CP190* a l’espai entre els dos.

S'ha vist que aquest és un cas molt freqüent a *Drosophila* i que BEAF està clarament enriquida a aquestes regions (Yang et al., 2012). Així s'ha proposat que BEAF pugui estar duent a terme una funció aïlladora en promotors que són molt pròxims i que han de mantenir els seus propis estats d'expressió. Ens hem preguntat si aquesta és també una característica de lbf-1, lbf-2 o CP190.

Nosaltres hem fet una classificació dels gens head-to-head (direcció de transcripció divergent) i tail-to-tail (direcció de transcripció convergent) propers (a menys d'1Kb de separació uns dels altres) a *Drosophila*. Vam veure que efectivament lbf-1, lbf-2 i CP190 es troben enriquits als espais que hi ha entre dos promotors head-to-head en relació al que s'esperaria per atzar (p -valors < 0.05 en tots tres casos) però no als espais que hi ha entre dos gens convergents tail-to-tail (p -valors > 0.05).

4.2.3.13.6 lbf-1 i lbf-2 s'uneixen a una seqüència consens

Per a saber si lbf-1 i lbf-2 reconeixen una seqüència determinada a l'hora d'unir-se al DNA, si és que s'hi uneixen directament, hem realitzat una cerca de motius enriquits dins dels intervals delimitats pels pics obtinguts al ChIP-seq utilitzant rGadem (Figura 63). Com era d'esperar, alguns dels motius enriquits que trobem (Figura 63-c i d) ja eren coneguts i identificats com a llocs d'unió d'altres proteïnes d'*insulator* (dCTCF, Su(Hw)). Això és normal i valida la cerca, ja que aquestes proteïnes coincideixen amb molts pics de lbf-1 i lbf-2 i per tant les seves respectives seqüències consens es trobaran a bona part dels pics identificats per nosaltres.

També hem trobat altres seqüències consens que no esperàvem. Un exemple és el que s'anomena G-tract (Figura 63-e), una seqüència de 10-20 Guanines seguides que són capaces de formar una estructura secundària del DNA coneguda com a G-quadruplex (G4). El rol d'aquestes G-quadruplex es relaciona amb seqüències reguladores de la transcripció i de la replicació (Bochman et al., 2012). També hi ha un Mirror Repeat (Figura 63-f), que és una seqüència formada per repeticions que són imatges especulars l'una de l'altra i que s'ha relacionat amb la formació de H-DNA, una

estructura de triple hèlix del DNA. Els mirròr repeats s'associen a regulació de la transcripció i a orígens de replicació.

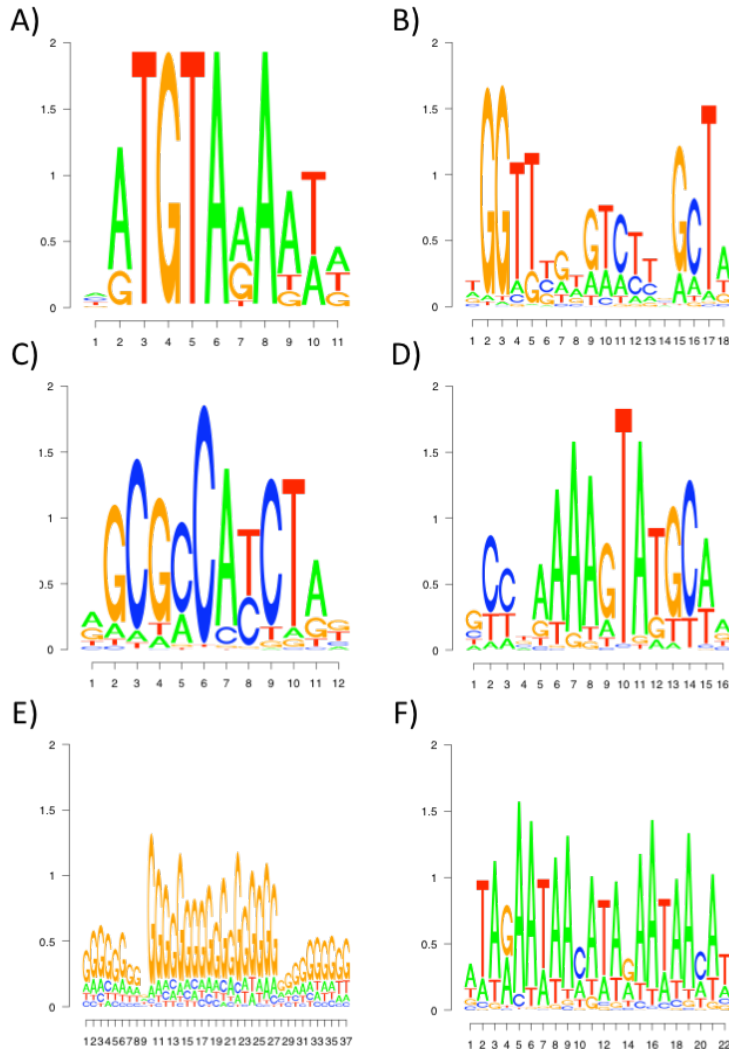
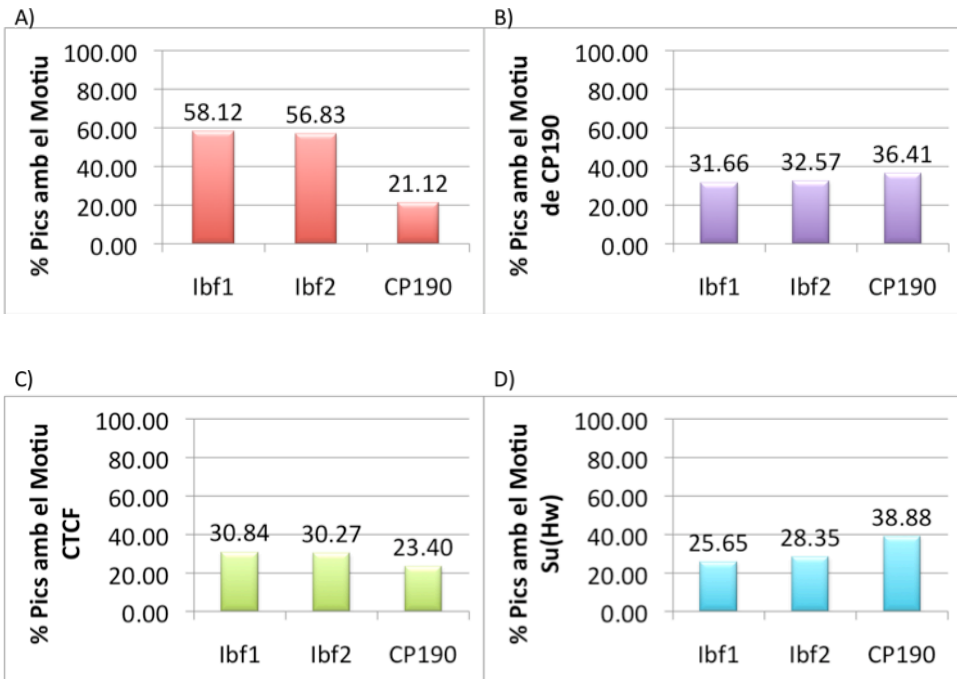


Figura 63.- Seqüències de DNA enriquides als llocs d'unió d'Ibf-1, Ibf-2 i CP190. A) Seqüència no caracteritzada amb anterioritat; B) seqüència prèviament identificada com a possible lloc d'unió de CP190; C) seqüència d'unió de CTCF; D) seqüència d'unió de Su(Hw); E) *G-tract* i F) *mirror repeat*. La cerca i els logos han estat generats amb rGadem.

També trobem una seqüència que no havia estat mai descrita fins ben recentment (Schwartz et al., 2012; Van Bortle et al., 2012) i descoberta en tots dos estudis a través de CHIP-seq d'altres proteïnes d'*insulator* (Figura 63-b). Ha estat atribuït a CP190 sense basar-se en cap evidència molecular, per tant resta per ser corroborada la seva afinitat. Finalment trobem una seqüència (Figura 64-a) que no ha estat mai descrita i que crida la nostra atenció al mirar el quadre on es mostra el nombre d'ocurrències de cada seqüència (Figura 64-a). Apareix com la seqüència més freqüent a Ibf-1 i Ibf-2, mentre que no és de les seqüències més enriquides de CP190. Això ens fa pensar que potser és una seqüència reconeguda per Ibf-1 i Ibf-2 per unir-se al DNA, l'anomenem Motiu Ibf.



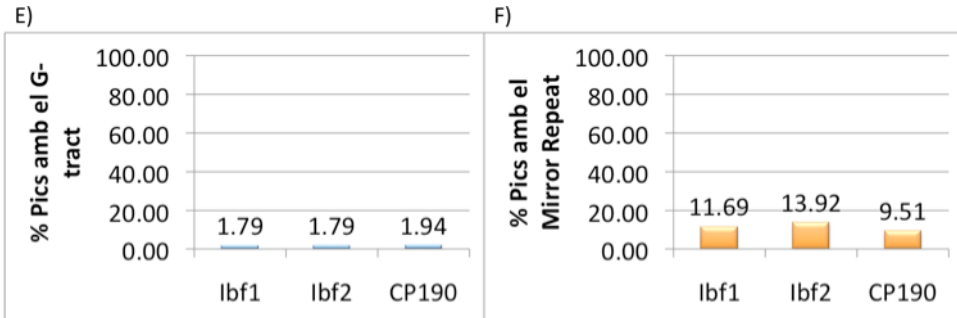


Figura 64.- Percentatge de pics en cada CHIP-seq que contenen cadascun dels motifs. Els gràfics mostren el percentatge de pics contenint com a mínim un cop cadascuna de les seqüències trobades a la cerca.

Per a comprovar-ho realitzem un assaig in vitro, Electrophoretic Mobility Shift Assay (EMSA) o gel de retard. Consisteix en córrer en un gel d'acrilamida una sonda de DNA marcada radioactivament en presència i en absència d'una proteïna. Si no hi ha unió, la sonda correrà pel gel a la mateixa velocitat en tots dos casos. Ara, si hi ha unió, s'observa que la sonda que té presència de proteïna corre més lentament que la sonda sola. A la figura 65 s'observa com la sonda que correspon al Motiu lbf corre més lentament en presència d'extracte de proteïna lbf-1 i també de lbf-2.

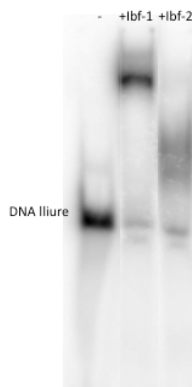


Figura 65.- Gel de retard amb el Motiu lbf. Al carril de l'esquerra es corre la sonda de DNA sense proteïna. Al carril del mig es corre en presència d'extracte de lbf-1, i al de la dreta amb lbf-2. En tots dos casos observem un retard respecte el carril de l'esquerra, indicant unió DNA-proteïna.

Per a verificar l'especificitat de la unió repetim l'experiment (Figura 65) en presència de sonda competidora (en excés) i freda (no radioactiva). Utilitzem també una sonda que és la mateixa però que té una mutació de seqüència al Motiu lbf (competidor inespecífic). Veiem com només en el cas de competició específica es perd el retard en la banda, ja que la proteïna s'ha unit al competidor fred, cosa que no ha passat en el cas del competidor inespecífic demostrant l'especificitat de l'afinitat de lbf-1 i lbf-2 per aquest Motiu. Quan hem mirat quina afinitat tenen lbf-1 i lbf-2 per altres Motius trobats en la cerca realitzant el mateix experiment el resultat és que només semblen tenir una clara afinitat pel Motiu (Figura 67).

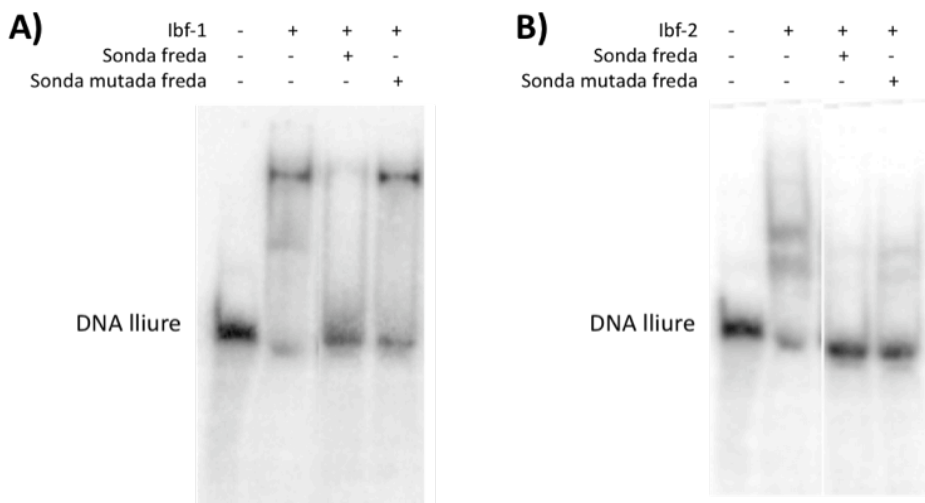


Figura 66.- Gel de retard amb el Motiu lbf amb competidors. Hem fet un gel de retard per provar l'especificitat de la unió de lbf-1 (A) i lbf-2 (B) amb el Motif lbf. Al carril de l'esquerra s'hi corre la sonda lbf sense proteïna, al segon carril s'hi afegeix la proteïna, al tercer carril s'hi afegeix la proteïna i sonda en excés i no radioactiva, al quart carril en lloc d'aquesta sonda se n'afegeix una que està mutada al lloc d'unió i que per tant es pot considerar inespecífica, també freda.

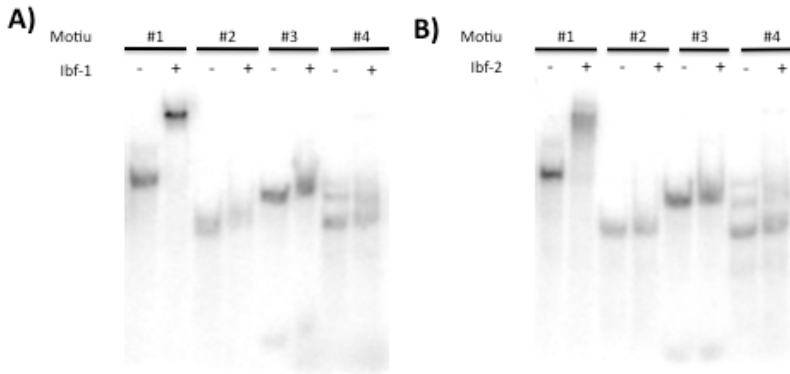


Figura 67.- Gel de retard de diferents seqüències d'unió. Per cada seqüència testada es mostra lliure (esquerra) i amb proteïna (dreta). La sonda #1 és una seqüència que conté 3 Motifs lbf. La sonda #2 és la seqüència mutada que s'ha utilitzat de competidor en la figura anterior. La sonda #3 és el motiu (B) de la figura 63 que ha estat proposat com a motiu d'unió de CP190. La sonda #4 és el motiu (F) de la figura 63, el mirror repeat.

Quan analitzem quantes vegades coincideix aquest motiu amb els demás motius d'*insulator* identificats veiem que amb qui més vegades coincideix és amb el de CTCF, seguit del recentment descobert putatiu de CP190 i finalment pel de Su(Hw) (no mostrat). Si analitzem els pics de lbf-1 i 2 veiem com el Motiu hi és present independentment de la presència de CP190 (Taula 8-b). En canvi si n'analitzem la presència als pics de CP190, hi és en un 53.9% a aquells que coincideixen amb lbf1 i lbf-2, mentre que només es troba amb un 7% a la resta, confirmant l'afinitat d'aquest motiu pels llocs lbf. A la Taula 8-a veiem que de mitjana hi ha al voltant de 1.5 motius a cada pic, volent dir que en molts casos hi serà per duplicat en un mateix lloc d'unió.

Taula 8.- A) Nombre mitjà de Motif lbf a cada pic en cada CHIP-seq; B) Presència del Motiu lbf en funció de la presència de CP190 i de lbf1 i 2.

A)

	binding sites/pic
lbf-1	1.56
lbf-2	1.63
CP190	1.43

B)

	% de pics amb Motiu Ibf
Ibf1/2 sense CP190	62.73
Ibf1/2 amb CP190	55.30
CP190 sense Ibf1/2	7.01
CP190 amb Ibf1/2	53.87

4.2.3.14 Activitat enhancer-blocking d'Ibf-1 i Ibf-2

Per analitzar si Ibf-1 i Ibf-2 són necessàries per l'activitat enhancer-blocking hem fet ús dels dos mutants dels que disposem. El mutant d'Ibf-1, Δ Ibf-1, és una deleció que també elimina el gen adjacent VhaM8.9, i és letal en homozigosi. El mutant d'Ibf-2 prové d'una inserció a dins de l'exó del gen que impedeix que es produeixi la proteïna.

Hem creuat els mutants amb diverses línies transgèniques que porten constructes amb l'enhancer i el gen mini-white separats per un insulator que atenua l'efecte de l'enhancer sobre l'expressió del gen. D'aquesta manera podem monitoritzar el funcionament de l'insulator a través de la pigmentació dels ulls. Si una proteïna que és necessària pel correcte funcionament de l'insulator està mutada o deplecionada llavors aquest no podrà dur a terme l'enhancer-blocking i s'incrementarà l'expressió del mini-white i la pigmentació de l'ull augmenta. Per tant nosaltres volem saber si en creuar aquestes línies transgèniques amb els mutants d'Ibf-1 i Ibf-2 la pigmentació de l'ull augmenta.

Hem utilitzat línies que porten insulators de la regió reguladora d'Abd-B i de les que ja se sabia que necessiten CP190 per funcionar. Hem vist que Ibf-1 és necessari pel funcionament dels insulators Fab-6, Fab-7 i Fab-8 (Figura 68) i quant a Ibf-2 hem vist que és necessari pel Fab-7 i Fab-8 (Figura 69). També hem combinat les mutacions de

Ibf-1 i CP190 (Figura 68) i hem vist que, sobre l'insulator del Fab-7, hi ha una interacció que fa que l'efecte de la combinació sigui més intens que cada mutació per separat.

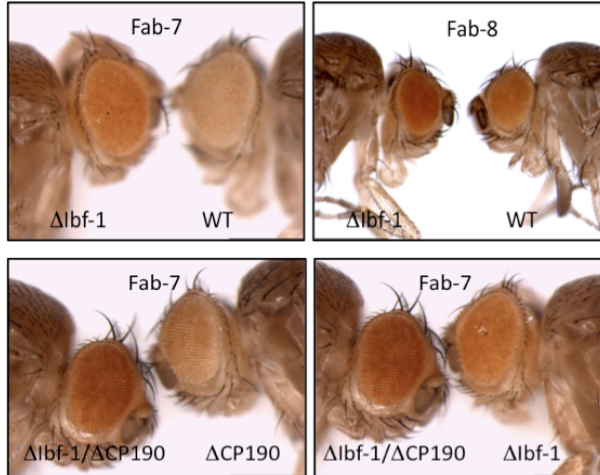


Figura 68.- Ibf-1 és necessari per l'enhancer-blocking al Fab-7. A) La línia Fab-7 té més pigmentació a l'ull si és mutant d'Ibf-1; B) La combinació de mutants d'Ibf-1 i CP190 té més pigmentació que el mutant d'Ibf-1; C) la mateixa combinació també té més pigmentació que el mutant de CP190.

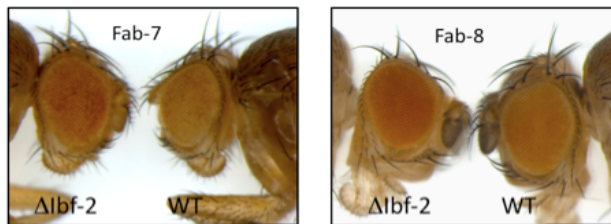


Figura 69.- Ibf-2 és necessari per l'enhancer-blocking al Fab-7 i Fab-8. A) La mutació d'Ibf-2 augmenta la pigmentació de l'ull en línies amb l'insulator Fab-7 i també Fab-8 (B).

Concloem que Ibf-1 i Ibf-2 són necessàries per l'enhancer-blocking dels insulators de la regió reguladora de Abd-B, confirmant que aquestes dues proteïnes presenten una activitat característica de les proteïnes d'insulator.

5.- DISCUSSIÓ

5.1 La regió reguladora d'Abdominal-B

5.1.1 El Fab-6 i l'iab6PRE

El Complex Bitòrax i més concretament la regulació del gen homeòtic Abdominal-B ha estat un dels focus d'atenció de l'estudi del desenvolupament de *Drosophila* des de fa temps (Casanova et al., 1986; Celniker et al., 1989; Boulet et al., 1991). Abdominal-B és un gen model que ha permès entendre mecanismes fonamentals del desenvolupament antero-posterior de l'organisme gràcies a les eines genètiques de les que es disposa amb *Drosophila*. La determinació de la identitat dels segments, la colinearitat o l'expressió additiva de gens homeòtics al llarg de l'eix antero-posterior són conceptes que han estat, si més no parcialment, descrits utilitzant el Complex Bitòrax de *Drosophila* com a model.

Però més enllà de la seva aportació a l'estudi del desenvolupament, la regió reguladora d'Abdominal-B s'ha revelat com un excel·lent model d'estudi dels mecanismes moleculars que sostenen la regulació adequada del gen a cada segment. Aquests mecanismes permeten que una mateixa regió reguladora d'unes 100 kb activi l'expressió d'un sol gen de manera particular a cadascun dels cinc segments a on s'expressa.

Nosaltres hem descrit i caracteritzat alguns dels elements que conformen aquesta regió reguladora. Fins al moment s'havien descrit els enhancers presents a cada iab i els elements Mcp, Fab-7 i Fab-8, que contenen insulators, PREs i PTS (Chen et al., 2005; Maeda and Karch, 2006). Mancava doncs la informació relativa a la separació entre els iab-5 i iab-6 —no se n'havia descrit amb anterioritat ni insulator ni PRE— malgrat ja s'havia especulat amb la seva existència ja que era l'únic límit entre dos iab sense Fab descrit i sí que s'havia detectat presència de CTCF per CHIP al lloc que ha acabat corresponent al Fab-6 (Holohan et al., 2007).

Els assaigs d'hipersensibilitat a la DNAsal fets al laboratori van demostrar que hi havia varies regions hipersensibles entre l'iab-5 i l'iab-6. Això no és rar tenint en compte que el Fab-7, Fab-8 i Mcp estan compostats per més d'una regió hipersensible (Karch et al., 1994; Zhou et al., 1996; Barges et al., 2000). La regió HS4 correspon per localització a l'enhancer ja descrit IAB-6. Aquest ha estat caracteritzat com a element iniciador, que no és suficient per mantenir l'expressió correcta tot al llarg del desenvolupament (Mihaly et al., 2006).

Els treballs previs de la Sílvia Pérez-Lluch al laboratori mostraven que en assaigs de constructes transgènics on es posava l'HS1, l'HS2 i l'HS3 entre un enhancer i un promotor de mini-white, s'observava una clara disminució de la pigmentació dels ulls en els dos primers casos, però no en HS3. A més d'aquest silenciament es va observar pairing-sensitive silencing, altra vegada tant a HS1 com HS2, i les dues activitats depenien de les proteïnes del grup Polycomb. Tot això indicava que les dues regions tenien activitat PRE. La unió de Polycomb i de GAGA a elements reguladors del BX-C havien estat identificades prèviament com a necessaris per la correcta funció de silenciament al PRE del Fab-7 (Hagstrom et al., 1997; Strutt et al., 1997; Schweinsberg et al., 2004) i també al Mcp (Busturia et al., 2001). En els treballs mostrats en aquesta tesi vam demostrar per ChIP que tant Polycomb com GAGA s'uneixen a l'HS1 i al l'HS2, de la mateixa manera que ho fan al Mcp, Fab-7 i Fab-8. Per tant un cop demostrat que totes dues regions eren PRE, calia veure si també tenien activitat insulator i si una de les dues corresponia al Fab-6.

L'assaig d'enhancer-blocking en constructes transgènics té la limitació que el que es pot observar és una disminució de l'expressió del mini-white a través de la pigmentació de l'ull, i això pot correspondre tant a un silenciament del tipus Polycomb com a un bloqueig de l'enhancer típic d'insulator. Tots dos mecanismes, de naturalesa molecular diferent, porten al mateix resultat observable. D'aquesta manera, per tal de dilucidar si HS1 o HS2 tenien activitat enhancer-blocking a part de PRE (que és detectable pel pairing-sensitive

silencing), vam haver de posar aquests constructes en un fons mutant de proteïna d'insulator. Els assaigs publicats amb constructes de Fab-7 i Fab-8 demostraven que aquests perdien capacitat enhancer-blocking si es trobaven en fons mutant de CTCF o de CP190 (Moon et al., 2005; Mohan et al., 2007). Al posar nosaltres els constructes enhancer-blocking en fons mutant de CP190 vam observar que només l'HS1 perdia activitat enhancer-blocking (augmentava la pigmentació de l'ull) i en posar-lo en fons mutant de CTCF i s'observava el mateix resultat, demostrant la presència d'un insulator.

D'aquesta manera quedava demostrat que l'HS1 correspon al Fab-6 i que té activitat insulator i PRE, mentre que l'HS2 només té activitat PRE i corresponia a l'iab6PRE. El Fab-6 de fet és una regió hipersensible gran, comparable a la del Fab-7, i es troba aproximadament equidistant de Mcp i de Fab-7, a 13 kb i a 14 kb respectivament. El fet que el Fab-6 contingui activitat insulator i activitat PRE no és quelcom extrany ja que Mcp, Fab-7 i Fab-8 tenen també un insulator i un PRE. No obstant això, en aquests tres casos s'han pogut delimitar en l'espai les dues activitats. Tant al Fab-7 com al Fab-8 el PRE i l'insulator són fàcils de distingir ja que es tracta de dues regions hipersensibles diferents, tot i que adjacents i molt pròximes (Hagstrom et al., 1996, 1997; Barges et al., 2000). En el cas de l'Mcp no són dues regions hipersensibles diferents però igualment s'han pogut separar a través d'assaigs amb constructes transgènics en els que s'han utilitzat diverses seqüències dins de l'element fins que s'han trobat les seqüències mínimes necessàries per a dur a terme cada funció (Muller et al., 1999; Gruzdeva et al., 2005).

En el nostre cas no vam poder delimitar les dues activitats. D'entrada no és fàcil perquè, a diferència del que passa al Fab-7 i Fab-8, PRE i insulator es trobarien dins de la mateixa regió hipersensible. Vam provar de dividir-la en dos, però totes dues meitats continuaven presentant pairing-sensitive silencing, de manera que no havíem aïllat bé el PRE. Aquesta és doncs una tasca que es podria dur a terme en pròxims experiments, intentar delimitar les dues activitats dins del Fab-6 però aquesta vegada fragmentant-lo en seqüències més petites i

basant-se en les dades actuals de motius consens de proteïnes d'insulator i de Polycomb, per tal d'intentar anticipar els possibles llocs d'unió i la funció de la seqüència. No obstant això, no pensem que hi hagi d'haver forçosament una clara delimitació entre totes dues funcions, les proteïnes d'insulator i les de Polycomb podrien unir-se a llocs molt junts (tan junts que no seria possible separar-los en constructes transgènics) en el Fab-6 i dur a terme igualment la seva funció.

Posteriorment a la publicació dels nostres resultats caracteritzant el Fab-6 (Pérez-Lluch et al., 2008) han aparegut diversos estudis que l'utilitzaven com a model d'estudi. Se n'ha descrit l'unió de CTCF (Smith et al., 2009); s'ha descrit que contacta amb el promotor d'Abdominal-B i amb els altres Fabs de manera CTCF-dependent (Kyrchanova et al., 2011); se n'ha descrit la unió d'Argonaute2 de manera RNAi-independent (Moshkovich et al., 2011); i se n'ha fet un estudi de deleccions utilitzant el mètode InSIRT (Iampietro et al., 2010). Aquesta tècnica (*In situ Integration for Repeated Targeting*) consisteix en la recombinació homòloga de tota una regió genòmica (en aquest cas l'iab-6 sencer, de 19 kb) i la seva substitució per un lloc d'integració del sistema ϕ C31. Un cop s'ha reemplaçat, es reintegra amb la recombinasa ϕ C31 la regió estreta però mutada *in vitro* als llocs d'interès (en el cas de l'estudi, amb deleccions parcials al llarg de l'iab-6, des del Fab-6 fins al Fab-7). Així es van generar deleccions tot al llarg de l'iab-6 i se'n van estudiar els fenotips.

Van generar una deleció que eliminava el Fab-6 (Fab-6¹) i una altra que eliminava el Fab-6 i l'iab6PRE (Fab-6²). La primera, que elimina el Fab-6, donava lloc a una barreja de transformacions homeòtiques de guany de funció (A5 a A6) i de pèrdua de funció (A5 a A4). En el primer cas es deu a que els elements activadors de l'iab-5 s'extenen a l'iab-6 a l'A5, i per tant té fenotip de A6. En canvi en el segon cas, la pèrdua de funció, el que passa és que com que el iabPRE-6 encara és funcional i ja no hi ha un insulator que en delimiti el radi d'acció, el silenciament de l'iab-6 a l'A5 es transmet a tot l'iab-5 i llavors agafa fenotip d'A4. En canvi la deleció que elimina tant el Fab-6 com el iabPRE6, tot i

que de fenotip similar, té més tendència a la transformació de guany de funció, ja que la pèrdua de funció implica el silenciament per part del PRE que en aquest cas ha estat eliminat. Una tercera deleció, *iab-6^{A7}*, que elimina només l'*iab6PRE* però no el Fab-6, no provoca cap transformació homeòtica aparent.

Aquestes observacions són coherents amb els nostres resultats. Es podria argumentar que, donat que hi ha un PRE al Fab-6, la mutació que elimina el Fab-6 no podria generar fenotip de pèrdua de funció a l'A5 ja que es perd el PRE que hi ha junt a l'insulator i amb ell l'efecte silenciador. Aquest no és el cas perquè hi ha l'*iab6PRE*, no afectat per la deleció, que continua silenciament i per això la pèrdua de funció. Llavors, com és que quan s'elimina l'*iab6PRE* no hi ha cap transformació? Doncs segons les nostres observacions és precisament perquè hi ha l'altre PRE al Fab-6 que s'continua fent l'efecte silenciador del domini a l'A6. És a dir, tot plegat suggereix un model en el que l'*iab-6* conté dos PREs de funció aparentment redundants, un d'ells més o menys a la meitat del domini (*iab6PRE*) i l'altre sobreposat o molt junt amb l'insulator que separa l'*iab-6* de l'*iab-5* (Fab-6). Donada l'extrema importància de la correcta expressió espacial d'Abd-B pel desenvolupament de l'organisme, no és incoherent la presència d'elements redundants que hagin estat mantinguts per selecció per assegurar que el domini només s'expressi als segments que li corresponen.

5.1.2 Transcripció als elements reguladors d'Abdominal-B

Els diferents dominis reguladors d'AbdB es transcriuen i ho fan de manera colinear, es detecten els seus trànscrips al segment en el que el domini controla l'expressió del gen, que estan disposats en el mateix ordre al genoma que els segments a l'organisme (Sánchez-Herrero and Akam, 1989). Aquests trànscrips sempre són en el mateix sentit de transcripció que AbdominalB. La funció i naturalesa d'aquests trànscrips continua essent una qüestió de debat. Alguns estudis els defineixen com a trànscrips discrets i separats els uns dels altres pels insulators (Bae et al., 2002; Drewell et al., 2002). La transcripció al domini estaria correlacionada amb un estat actiu d'aquest que li permetria

controlar l'expressió d'AbdominalB, o alternativament la transcripció seria una conseqüència de l'estat actiu del domini però sense rellevància per si mateixa. S'ha correlacionat l'estat actiu dels elements PRE/TRE (que són presents a cadascun dels dominis) amb el pas de la maquinària transcripcional pel seu damunt (Rank et al., 2002; Schmitt et al., 2005). Això podria ser una de les explicacions de la transcripció als dominis actius i no als dominis silencisats.

Un estudi recent suggereix que al segment A8 el trànscrit fa 92 kb, des de AbdB fins a abdominalA (també en el mateix sentit de transcripció que AbdB), i que per tant no s'interromp la transcripció als Fabs sinó que tota la regió reguladora es transcriu de manera contínua i es crea el trànscrit no codificant iab-8 (Gummalla et al., 2012). La funció d'aquest trànscrit és la de silenciar en *cis* l'expressió d'abdominalA mitjançant un mecanisme d'interferència transcripcional. Aquesta funció també s'havia trobat anteriorment pel trànscrit no codificant bxd que silencia en *cis* el gen Ultrabithorax, també al complex Bitòrax (Petruk et al., 2006; Mazo et al., 2007).

En qualsevol dels casos nosaltres vam voler comprovar que la transcripció també tenia lloc al Fab-6 i que es donava, com havia estat descrit pels demés Fabs, en el mateix sentit de transcripció que AbdominalB (sense). Això ho vam demostrar utilitzant primers específics per fer la retrotranscripció que només serveixen de motlle per una orientació. A larves i a embrions vam demostrar que hi havia trànscrit sense al Fab-6 però no anti-sense, és a dir que la maquinària transcripcional estava passant per sobre de l'element en la mateixa orientació que AbdB.

L'ús de cèl·lules en cultiu ens va servir per correlacionar la transcripció dels Fabs amb la d'AbdB. Les cèl·lules S2, on AbdB no s'expressa, no presentaven tampoc transcripció del Fab7, i en canvi les SG4 on AbdB sí que s'expressa mostraven transcripció al Fab7. Per tant és coherent amb la hipòtesi que la transcripció als dominis reguladors està correlacionada amb l'expressió d'AbdB, per bé que no podem concloure res sobre el mecanisme que les correlaciona.

Hem observat no obstant que la manca de CTCF i de JIL-1 redueix significativament la transcripció al Fab6 i Fab7. JIL-1 està relacionada amb transcripció activa, per tant és possible que senzillament el procés transcripcional s'hagi vist reduït perquè JIL-1 n'és necessària. Que la manca de CTCF també en redueixi la transcripció pot tenir diverses explicacions. Una seria que sigui una alteració indirecta ja que CTCF és important per la regulació de molts gens i que algun d'aquests que es veu afectat acabi alterant la transcripció als Fabs. Una altra possibilitat seria una afectació directa, en la que si l'insulator no funciona del tot bé degut a la manca de CTCF, tampoc no es transcriu correctament.

Que la transcripció de l'element insulator sigui important per la seva activitat pot ser degut a varis motius. Podria ser que el trànscrip fos important per la unió de l'insulator amb el promotor que controla o amb els altres insulators, de manera que si s'eliminés el trànscrip per exemple amb RNAi s'hauria de veure una disfunció de l'insulator en forma d'alteració de l'expressió d'AbdB al segment que correspongui a l'insulator i en conseqüència una transformació homeòtica. Però també podria ser que el trànscrip no sigui important en sí, sinó que l'important sigui el procés transcripcional sobre l'insulator, possiblement per mantenir-ne la cromatina en un estat descondensat que faciliti la unió de proteïnes d'insulator. Finalment també cabria la possibilitat que la transcripció als dominis actius no sigui més que una conseqüència colateral i inòcua d'un estat descondensat de la cromatina necessari pel funcionament dels elements presents al domini.

5.1.3 JIL-1 als insulators

Hem observat que la manca de JIL-1 en mosques que contenen una construcció transgènica amb un enhancer i un gen mini-white separats per un insulator provoca que l'insulator funcioni pitjor, i això és observable per un increment de la pigmentació de l'ull. El motiu pel qual JIL-1 pot afectar la funció insulator no és clar.

JIL-1 ha estat tradicionalment associat a la cromatina activa donada la seva presència a interbandes de cromosomes politènics, és necessària per mantenir l'estructura de la cromatina i és la quinasa responsable de la fosforilació de la serina 10 de la histona H3 a interfase (Jin et al., 1999, 2000; Wang et al., 2001). La contribució de JIL-1 a la regulació de l'expressió gènica no està clara. S'ha proposat que és necessària per l'alliberament de la polimerasa pausada i la represa de l'elongació (Ivaldi et al., 2007); un altre estudi proposa que la presència de JIL-1 és tot al llarg del gen i que no és proporcional als nivells de transcripció ni a la densitat de polimerasa (Regnard et al., 2011), en canvi observen una correlació entre la presència de JIL-1 i la trimetilació de la lisina 36 de la H3, marca que s'associa a elongació transcripcional.

A més de la serina 10, també pot fosforilar la serina 28 de la H3, i és necessari per l'acetilació de la lisina 27 per part de la CREB binding protein, una marca característica dels enhancers (Kellner et al., 2012). Així, s'ha detectat la presència de JIL-1, H3S10ph, H3S28ph i H3K27ac a molts gens de *Drosophila*, localitzant tant als promotors com als enhancers, obrint la possibilitat que JIL-1 sigui requerit per establir el contacte entre enhancer i promotor, possiblement a través de la unió de les proteïnes 14-3-3 que reconeixen aquestes modificacions de la H3. A més, la presència de JIL-1 i aquestes tres modificacions estan correlacionats positivament amb l'expressió gènica. Aquests contactes enhancer-promotor dependents de JIL-1 han estat verificats per Chromosome Conformation Capture (Kellner et al., 2012).

Un altre estudi proposa que la funció de JIL-1 consisteix en mantenir els límits de les regions eucromàtiques i impedir l'expansió de l'heterocromatina mediada per Su(var)3-9 (Ebert et al., 2004; Zhang et al., 2006). A més hi ha evidències que JIL-1 interacciona amb un component estructural de la membrana interna nuclear com és la Lamina Dm0 (Bao et al., 2005).

Totes aquestes evidències —formació de contactes entre promotor i enhancer; barrera contra la heterocromatina; interacció amb elements

estructurals del nucli— són elements que posen a JIL-1 en comú amb els insulators, són per tant coherents amb la participació de JIL-1 a l'activitat insulator i poden donar pistes sobre com hi pot estar contribuint. És possible que JIL-1, a través de la fosforilació de la serina 10 de la H3, no només tingui un rol durant el procés transcripcional a l'elongació sinó que jugui un paper en el manteniment i formació de super-estructures de la cromatina i així la seva manca pugui alterar les interaccions entre elements gènics distants, tal vegada resultant en la incapacitat dels insulators de formar les estructures de cromatina adients que aïllin regions gèniques. L'altra possibilitat és que JIL-1 actuï només promovent la formació d'un estat descondensat de la cromatina i sigui per aquest motiu que quan falta els insulators funcionen pitjor.

Una conseqüència que hauria de tenir la falta de JIL-1 si és requerida per a la formació de contactes entre insulators i altres elements és al locus d'Abd-B, ja que els diferents Fab no podrien contactar entre ells o amb el promotor del gen. Efectivament s'ha demostrat que els mutants de JIL-1 exhibeixen una transformació homeòtica A6 a A5, que també s'observa en delecions de l'insulator del Fab-7 (Zhang et al., 2003).

Per tal d'escatir el mode de funcionament de JIL-1 serà necessari mapar bé al genoma els llocs de fosforilació de la serina 10. Nosaltres hem ho hem fet a nivell de Complex Bitòrax i hem vist que hi ha fosforilació als insulators, malgrat que vam haver de repetir diverses vegades l'experiment de CHIP i la majoria d'elles no aconseguíem una senyal específica. Sí que disposem però del mapa dels llocs que ocupa JIL-1 al genoma, publicats a ModEncode, i hem vist com no només coincideix amb insulators del Complex Bitòrax sinó que coincideix amb un gran nombre de llocs on hi ha proteïnes d'insulator. Això sumat a l'efecte enhancer-blocking que havíem vist amb els constructes posen de manifest contundentment que d'alguna manera JIL-1 participa de l'activitat insulator. En un futur es podrien fer experiments 3C en mutants de JIL-1 per veure si efectivament és necessari per a que es produeixi el contacte entre insulators i altres elements.

5.2 Noves proteïnes d'insulator

5.2.1 El complex de CP190

Amb l'objectiu d'aprofundir en el coneixement del funcionament i característiques dels insulators de *Drosophila*, vam dissenyar un seguit d'experiments que conduïssin a identificar nous factors necessaris per a dur a terme l'activitat insulator. L'estratègia partia d'aïllar una proteïna coneguda d'insulators i analitzar amb quines proteïnes interacciona formant un complex proteic. Com a proteïna *bait* (aquella que s'utilitza per immunoprecipitar el complex i identificar-ne els components) vam utilitzar CP190. El motiu principal per escollir-la és que se'n coneixia la presència a tots els tipus d'insulator caracteritzats fins aquell moment: Su(Hw), BEAF i CTCF (Bushey et al., 2009), i això feia pensar que si hi havia algun altre tipus d'insulator no conegut fins al moment la que tenia més possibilitats d'interaccionar-hi era CP190.

Per a purificar CP190 vam decidir expressar-la en un cultiu de cèl·lules S2, que en permetia l'expressió per inducció (evitant possibles efectes adversos de la sobreexpressió constitutiva), la obtenció de grans quantitats de proteïna gràcies al creixement exponencial de les cèl·lules i la facilitat d'obtenció de l'extracte nuclear degut a la falta de teixits extracel·lulars i cutícules. El tag al que vam fusionar CP190 va ser el TAP tag (Puig et al., 2001), utilitzat habitualment per fer doble purificació (*tandem purification*) però amb el que nosaltres vam fer purificació d'un sol pas. El fet que ja s'hagués utilitzat al laboratori amb resultats satisfactoris va fer que ens hi decantéssim per damunt d'altres tags de característiques similars. En el nostre cas també va donar bons resultats si ens atenim a la proporció de proteïnes identificades que hem demostrat posteriorment que eren veritables interactors específics de CP190 en relació a les que semblaven ser interaccions inespecífiques o contaminacions, minoritàries i de poca robustesa als anàlisis d'espectrometria de masses.

Entre les proteïnes que esperavem trobar a la purificació hi ha aquelles que ja havia estat publicada la seva interacció amb CP190, com Map60 (Kellogg et al., 1995; Oegema et al., 1997) i les proteïnes d'insulator Su(Hw), Mod(mdg4) (Pai et al., 2004; Gerasimova et al., 2007) i CTCF (Mohan et al., 2007). La interacció de CP190 amb la proteïna d'insulator BEAF no havia estat mai demostrada prèviament, sinó que l'evidència en que es basen els estudis publicats per dir que CP190 i BEAF col·laboren en un mateix tipus d'insulator són evidències de colocalització ja que coincideixen a molts llocs en el genoma tal com s'ha vist per CHIP on CHIP (Bushey et al., 2009). A diferència del que passa amb CTCF, Su(Hw), Mod(mdg4) i CP190, no s'ha publicat mai que BEAF formi part dels insulator bodies a on s'aglomeren les demés proteïnes d'insulator a nuclis diploïdes. A més la distribució de BEAF al genoma es diferencia de les demés en tant que la seva presència a promotors és força més alta (Jiang et al., 2009; Yang et al., 2012). La seva activitat enhancer-blocking no obstant és clara com es va demostrar als insulators scs (Hart et al., 1997).

Podria ser doncs que BEAF sigui una proteïna amb una activitat insulator particular, molt esbiaixada cap a la regulació de l'expressió gènica a nivell de promotor i no tant a nivell d'elements reguladors situats en regions intergèniques. Així, s'ha relacionat BEAF amb gens actius transcripcionalment tant en CHIP com per la localització a interbandes a cromosomes politènics (Zhao et al., 1995; Jiang et al., 2009). Es correlaciona amb la presència d'H3.3 i de H3K4me3, marques associades a l'activitat transcripcional, i s'ha proposat que és necessària per impedir la deposició de metilació a la H3K9, modificació associada al silenciament (Emberly et al., 2008), a través de la qual activaria l'expressió de diversos gens del cicle cel·lular antagonitzant la regulació del factor transcripcional DREF (Hart et al., 1999). S'ha detectat que la meitat dels llocs en que BEAF es troba a prop d'un promotor són llocs head-to-head (dos gens pròxims i de direcció oposada), suggerint que BEAF es requeriria per mantenir alts nivells d'expressió al crear l'entorn necessari perquè cada promotor no interfereixi amb un promotor adjacent. La manera com col·laboraria amb CP190

no està clara, però el cert és que CP190 també té una forta presència a promotors de gens actius (Bartkuhn et al., 2009), a molts dels quals hi coincideix amb BEAF.

Vam detectar interacció de CP190 amb Nurf-38, una pirofosfatasa inorgànica que forma part del complex remodelador de la cromatina NURF (Gdula et al., 1998). Els insulators necessiten una estructura descondensada de la cromatina per funcionar i per tant necessiten un complex remodelador de la cromatina que reestructuri els nucleosomes. Un estudi en constructes transgènics a cèl·lules i embrions va demostrar que l'efecte enhancer-blocking dels insulators del complex Bitòrax és dependent de l'acció de NURF sobre la cromatina, que a més ISWI —el component ATPasa del complex— es troba present als insulators endògens del complex Bitòrax (Li et al., 2010). Si s'elimina NURF s'observa un increment de la densitat de nucleosomes als insulators, confirmant que NURF és necessari per mantenir un estat descondensat dels insulators. Dins del marc d'aquestes observacions interpretem que el fet que un component de NURF formi complex amb CP190 suggereix que com a mínim una part substancial dels insulators en els que participa CP190 —que són la majoria dels descrits fins ara— necessiten la remodelació de cromatina ATP-dependent de NURF i reenfoca la idea que els insulators són actius en estats descondensats de cromatina.

Z4 (també anomenada putzig o pzg en algunes publicacions) és una proteïna amb zinc fingers que s'uneix a interbandes a cromosomes politènics i que és requerida per mantenir-ne l'estructura. El seu rol concret no és clar, però s'ha vist que recluta JIL-1 a la cromatina (Gan et al., 2011), que interacciona amb NURF (Kugler and Nagel, 2010) o amb DREF (Hochheimer et al., 2002). Un interactor habitual de Z4 amb qui colocalitzen totalment a politènics és Chromator (o Chriz en algunes publicacions), de qui s'ha publicat que està enriquit a llocs d'unió de CTCF (Celniker et al., 2009) i que interacciona amb una altra proteïna amb funció insulator, L(3)Mbt (Richter et al., 2011), i que a més el complex Z4-Chromator cooperen amb BEAF (Gan et al., 2011).

Nosaltres hem detectat Z4 entre les proteïnes que interaccionen amb CP190. A més en algunes línies hem vist que Z4 és necessària per l'enhancer-blocking en assaigs transgènics. La colocalització que mostren a politènics és parcial. Hem utilitzat unes mosques que no expressen Chromator que sabíem que es necessària per la unió de Z4 a la cromatina (Gan et al., 2011) per comprovar que ni Z4 ni Chromator són requerides per la unió de CP190. Tot plegat suggereix que Z4 està interaccionant amb CP190 a alguns llocs i que d'alguna manera la regulació de l'estructura de la cromatina que se li atribueix, contribuint a generar regions descondensades, és necessària per la funció insulator. De la mateixa manera que JIL-1, proteïna associada a cromatina descondensada, i NURF (totes dues interactores demostrades de Z4) que promou el moviment de nucleosomes per descondensar-la, suggereixen un model en que els insulators —potser per la necessitat de formar superestructures de cromatina per posar en contacte regions distants— requereixen baixa densitat de nucleosomes, i per això es troba CP190 associat a aquestes proteïnes que són les encarregades de mantenir aquesta baixa densitat.

5.2.2 Caracterització d'Ibf-1 i Ibf-2

5.2.2.1 Conservació i dominis

Entre les proteïnes que hem identificat formant complex amb CP190 s'hi trobaven dos gens no caracteritzats fins al moment, CG8436 i CG9740 (en la nomenclatura que s'utilitza pels gens sistemàticament anotats a partir de la seqüenciació del genoma de *Drosophila melanogaster*) que nosaltres hem anomenat Ibf-1 i Ibf-2 respectivament (per Insulator Binding Factor 1 i Insulator Binding Factor 2). Quan les hem co-immunoprecipitat hem comprovat que no només formen complex cadascuna d'elles amb CP190 sinó també entre elles, per tant hi ha una interacció de totes tres.

No són però proteïnes semblants a CP190, d'entrada són força més petites: CP190 fa 121 KDa, Ibf-1 27.7 KDa i Ibf-2 21.9 KDa. CP190 té dominis BTB

i zinc finger C₂H₂ i lbf-1 i lbf-2 no. En canvi lbf-1 i lbf-2 sí que són semblants entre elles, amb un 18% de similaritat de seqüència polipeptídica essent cadascuna la proteïna més semblant a l'altra dins del genoma de *Drosophila melanogaster*. Els dos gens es troben l'un al costat de l'altre al cromosoma 3 i es transcriuen en direccions oposades, els seus TSS estan separats per 470 pb, obrint la possibilitat de que estiguin co-regulats. El fet que els dos gens no només tinguin seqüències similars sinó que a més siguin adjacents al genoma suggereix que poden provenir d'una duplicació d'un sol gen ancestral.

Quan un gen es duplica pot donar lloc a diferents situacions (Prince and Pickett, 2002). En primer lloc, si el nou gen ni és positiu ni negatiu per l'organisme i és una duplicació recent a escala evolutiva, pot donar-se redundància: tots dos gens fan la mateixa funció. De vegades la redundància pot ser beneficiosa i quedar fixada per la selecció positiva, sobretot en casos de gens que s'han d'expressar en gran intensitat. El segon cas és quan aquesta redundància no aporta res a l'organisme, el nou gen pot degenerar a pseudogen ja que la selecció només actuarà sobre un dels dos. El tercer cas és la neofuncionalització: el nou gen duplicat adquireix una nova funció. Un quart cas és la subfuncionalització. En aquesta situació, cadascun dels dos gens originats adquireix mutacions deletèries que afecten a sub-funcions diferents, i tots dos gens es complementen: partir d'aquí els dos gens seran seleccionats positivament. Aquestes sub-funcions tant poden ser dominis independents dins de la seqüència del gen com elements reguladors com enhancers. Per exemple mutacions deletèries en enhancers d'un i altre gen que fan que finalment els dos duplicats responguin a enhancers diferents i la seva expressió en el temps o l'espai sigui complementària. lbf-1 i lbf-2 semblen són funcionals tots dos, i no és probable que siguin redundants tenint en compte que lbf-2 és necessari per a que lbf-1 s'uneixi a la cromatina. Una neofuncionalització és possible, però tenint en compte que totes les dades de localització i interaccions són gairebé idèntiques no sembla que fagin funcions diferents. El més plausible sembla una subfuncionalització en la que els dos gens duen a terme la mateixa funció però

alguna diferència a nivell de seqüència o de regulació fa que tots dos es complementin i es necessitin.

Ibf-1 i Ibf-2 són proteïnes conservades, al menys, entre la majoria d'espècies de *Drosophila*. El nostre anàlisi no ha estat exhaustiu i ha estat fet amb mètodes de cerca bàsics, per tant no descartem que hi hagi homòlegs més allunyats i amb no tanta similaritat de seqüència que no haguem identificat. De la mateixa manera, no hem cercat a grups ancestrals als drosòfílids un possible gen ancestral a Ibf-1 i Ibf-2. El que sí que podem dir és que totes les espècies del subgènere *Sophophora* tenen homòlegs de tots dos gens de manera que la duplicació ha de ser com a mínim anterior a la formació d'aquest grup. Inclús en una espècie, *Drosophila ananassae*, Ibf-2 ha patit una duplicació.

El fet que no haguem trobat homòleg d'aquestes dues proteïnes fora de *Drosophila* no és, en el marc de les proteïnes d'insulator, extrany. Només CTCF té homòlegs a vertebrats. De fet té homòlegs a gairebé tots els llinatges de bilaterals, i se n'ha proposat una posició fonamental en la seva evolució des de l'explosió càmbrica a partir del seu rol en la organització dels clusters de gens homeòtics (Heger et al., 2012). CP190, Su(Hw) i Mod(mdg4) no tenen homòlegs més enllà dels insectes, i BEAF només té homòlegs en el gènere *Drosophila* (Schoborg and Labrador, 2010). A BEAF se li ha suggerit un rol important i específic en la regulació del genoma de *Drosophila* degut particularitats d'aquest com l'alta densitat gènica i l'alta presència de promotors compartits en gens disposats head-to-head (Yang et al., 2012). A falta doncs d'anàlisis més elaborats sobre la distribució i orígens filogenètics d'Ibf-1 i Ibf-2, no podem descartar que siguin proteïnes que, com BEAF, tenen un rol específic en la delimitació de regions reguladores al genoma de *Drosophila*.

De fet hi ha un altre aspecte que relaciona Ibf-1 i Ibf-2 amb BEAF, que és la presència d'un domini d'unió al DNA BED finger a la seva seqüència (extrem N-terminal). Els dominis BED van ser identificats a partir de l'anàlisi de seqüència de BEAF i del factor de transcripció DREF (Aravind, 2000), tots dos compartint un

mateix motiu d'unió al DNA pel que competeixen (Hart et al., 1999). No són però les úniques proteïnes de *Drosophila* amb aquest motiu (Taniue et al., 2010).

Diverses proteïnes de recent descripció a vertebrats contenen aquest motiu (Andersson et al., 2010; Mokhonov et al., 2012) entre elles un homòleg de DREF a vertebrats (Ohshima et al., 2003) i de fet hi ha proteïnes amb BED a tots els eucariotes, des de plantes a fongs (Aravind, 2000). Es tracta majoritàriament de proteïnes d'unió al DNA —transposases i factors de transcripció. S'ha suggerit que aquests dominis tenen un origen evolutiu en les transposases que aïllaven els transposons dels efectes de l'entorn cromosomal, de manera que l'organisme hoste va acabar reclutant aquests dominis en proteïnes que duguessin a terme funcions cel·lulars com aïllar regions reguladores.

El fet que Ibf-1 i Ibf-2 —i tots els seus homòlegs— no només continguin un domini proteic d'unió al DNA, sinó que a més aquest sigui un domini relacionat evolutivament amb l'aïllament de regions cromosòmiques ens porta a pensar que el complex que formen amb CP190 ha d'estar relacionat amb els insulators i el reconeixement d'algun tipus de seqüència. Cal dir que la regió C-terminal de Ibf-1 i Ibf-2 també està ben conservada, tant entre elles dues com amb cadascun dels seus homòlegs, suggerint que conté algun tipus de domini important per la proteïna. Amb les eines de cerca de que disposem no hem estat capaços d'identificar-ne un domini amb robustesa, caldria doncs explorar-ho millor perquè podria donar més pistes sobre el seu mode de funcionament.

5.2.2.2 Localització cel·lular d'Ibf1 i Ibf-2

Les immunotincions de cèl·lules S2 mostren que Ibf-1 i Ibf-2 es troben al nucli en aquesta línia cel·lular. També són nuclears les immunotincions amb Ibf-2 a tots els estadis del desenvolupament embrionari. El cas d'Ibf-1 no és clar ja que la senyal als embrions no mostra el patró nuclear marcat pel DAPI sinó que sembla estar disseminat per tot l'embrió. Això es pot deure a dues coses: o bé el patró de localització és nuclear i citoplasmàtic a la vegada, o bé l'anticòs s'uneix inespecíficament a proteïnes del citoplasma i llavors la senyal del citoplasma és

un artefacte. Dos motius donen pes a aquesta última opció: el fet que a cèl·lules S2 (que és una línia d'origen embrionari) el patró és nuclear i no citoplasmàtic, i que a totes les immunotincions i Western Blots que hem realitzat sol tractar-se d'un anticòs que, fora d'una estreta finestra de condicions, genera un cert soroll inespecífic.

Quant a les mitosis no podem concloure res amb robustesa. En la majoria de tincions mostren un patró difuminat sobre tota l'àrea que ocupen els cromosomes mitòtics, com si al trencar-se l'embolcall nuclear difonguessin. Ara bé, en algunes tincions es veu una imatge més precisa de senyal al llarg del cromosoma mitòtic amb particular intensitat a la zona centromèrica. De ser així, el rol que podrien tenir als centròmers mitòtics ens és desconegut i no hi tenim més pistes al respecte. En alguns casos es veu a més a més colocalitzant amb CP190 als centrosomes. La funció de CP190 als centrosomes és encara desconeguda, se sap però que no és essencial per les divisions cel·lulars (Butcher et al., 2004), i no es té coneixement que cap de les altres proteïnes d'insulator tingui el mateix comportament. Per tant, en cas que sigui una localització verídica i no un artefacte degut a una inespecificitat de l'anticòs, s'obren dues possibilitats. Una és que tinguin algun paper en la formació, manteniment o funció centrosomal; l'altre és que localitzin als centrosomes degut a la seva afinitat per CP190 a la que d'alguna manera segueixin, però que no hi tinguin cap funció en particular.

A nuclis de cèl·lules de cervell tenyits tant amb Ibf-1 com amb Ibf-2 observem la seva presència a insulator bodies, a on colocalitza també CP190. Els insulator bodies havien estat descrits prèviament com a agregats de proteïnes d'insulator a llocs del nucli on es formen loops de les fibres de cromatina que se suposa que separen dominis independents d'expressió gènica (Byrd and Corces, 2003; Gerasimova et al., 2007). S'ha trobat que també CTCF, Su(Hw) i Mod(mdg4) colocalitzen amb CP190 en aquests punts (Pai et al., 2004; Gerasimova et al., 2007). Per tant nosaltres ho interpretem com una evidència de la participació d'Ibf-1 i Ibf-2 en conjunció amb CP190 als insulators.

Les immunotincions de cromosomes politènics de nuclis de glàndula salival mostren que tant lbf-1 com lbf-2 s'uneixen a la cromatina. S'hi uneixen coincidint en gran mesura amb la senyal de CP190, coherentment amb el fet que formen un complex plegats. Ara bé, el nombre de bandes de CP190 és molt major que el de lbf-1 o lbf-2, de manera que molts dels llocs de CP190 no són llocs a on formi complex amb lbf-1 o lbf-2. Inversament, la gran majoria de llocs lbf-1 o lbf-2 estan ocupats també per CP190, malgrat que hi ha una petita part en la que no és així. No podem treure conclusions sobre la naturalesa d'aquests llocs degut a la dificultat d'identificar els loci a cromosomes politènics. Tot plegat ens confirma no obstant que hi ha una extensa colocalització de totes tres proteïnes a politènics.

Els cromosomes politènics de larves mutants de CP190 no mostren que hi hagi una baixada significativa de la unió de lbf-1 o lbf-2 a la cromatina, com si la unió d'aquestes en fos independent. Als mutants de lbf-2 no hi veiem cap canvi a gran escala de la unió de CP190, tot i que com que la majoria de llocs de CP190 no són llocs on es formés el complex amb lbf-2 al wild-type, podria ser que hi hagués hagut una lleugera disminució dels llocs de CP190 (aquells a on hi hauria lbf-2) i que a gran escala fos imperceptible. La naturalesa de la tècnica fa que les tincions a politènics no siguin una bona estratègia per detectar canvis parcials o minoritaris d'unió d'una proteïna a la cromatina. Per a estar-ne segurs s'hauria d'adoptar una altra estratègia, com immunoprecipitació de cromatina en mutants seguida de PCR a llocs d'interès o seguida de seqüenciació total.

Els politènics mutants de lbf-2 sí que revelen una davallada completa de la unió de lbf-1 a la cromatina. Una primera possibilitat seria que una mutació d'lbf-2 fa baixar els nivells d'lbf-1 a la cèl·lula, ja sigui a nivell transcripcional com a nivell d'estabilitat de la proteïna. Aquesta possibilitat va quedar descartada tant pels alts nivells de transcrit d'lbf-1 detectats al microarray com per Western Blots (no mostrat) que corrobora que en mutants d'lbf-2 no hi baixa el nivell de lbf-1. Així, s'obre la possibilitat que lbf-2 sigui qui s'uneixi primàriament a la cromatina i lbf-1 ho fagi a través d'lbf-2. Alternativament lbf-1 i lbf-2 es

necessiten mútuament per unir-se a la cromatina en interdependència i actuarien com a dímer a l'hora d'unir-s'hi. La millor manera de saber-ho seria fent immunotincions d'Ibf-2 en politènics mutants de Ibf-1, però malauradament el mutant $\Delta RB1$ és letal en homozigosi i no hi ha larves per comprovar-ho.

5.2.2.3 Canvis de l'expressió gènica

Del fet que els insulators són importants per la correcta disposició de moltes regions reguladores al genoma se'n deriva que la falta de proteïnes d'insulator pot causar alteracions en l'expressió gènica. Com que no estan funcionalment lligats ni a l'activació ni a la repressió sinó més aviat a la separació entre dominis diferencialment expressats, ens podem esperar que aquestes alteracions siguin en tots dos sentits: gens que augmenten l'expressió (upregulats) i gens que la redueixen (downregulats).

Dels resultats que obtenim amb els Microarrays fets amb RNA de cultius cel·lulars S2 tractats amb RNAi de Ibf-1, Ibf-2 i CP190 observem que aquest és efectivament el cas. El nombre de gens diferencialment expressats varia en cada cas i és subjecte principalment a dues fonts de variabilitat. Per un cantó hi ha els efectes intrínsecs de l'eliminació de la proteïna. Per altre hi ha l'eficàcia en l'eliminació d'aquesta amb l'RNAi. Teòricament els experiments han estat fets de manera que tota la resta de fonts de variabilitat siguin iguals per a totes tres condicions. Però la segona font, l'eficàcia de l'RNAi, no és quelcom que es pugui controlar completament ja que depèn de característiques del fragment de dsRNA utilitzat, de la seva integració a les cèl·lules i de la seva captació per la maquinària RNAi.

D'aquesta manera el knock-down de Ibf-1 és el que menys efecte ha tingut sobre el seu transcrit. Molt probablement en conseqüència a això és també la condició que menys gens diferencialment expressats causa. La xifra de gens diferencialment expressats és de totes maneres relativa al lliniar de robustesa que s'esculli, i aquesta robustesa estadística ha quedat limitada des del moment en que només estem tenint en compte dos rèpliques de

l'experiment, ja que la tercera la vam haver d'eliminar perquè per algun factor experimental no controlat va comportar-se de manera massa diferent a les altres dues.

Ibf-2 i CP190 van tenir un knock-down efectiu que va reduir els nivells de transcrit més de quatre vegades. CP190 va alterar l'expressió de molts més gens que no pas Ibf-2. Això és un resultat que ens podíem esperar atès que CP190 es troba a molts més llocs del genoma com hem vist a politènics i per tant la seva manca es fa notar a molts més gens.

Cal subratllar el fet que en el knockdown d'Ibf-1 augmenta la transcripció d'Ibf-2, i en el de Ibf-2 augmenta la transcripció d'Ibf-1. Podria extreure-se'n doncs que, de normal, actuen com a reguladors negatius l'un de l'altre. Sabem que si no hi ha Ibf-2 no hi ha unió d'Ibf-1 a la cromatina, per tant en aquest cas Ibf-1 també seria un regulador negatiu del seu propi gen. En la hipòtesi que els dos gens tinguessin funcions redundants, l'augment d'expressió d'un d'ells a falta de l'altre podria ser entès com a un mecanisme compensatori. No obstant aquesta opció no és la més probable, d'entrada perquè si Ibf-2 és necessari per la unió a cromatina d'Ibf-1, no té cap efecte compensatori l'augment de la transcripció d'aquest.

Una altra perspectiva d'analitzar aquests augments de transcripció és no interpretant Ibf-1 i Ibf-2 com a reguladors negatius o positius sinó simplement com a organitzadors de l'estructura de la cromatina a regions reguladores. Convé recordar que aquests dos gens estan en disposició head-to-head: comparteixen la mateixa regió a 5' del gen, a on s'acostumen a trobar els promotors proximals o altres seqüències reguladores. D'aquesta manera, en la hipòtesi que aquestes dues proteïnes actuen com a dímer a la cromatina, el que estaria passant és que a l'eliminar-les es produeix alguna malfunció en la seva regulació que faria que s'expressin més del normal. En un plantejament hipotètic, si hi hagués algun enhancer a l'entorn dels dos gens i un insulator a 5' dels gens que n'impedeix l'activació inapropiada, al faltar una de les dues

proteïnes Ibf-1 o Ibf-2 l'insulator deixaria de funcionar i els promotors dels gens respondrien a l'activació per part de l'enhancer.

Una predicció que es podria fer del fet que CP190, Ibf-1 i Ibf-2 formin un complex és que els gens que es vegin afectats en tots tres casos ho seran en la mateixa direcció. És a dir —al marge dels gens que estan afectats per CP190 i no per Ibf-1 o Ibf-2— aquells que estan afectats pel complex estan co-regulats ja que les tres proteïnes actuen en conjunció per dur a terme una mateixa funció. Hem comprovat i verificat aquesta hipòtesi calculant una correlació d'Spearman per a cada parella de microarrays, essent 0 gens de correlació i 1 màxima correlació. Ara bé, cal tenir en compte que les condicions experimentals generen una variabilitat que fa que aquesta correlació màxima teòrica d'1 no sigui la real donades les nostres dades. Es van generar simulacions amb els paràmetres estadístics dels arrays reals per tenir una estimació d'a quin interval de valors d'Spearman correspon la màxima correlació, sota l'assumpció de coregulació total, i la mínima correlació, sota l'assumpció que no hi ha co-regulació en absolut. D'aquesta manera es va concloure que en totes les comparacions (Ibf-1 amb Ibf2; Ibf-1 amb CP190; Ibf-2 amb CP190) hi havia co-regulació significativa. Concloem doncs que Ibf-1, Ibf-1 i CP190 cooperen funcionalment en la regulació gènica d'un conjunt de gens en la mateixa direcció.

Els resultats de l'anàlisi de clustering als cromosomes tenia l'objectiu de saber si els gens alterats es trobaven a prop entre ells al genoma agrupats en clusters funcionals. El resultat va ser negatiu, volent dir que les proteïnes el complex afecten a gens tot al llarg del genoma independentment de si formen part d'un cluster, apropant-les més al seu rol mecanístic i d'organització cromosomal que no pas a un rol funcional en la regulació d'un tipus de gens particular. Coherentment amb aquesta observació, l'anàlisi de Gene Ontology va revelar una gran diversitat de grups de gens, no mostrant cap grup de gens particularment més enriquit que la resta.

5.2.2.4 Localització al genoma d'Ibf-1, Ibf-2 i CP190

Les tincions de cromosomes politènics ja ens havien donat alguna informació a gran escala de la distribució de les tres proteïnes del complex al genoma i del solapament entre elles, però la resolució que ofereixen els politènics és limitada i per conèixer al detall la localització al genoma vam utilitzar CHIP-Seq. El CHIP-Seq és una tècnica que ha esdevingut en els últims temps d'ampli ús en el camp de l'estudi de la cromatina, degut principalment a l'avanç en les tecnologies de seqüenciació. Té algunes aventatges sobre el CHIP-on-CHIP, la informació que s'obté és completa i abarca tot el genoma ja que es seqüencia tot, enlloc de dependre de l'abast de sondes d'un microarray. En conseqüència, l'amplitud dels pics d'enriquiment s'ajusta més a la localització real de la proteïna sobre la cromatina, són pics més discrets i la resolució augmenta.

La qualitat de les seqüències que vam obtenir va ser bona, l'amplitud mitjana dels pics en els tres experiments es va mantenir força constant i discreta. Observem que, coherentment amb les observacions a politènics, la proteïna que presenta més unions a la cromatina és CP190, i Ibf-2 en presenta més que Ibf-1. Aquesta diferència entre Ibf-1 i Ibf-2 no descartem no obstant que sigui deguda a la ja comentada menor especificitat de l'anticòs d'Ibf-1, que com en la resta d'experiments realitzats ha funcionat pitjor que el d'Ibf-2. En una primera visió general dels pics sobre el genoma ja s'observen un seguit de trets que han hagut de ser examinats amb més detall. Per exemple, el posicionament de les proteïnes del complex a promotors head-to-head, o la diversitat de coincidències amb les demés proteïnes d'insulator. Un tret que s'observa amb freqüència són els pics desdoblats, característics d'Ibf-1 i Ibf-2. Podria ser un artefacte de la tècnica o podria ser que hi ha llocs d'unió molt propers que semblen un sol interval amb pic desdolat. El motiu pel qual s'observen amb freqüència s'haurà d'estudiar millor analitzant la naturalesa de la unió d'Ibf-1 i Ibf-2 al DNA, fins a quin punt es necessiten l'una a l'altra i si s'uneixen totes dues al DNA o només una i l'altra a aquesta.

CP190 és menys intergènica del que es deia al CHIP-on-ChIP de Bushey et al. (Bushey et al., 2009) però més que no pas el que es deia al de Bartkuhn et al. (Bartkuhn et al., 2009). És efectivament una proteïna amb alta presència a regions intergèniques, gairebé la meitat de tots els seus llocs. L'altra meitat es reparteix sobretot en voltants del TSS i interior dels gens. La fracció més minoritària es troba al voltant de l'extrem 3' dels gens.

En comparació, Ibf-1 i Ibf-2 —amb una localització gairebé idèntica entre elles— tenen molta més tendència a les regions intergèniques, al voltant del 60% de llocs. Això indicaria una tendència més pronunciada a unir-se a elements reguladors distals en detriment dels promotors i elements reguladors proximals. De la fracció gènica, destaca en primer lloc el posicionament dins dels gens, observació que podríem lligar a un possible rol en promotors interns o promotors alternatius, sense obviar una forta presència també al voltant dels TSS i una minoritària als extrems 3'.

Aquest posicionament lleugerament diferent en regions gèniques de CP190 respecte Ibf-1 o Ibf-2 és també el que interpretem del gràfic sobre un gen estandaritzat, en el que CP190 té un pic més alt a 5', i per contra la corba Ibf-1 i Ibf-2 supera a la de CP190 dins del gen. Tot plegat suggereix algun rol d'Ibf-1 i Ibf-2 a l'interior dels gens. La significativa correlació positiva de la presència de les proteïnes del complex a gens que contenen promotors alternatius suggereix que aquest rol podria tenir a veure amb la regulació de l'expressió d'un mateix gen per diferents promotors. Aquesta regulació diferencial, que té una dimensió temporal i també espacial dins l'organisme, duu a l'expressió de transcrits diferents i és de vital importància pel desenvolupament (Schibler and Sierra, 1987; Ayoubi and Van De Ven, 1996) i és fins i tot més freqüent que l'splicing alternatiu (Pal et al., 2011). La manera com Ibf-1 i Ibf-2 podrien regular l'expressió de gens amb promotors alternatius és quelcom que s'hauria d'investigar més a fons. Tenint en compte que es tracta de proteïnes d'insulator, és plausible un escenari en el que determinessin l'activitat d'un o altre promotor

a través de loops de cromatina que connectessin l'enhancer amb el promotor adequat a cada moment del desenvolupament.

Ibf-1 i Ibf-2 colocalitzen majoritàriament amb CP190, amb qui formen complex. No obstant això, també colocalitzen amb les altres proteïnes d'insulator, encara que menys. D'entre elles, amb qui més coincidències mostren és amb CTCF, que com Ibf-1 i Ibf-2 mostren un patró de localització repartit entre l'intergènic i el gènic, amb una mica més de la meitat a intergènic (Bushey et al., 2009). De totes les combinatòries possibles entre proteïnes d'insulator descrites (Schwartz et al., 2012), amb la que més freqüentment coincideixen és amb la combinació CP190 + CTCF, seguida de CP190 + CTCF + Su(Hw) + Mod(mdg4). En tercer lloc i vindria la situació en que només hi són elles i CP190, i després la combinació CP190 + BEAF. El solapament amb CTCF no és, malgrat tot, prou extens com per considerar que Ibf-1 i Ibf-2 formarien part d'un sub-grup d'insulators CTCF. Així, considerem que Ibf-1 i Ibf-2 formarien un nou sub-grup d'insulators amb la seva pròpia xarxa de llocs d'unió en conjunció amb CP190. Degut a que ni són insulators de naturalesa intergènica com els de Su(Hw) ni tampoc tan intensament gènica com els de BEAF, coincideixen amb un major grau amb la xarxa d'insulators de CTCF. Les interaccions físiques i funcionals de Ibf-1 i Ibf-2 amb les demés proteïnes d'insulator resten, de moment, per explorar.

Entre els motius de DNA enriquits als intervals genòmics de les proteïnes del complex hi hem trobat motiu no descrit en la literatura i hem demostrat que Ibf-1 i Ibf-2 s'hi uneixen in vitro de manera específica. Són, al menys en els assaigs in vitro, capaces d'unir-s'hi de manera independent l'una de l'altra. Aquesta observació entra en aparent discordància amb la incapacitat que hem observat d'Ibf-1 per unir-se a la cromatina en mutants d'Ibf-2 i viceversa. És a dir, que tot i que totes dues tenen afinitat per aquest motiu per separat, Ibf-1 no s'hi uneix a menys que hi hagi present Ibf-2 i Ibf-2 no s'hi uneix si no és en presència d'Ibf-1. Aquest és un subjecte que s'hauria d'estudiar més a fons in vitro, realitzant EMSA amb combinacions de les dues proteïnes.

Una altra qüestió que queda per aprofundir és fins a quin punt Ibf-1 i Ibf-2 són necessàries per a que CP190 s'uneixi a la cromatina als llocs en que coincideixen totes tres. Hem vist a politènics mutants d'Ibf-2 que CP190 continua mostrant un extens patró de bandes, però la pregunta que ens estem fent sobrepassa amb escreix la resolució dels politènics. És a dir, en el cas que CP190 perdés alguns dels seus llocs d'unió degut a la manca d'Ibf-1 i 2, possiblement no ho podrem detectar per politènics ja que CP190 té molts altres llocs d'unió al genoma i identificar banda per banda seria una tasca no només complicada sinó també imprecisa. A més, és plausible que no sempre que Ibf-1 i 2 coincideixen amb CP190 fossin aquestes dues les que s'uneixen al DNA i CP190 a elles. Pot ser també que, al menys en alguns casos, sigui CP190 qui s'uneix al DNA i Ibf-1 i 2 a CP190. I aquí aflora una de les imprecisions de la tècnica del CHIP-Seq: els *reads* significativament enriquits respecte l'input són tractats com a llocs de la cromatina on hi ha la proteïna en qüestió, però això pot ser degut a que 1) la proteïna s'uneix a aquella seqüència de DNA, o 2) la proteïna està unida a una proteïna o complex que directa o indirectament estan units a aquella seqüència. En el cas de Ibf-1, Ibf-2 i CP190, molt possiblement els pics identificats són una barreja dels dos casos. Una opció per discriminar aquells pics en els que la proteïna s'uneix al DNA dels que la proteïna s'hi uneix a través d'altres proteïnes és només tractar com a llocs d'unió *bona fide* aquells en que hi ha el motiu consens d'aquella proteïna, tal com es proposa en un estudi recent (Van Bortle et al., 2012).

En qualsevol cas fóra interessant estudiar per CHIP fins a quin punt CP190 necessita Ibf-1 i 2 per unir-se a la cromatina. Si es fa per CHIP, en mosques mutants o en cèl·lules S2 amb RNAi, es poden escollir llocs concrets del genoma en base a les dades que ja tenim d'unió i en base a la presència del motiu consens que hem identificat per llavors analitzar el comportament de CP190.

Hem detectat també enriquiment d'altres motius, com els de CTCF i Su(Hw). Donada l'alta freqüència de solapament amb aquestes proteïnes, era quelcom que ja esperàvem, i que en certa manera valida la cerca de motius que

hem realitzat. Hi ha un motiu enriquit que ha cridat la nostra atenció perquè ha estat recentment descrit en dos articles que l'interpreten de manera diferent. Per una banda, Schwartz et al. (Schwartz et al., 2012) suggereixen que es tracta d'un motiu d'unió de CP190 al DNA, mentre que Van Bortle et al. (Van Bortle et al., 2012) el defineixen com un segon —i menor— motiu d'unió de CTCF. Nosaltres hem vist que ni Ibf-1 ni Ibf-2 s'hi uneixen *in vitro*. I tant nosaltres com els dos estudis mencionats el troben enriquit a molts pics de CP190. Per tal resoldre-ho s'hauria de fer un EMSA amb CP190 i CTCF amb aquest motiu per tal de veure si alguna de les dues hi té afinitat, tenint present la possibilitat gens menyspreable que sigui en realitat el motiu d'unió d'una altra proteïna d'insulator encara per descriure.

Analitzar el funcionament i mecanismes de proteïnes d'insulator no és evident, d'entrada perquè el que fan —separar regions reguladores— és un rol mecanístic dins del nucli, i no està per tant lligat a un grup de gens en particular o a una via de senyalització en particular sinó que afecta a regions molt diverses, i llavors a nivell de fenotip no té per què donar indicacions clares. La manera que s'utilitza és en assaigs amb constructes transgènics per determinar la contribució a l'enhancer blocking, que és una característica dels insulators. Nosaltres hem vist que Ibf-1 i Ibf-2 són necessàries per l'activitat enhancer-blocking dels insulators del Complex Bitòrax, que són els més estudiats de *Drosophila*. Això, juntament amb la seva extensa i clara colocalització al genoma amb els llocs on s'uneixen les altres proteïnes d'insulator (inclosos els insulator bodies al nucli), juntament amb la seva capacitat d'unir-se al DNA, i juntament amb l'alteració de l'expressió gènica que provoca la seva manca, considerem que Ibf-1 i Ibf-2 són dos proteïnes d'insulator, considerant proteïnes d'insulator aquelles que s'uneixen als insulators i que la seva funció principal n'és el correcte funcionament.

S'ha d'estudiar amb profunditat la manera com aquestes proteïnes contribueixen al funcionament dels insulators. D'entrada, se n'ha d'estudiar el per què de la interacció amb CP190. Es creu que CP190 fa d'engranatge entre les

proteïnes d'insulator unides al DNA i això genera loops i barreres entre regions reguladores. Si aquest fos el cas, podria ser que Ibf-1 i Ibf-2 s'unissin al DNA a través del reconeixement del seu motiu consens i que amb la interacció amb CP190 acabessin formant loops entre diferents llocs que tenen aquest motiu. Una altra possibilitat és que CP190 s'unís als seus propis llocs d'unió, i que es formés un loop amb els llocs d'unió Ibf a través de la interacció entre les tres. Podria fins i tot ser que aquest escenari també es donés amb CTCF i que les Ibf interaccionessin amb CTCF. En qualsevol dels casos seria molt informatiu fer experiments d'interaccions intra- i inter-cromosòmiques (els anomenats 3C i totes les seves variants) comparant wild-type i mutant de cadascuna de les proteïnes conegudes d'insulator, per així determinar quina és la contribució de cadascuna a la formació de contactes entre seqüències. Una altra possibilitat és el ChiA-PET, en que es determina la contribució d'una proteïna en particular a la formació de xarxes de contactes entre seqüències.

En el cas de Ibf-1 i Ibf-2, com són els insulators que estan regulant? Analitzant els seus llocs d'unió hem detectat un fort enriquiment a promotors head-to-head pròxims. La seva presència a aquests llocs podria deure's a la necessitat que té el genoma de mantenir aïllada la regulació d'un i altre promotor malgrat la proximitat (com s'ha suggerit per BEAF). Un altre lloc on es troben preferentment és a gens amb promotors alternatius. En aquest cas, potser utilitzant el mateix mecanisme però amb un escenari diferent, s'encarreguen de mantenir separada la regulació dels diferents promotors d'un gen. Així, s'hauria de veure si degut a la manca d'Ibf-1 i Ibf-2 es produeixen canvis en l'expressió relativa dels diferents trànscrips sorgits dels diferents promotors. Si es confirma, s'hauria d'intentar saber com els mantenen aïllats. Si és a través de loops que deixen fora el promotor que no ha d'estar actiu (o viceversa) es podria veure per 3C. Podria ser però que no fos a través de la formació de loops que actuen, sinó reclutant altres proteïnes que desconexim. És per això que en un moment donat pot ser necessari purificar i analitzar el

complex format per Ibf-1 i Ibf-2 i veure amb qui més interaccionen a part de CP190.

Conèixer el funcionament d'aquestes dues proteïnes no només pot contribuir a ampliar els coneixements sobre com funciona el nucli de les cèl·lules de *Drosophila* sinó que contribueix a la comprensió de la regulació del genoma humà. Les proteïnes d'insulator de *Drosophila*, tret de CTCF, no s'han conservat fins a vertebrats, però la seva funció i mecanismes d'acció sí. CTCF a vertebrats també aïlla regions reguladores formant loops i interaccions inter-cromosòmiques, i com Ibf-1 i Ibf-2, s'ha vist enriquida a gens amb promotors alternatius (Kim et al., 2007). La manera com els regulen i en general com funcionen els insulators podria ser similar i és, per varis motius, més accessible l'estudi de la genètica molecular a *Drosophila* que a vertebrats. Donada la gran importància que tenen els insulators sobre el desenvolupament de l'organisme serà important continuar aprofundint en el coneixement de Ibf-1, Ibf-2 i CP190.

6.- CONCLUSIONS

1. L'element Fab-6 de la regió reguladora del gen homeòtic Abdominal-B, que separa els dominis iab-5 i iab-6, conté seqüències corresponents a un insulator i a un *Polycomb-response element* (PRE). L'insulator presenta activitat enhancer-blocking dependent de CP190 i CTCF.
2. L'element iab6PRE, situat al domini iab-6 de la regió reguladora d'Abdominal-B, només conté seqüències corresponents a un PRE i no a un insulator.
3. Els insulators de la regió reguladora del gen Abdominal-B Fab-6 i Fab-7 es transcriuen, i ho fan en la mateixa direcció de transcripció que Abdominal-B. Només es transcriuen si també ho fa Abdominal-B, quan està silenciada també ho estan els insulators.
4. JIL-1 és necessària per l'activitat *enhancer-blocking* del Fab-6 i Fab-7.
5. CP190 forma complex amb les proteïnes d'insulator CTCF, Su(Hw) i Mod(mdg4) i també amb les proteïnes Nurf-38, Z4, Ibf-1 i Ibf-2.
6. Z4 és necessària per l'activitat *enhancer-blocking* del Fab-6 i Fab-7.
7. Ibf-1 i Ibf-2 són proteïnes similars i conservades dins de *Drosophila* i tenen un domini conservat BED zinc finger d'unió al DNA.
8. Ibf-1 i Ibf-2 són proteïnes de localització nuclear que s'uneixen al DNA amb afinitat per un motiu consens.
9. Ibf-1 i Ibf-2 es necessiten mútuament per unir-se a la cromatina. Totes dues s'uneixen a la cromatina independentment de CP190.
10. Ibf-1 i Ibf-2 colocalitzen extensament amb CP190 a la cromatina i parcialment amb les demés proteïnes d'insulator.
11. Ibf-1, Ibf-2 i CP190 estan enriquides a gens amb promotors alternatius i a parells de gens pròxims amb promotors divergents.
12. Ibf-1, Ibf-2 i CP190 són necessaris per la correcta expressió de gens de cèl·lules S2, a qui co-regulen en la mateixa direcció.
13. Ibf-1 i Ibf-2 tenen activitat insulator en assaigs enhancer-blocking a mosques adultes als Fab-6, Fab-7 i Fab-8.

7.- BIBLIOGRAFIA

- Adam, M., Robert, F., Laroche, M., and Gaudreau, L. (2001). H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. *Molecular and Cellular Biology* 21, 6270–6279.
- Adryan, B., Woerfel, G., Birch-Machin, I., Gao, S., Quick, M., Meadows, L., Russell, S., and White, R. (2007). Genomic mapping of Suppressor of Hairy-wing binding sites in *Drosophila*. *Genome Biology* 8, R167.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Molecular Biology of the Cell* (Sciences, Garland).
- Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25, 3389–3402.
- Andersson, L., Andersson, G., Hjälm, G., Jiang, L., Lindblad-Toh, K., Lindroth, A.M., Markljung, E., Nyström, A.-M., Rubin, C.-J., and Sundström, E. (2010). ZBED6: The birth of a new transcription factor in the common ancestor of placental mammals. *Transcription* 1, 144–148.
- Aravind, L. (2000). The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends in Biochemical Sciences* 25, 421–423.
- Arney, K.L., Bae, E., Olsen, C., and Drewell, R. a (2006). The human and mouse H19 imprinting control regions harbor an evolutionarily conserved silencer element that functions on transgenes in *Drosophila*. *Development Genes and Evolution* 216, 811–819.
- Ayoubi, T. a, and Van De Ven, W.J. (1996). Regulation of gene expression by alternative promoters. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 10, 453–460.
- Bae, E., Calhoun, V.C., Levine, M., Lewis, E.B., and Drewell, R. a (2002). Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the *Drosophila* bithorax complex. *Proceedings of the National Academy of Sciences of the United States of America* 99, 16847–16852.
- Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33, 729–740.
- Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308.
- Bao, X., Cai, W., Deng, H., Zhang, W., Krencik, R., Girton, J., Johansen, J., and Johansen, K.M. (2008). The COOH-terminal domain of the JIL-1 histone H3S10 kinase interacts with histone H3 and is required for correct targeting to chromatin. *The Journal of Biological Chemistry* 283, 32741–32750.
- Bao, X., Zhang, W., Krencik, R., Deng, H., Wang, Y., Girton, J., Johansen, J., and Johansen, K.M. (2005). The JIL-1 kinase interacts with lamin Dm0 and regulates nuclear lamina morphology of *Drosophila* nurse cells. *Journal of Cell Science* 118, 5079–5087.
- Barges, S., Mihaly, J., Galloni, M., Hagstrom, K., Müller, M., Shanower, G., Schedl, P., Gyurkovics, H., and Karch, F. (2000). The Fab-8 boundary defines the distal limit of the bithorax complex iab-7 domain and insulates iab-7 from initiation elements and a PRE in the adjacent iab-8 domain. *Development (Cambridge, England)* 127, 779–790.
- Barkess, G., and West, A.G. (2012). Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics* 4, 67–80.
- Bartkuhn, M., Straub, T., Herold, M., Herrmann, M., Rathke, C., Saumweber, H., Gilfillan, G.D., Becker, P.B., and Renkawitz, R. (2009). Active promoters and insulators are marked by the centrosomal protein 190. *The EMBO Journal* 28, 877–888.
- Batta, K., Zhang, Z., Yen, K., Goffman, D.B., and Pugh, B.F. (2011). Genome-wide function of H2B ubiquitylation in promoter and genic regions. *Genes & Development* 25, 2254–2265.
- Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 387–396.
- Van Bommel, J.G., Pagie, L., Braunschweig, U., Brugman, W., Meuleman, W., Kerkhoven, R.M., and Van Steensel, B. (2010). The insulator protein SU(HW) fine-tunes nuclear lamina interactions of the *Drosophila* genome. *PLoS One* 5, e15013.
- Bender, M., Horikami, S., Cribbs, D., and Kaufman, T.C. (1988). Identification and expression of the gap segmentation gene hunchback in *Drosophila melanogaster*. *Developmental Genetics* 9, 715–732.
- Bhaumik, S.R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. *Nature Structural & Molecular Biology* 14, 1008–1016.

- Blanton, J., Gaszner, M., and Schedl, P. (2003). Protein: protein interactions and the pairing of boundary elements in vivo. *Genes & Development* 664–675.
- Blobel, G.A., Kadauke, S., Wang, E., Lau, A.W., Zuber, J., Chou, M.M., and Vakoc, C.R. (2009). A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Molecular Cell* 36, 970–983.
- Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews. Genetics* 13, 770–780.
- Bonchuk, A., Denisov, S., Georgiev, P., and Maksimenko, O. (2011). Drosophila BTB/POZ domains of “ttk group” can form multimers and selectively interact with each other. *Journal of Molecular Biology* 412, 423–436.
- Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J.E., and Corces, V.G. (2012). Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Research* gr.136788.111–.
- Bouazoune, K., and Brehm, A. (2006). ATP-dependent chromatin remodeling complexes in Drosophila. *Chromosome Research : an International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* 14, 433–449.
- Boulet, a M., Lloyd, a, and Sakonju, S. (1991). Molecular definition of the morphogenetic and regulatory functions and the cis-regulatory elements of the Drosophila Abd-B homeotic gene. *Development (Cambridge, England)* 111, 393–405.
- Bracken, A.P., and Helin, K. (2009). Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nature Reviews. Cancer* 9, 773–784.
- Brown, D.T. (2001). Histone variants: are they functionally heterogeneous? *Genome Biology* 2, REVIEWS0006.
- Buratowski, S. (2003). The CTD code. *Nature Structural Biology* 10, 679–680.
- Bushey, A.M., Dorman, E.R., and Corces, V.G. (2008). Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Molecular Cell* 32, 1–9.
- Bushey, A.M., Ramos, E., and Corces, V.G. (2009). Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. *Genes & Development* 23, 1338–1350.
- Busturia, a, Lloyd, a, Bejarano, F., Zavortink, M., Xin, H., and Sakonju, S. (2001). The MCP silencer of the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. *Development (Cambridge, England)* 128, 2163–2173.
- Butcher, R.D.J., Chodagam, S., Basto, R., Wakefield, J.G., Henderson, D.S., Raff, J.W., and Whitfield, W.G.F. (2004). The Drosophila centrosome-associated protein CP190 is essential for viability but not for cell division. *Journal of Cell Science* 117, 1191–1199.
- Butler, J.E.F., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & Development* 16, 2583–2592.
- Byrd, K., and Corces, V. (2003). Visualization of chromatin domains created by the gypsy insulator of Drosophila. *The Journal of Cell Biology* 565–574.
- Cai, H., and Levine, M. (1995). Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo. *Nature* 376, 533–536.
- Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Current Opinion in Genetics & Development* 14, 155–164.
- Capelson, M., and Corces, V.G. (2004). Boundary elements and nuclear organization. *Biology of the Cell / Under the Auspices of the European Cell Biology Organization* 96, 617–629.
- Capelson, M., and Corces, V.G. (2005). The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Molecular Cell* 20, 105–116.
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. *Science (New York, N.Y.)* 309, 1559–1563.
- Casanova, J., Sánchez-Herrero, E., and Morata, G. (1986). Identification and characterization of a parasegment specific regulatory element of the abdominal-B gene of Drosophila. *Cell* 47, 627–636.
- Cavalli, G., and Paro, R. (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* 93, 505–518.
- Cavalli, G., and Paro, R. (1999). Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science (New York, N.Y.)* 286, 955–958.

- Celniker, S.E., Dillon, L.A.L., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M., et al. (2009). Unlocking the secrets of the genome. *Nature* **459**, 927–930.
- Celniker, S.E., Keelan, D.J., and Lewis, E.B. (1989). The molecular genetics of the bithorax complex of *Drosophila*: characterization of the products of the Abdominal-B domain. *Genes & Development* **3**, 1424–1436.
- Chen, Q., Lin, L., Smith, S., Lin, Q., and Zhou, J. (2005). Multiple Promoter Targeting Sequences exist in Abdominal-B to regulate long-range gene activation. *Developmental Biology* **286**, 629–636.
- Chiba, K., Yamamoto, J., Yamaguchi, Y., and Handa, H. (2010). Promoter-proximal pausing and its release: Molecular mechanisms and physiological functions. *Experimental Cell Research*.
- Chodagam, S., Royou, a, Whitfield, W., Karess, R., and Raff, J.W. (2005). The centrosomal protein CP190 regulates myosin function during early *Drosophila* development. *Current Biology : CB* **15**, 1308–1313.
- Chung, J.H., Whiteley, M., and Felsenfeld, G. (1993). A 5' Element of the Chicken β -Globin Domain Serves as an Insulator in Human Erythroid Cells and Protects against Position Effect in *Drosophila*. *Cell* **74**, 505–514.
- Ciavatta, D., Rogers, S., and Magnuson, T. (2007). *Drosophila* CTCF is required for Fab-8 enhancer blocking activity in S2 cells. *Journal of Molecular Biology* **373**, 233–239.
- Clancy, S. (2008). DNA Transcription. *Nature Education* **1**.
- Cléard, F., Moshkin, Y., Karch, F., and Maeda, R.K. (2006). Probing long-distance regulatory interactions in the *Drosophila melanogaster* bithorax complex using Dam identification. *Nature Genetics* **38**, 931–935.
- Corces, V.G., and Geyer, P.K. (1991). Interactions of retrotransposons with the host genome: the case of the gypsy element of *Drosophila*. *Trends in Genetics : TIG* **7**, 86–90.
- Cotney, J., Leng, J., Oh, S., DeMare, L.E., Reilly, S.K., Gerstein, M.B., and Noonan, J.P. (2012). Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Research* **22**, 1069–1080.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931–21936.
- Cuvier, O., Hart, C.M., and Laemmli, U.K. (1998). Identification of a class of chromatin boundary elements. *Molecular and Cellular Biology* **18**, 7478–7486.
- Dean, A. (2011). In the loop: long range chromatin interactions and gene regulation. *Briefings in Functional Genomics* **10**, 3–10.
- Delorenzi, M., and Bienz, M. (1990). Expression of Abdominal-B homeoproteins in *Drosophila* embryos. *Development (Cambridge, England)* **108**, 323–329.
- Deng, H., Bao, X., Cai, W., Blacketer, M.J., Belmont, A.S., Girton, J., Johansen, J., and Johansen, K.M. (2008). Ectopic histone H3S10 phosphorylation causes chromatin structure remodeling in *Drosophila*. *Development (Cambridge, England)* **135**, 699–705.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380.
- Djebali, S., Davis, C. a., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. *Nature* **489**, 101–108.
- Drewell, R. a, Bae, E., Burr, J., and Lewis, E.B. (2002). Transcription defines the embryonic domains of cis-regulatory activity at the *Drosophila* bithorax complex. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16853–16858.
- Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., and Reuter, G. (2004). *Su(var)* genes regulate the balance between euchromatin and heterochromatin in *Drosophila*. *Genes & Development* **18**, 2973–2983.
- egger, G., Liang, G., Aparicio, A., and Jones, P.A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**, 457–463.
- Eggert, H., Gortchakov, A., and Saumweber, H. (2004). Identification of the *Drosophila* interband-specific protein Z4 as a DNA-binding zinc-finger protein determining chromosomal structure. *Journal of Cell Science* **117**, 4253–4264.

- Egloff, S., Dienstbier, M., and Murphy, S. (2012). Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends in Genetics : TIG* *28*, 333–341.
- Emberly, E., Blattes, R., Schuettengruber, B., Hennion, M., Jiang, N., Hart, C.M., Käs, E., and Cuvier, O. (2008). BEAF regulates cell-cycle genes through the controlled deposition of H3K9 methylation marks into its conserved dual-core binding sites. *PLoS Biology* *6*, 2896–2910.
- Erokhin, M., Davydova, A., Kyrchanova, O., Parshikov, A., Georgiev, P., and Chetverina, D. (2011). Insulators form gene loops by interacting with promoters in *Drosophila*. *Development (Cambridge, England)* *138*, 4097–4106.
- Fan, H.-Y., He, X., Kingston, R.E., and Narlikar, G.J. (2003). Distinct strategies to make nucleosomal DNA accessible. *Molecular Cell* *11*, 1311–1322.
- Fiering, S., Whitelaw, E., and Martin, D.I. (2000). To be or not to be active: the stochastic nature of enhancer action. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* *22*, 381–387.
- Filion, G., Bommel, J. van, and Braunschweig, U. (2010). Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* *143*, 212–224.
- Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & Development* *17*, 1870–1881.
- Fonseca, J.P., Steffen, P.A., Müller, S., Lu, J., Sawicka, A., Seiser, C., and Ringrose, L. (2012). In vivo Polycomb kinetics and mitotic chromatin binding distinguish stem cells from differentiated cells. *Genes & Development* *26*, 857–871.
- Gan, M., Moebus, S., Eggert, H., and Saumweber, H. (2011). The Chr₂-Z4 complex recruits JIL-1 to polytene chromosomes, a requirement for interband-specific phosphorylation of H3S10. *Journal of Biosciences* *36*, 425–438.
- Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Reviews. Genetics* *7*, 703–713.
- Gaszner, M., Vazquez, J., and Schedl, P. (1999). The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes & Development* *13*, 2098–2107.
- Gause, M., Schaaf, C. a, and Dorsett, D. (2008). Cohesin and CTCF: cooperating to control chromosome conformation? *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* *30*, 715–718.
- Gdula, D. a, Gerasimova, T.I., and Corces, V.G. (1996). Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* *93*, 9378–9383.
- Gdula, D.A., Sandaltzopoulos, R., Tsukiyama, T., Ossipow, V., and Wu, C. (1998). Inorganic pyrophosphatase is a component of the *Drosophila* nucleosome remodeling factor complex. *Genes & Development* *12*, 3206–3216.
- Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., L, G., and Ge, Y. (2004). BioConductor: open software development for computational biology and bioinformatics. *Genome Biology* *5*.
- Gerasimova, T.I., and Corces, V.G. (2001). Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annual Review of Genetics* *35*, 193–208.
- Gerasimova, T.I., Lei, E.P., Bushey, A.M., and Corces, V.G. (2007). Coordinated control of dCTCF and gypsy chromatin insulators in *Drosophila*. *Molecular Cell* *28*, 761–772.
- Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korbil, J.O., Emanuelsson, O., Zhang, Z.D., Weissman, S., and Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Research* *17*, 669–681.
- Ghosh, D., Gerasimova, T.I., and Corces, V.G. (2001). Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. *The EMBO Journal* *20*, 2518–2527.
- Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K.B., Collins, J.B., Li, L., Gilmour, D.S., and Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes & Development* *22*, 1921–1933.
- Giresi, P.G., and Lieb, J.D. (2009). Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods (San Diego, Calif.)* *48*, 233–239.

- Golovnin, A., Mazur, A., Kopantseva, M., Kurshakova, M., Gulak, P. V., Gilmore, B., Whitfield, W.G.F., Geyer, P., Pirrotta, V., and Georgiev, P. (2007). Integrity of the Mod(mdg4)-67.2 BTB domain is critical to insulator function in *Drosophila melanogaster*. *Molecular and Cellular Biology* *27*, 963–974.
- Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006a). RNAi components are required for nuclear clustering of Polycomb group response elements. *Cell* *124*, 957–971.
- Grimaud, C., Nègre, N., and Cavalli, G. (2006b). From genetics to epigenetics: the tale of Polycomb group and trithorax group genes. *Chromosome Research: an International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* *14*, 363–375.
- Gruzdeva, N., Kyrchanova, O., Parshikov, A., Kullyev, A., and Georgiev, P. (2005). The Mcp element from the bithorax complex contains an insulator that is capable of pairwise interactions and can facilitate enhancer-promoter communication. *Molecular and Cellular Biology* *25*, 3682–3689.
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., De Klein, A., Wessels, L., De Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* *453*, 948–951.
- Gummalla, M., Maeda, R.K., Castro Alvarez, J.J., Gyurkovics, H., Singari, S., Edwards, K.A., Karch, F., and Bender, W. (2012). abd-A regulation by the iab-8 noncoding RNA. *PLoS Genetics* *8*, e1002720.
- Hagstrom, K., Muller, M., and Schedl, P. (1996). Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the *Drosophila* bithorax complex. *Genes & Development* *10*, 3202–3215.
- Hagstrom, K., Muller, M., and Schedl, P. (1997). A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the *Drosophila* bithorax complex. *Genetics* *146*, 1365–1380.
- Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W.H., Ye, C., Ping, J.L.H., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genetics* *43*, 630–638.
- Harrison, D.A., Geyer, P.K., Spana, C., and Corces, V.G. (1989). The gypsy retrotransposon of *Drosophila melanogaster*: mechanisms of mutagenesis and interaction with the suppressor of Hairy-wing locus. *Developmental Genetics* *10*, 239–248.
- Hart, C.M., Cuvier, O., and Laemmli, U.K. (1999). Evidence for an antagonistic relationship between the boundary element-associated factor BEAF and the transcription factor DREF. *Chromosoma* *108*, 375–383.
- Hart, C.M., Zhao, K., and Laemmli, U.K. (1997). The scs' boundary element: characterization of boundary element-associated factors. *Molecular and Cellular Biology* *17*, 999–1009.
- Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. *Proceedings of the National Academy of Sciences* 1–6.
- Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* *459*, 108–112.
- Herold, M., Bartkuhn, M., and Renkawitz, R. (2012). CTCF: insights into insulator function during development. *Development (Cambridge, England)* *139*, 1045–1057.
- Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nature Methods* *6*, 283–289.
- Hochheimer, A., Zhou, S., Zheng, S., Holmes, M.C., and Tjian, R. (2002). TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* *420*, 439–445.
- Holdridge, C., and Dorsett, D. (1991). Repression of hsp70 heat shock gene transcription by the suppressor of hairy-wing protein of *Drosophila melanogaster*. *Molecular and Cellular Biology* *11*, 1894–1900.
- Holohan, E.E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S., and White, R. (2007). CTCF genomic binding sites in *Drosophila* and the organisation of the bithorax complex. *PLoS Genetics* *3*, e112.
- Hou, C., and Corces, V.G. (2012). Throwing transcription for a loop: expression of the genome in the 3D nucleus. *Chromosoma* *121*, 107–116.

- Hou, C., Li, L., Qin, Z.S., and Corces, V.G. (2012). Gene Density, Transcription, and Insulators Contribute to the Partition of the *Drosophila* Genome into Physical Domains. *Molecular Cell* *45*, 1–14.
- Iampietro, C., Gummalla, M., Mutero, A., Karch, F., and Maeda, R.K. (2010). Initiator elements function to determine the activity state of BX-C enhancers. *PLoS Genetics* *6*, e1001260.
- Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research* *31*, e15.
- Ivaldi, M.S., Karam, C.S., and Corces, V.G. (2007). Phosphorylation of histone H3 at Ser10 facilitates RNA polymerase II release from promoter-proximal pausing in *Drosophila*. *Genes & Development* *21*, 2818–2831.
- Jack, J., Dorsett, D., Delotto, Y., and Liu, S. (1991). Expression of the cut locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer. *Development (Cambridge, England)* *113*, 735–747.
- Jiang, N., Emberly, E., Cuvier, O., and Hart, C.M. (2009). Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in *Drosophila melanogaster* links BEAF to transcription. *Molecular and Cellular Biology* *29*, 3556–3568.
- Jin, Y., Wang, Y., Johansen, J., and Johansen, K.M. (2000). JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *The Journal of Cell Biology* *149*, 1005–1010.
- Jin, Y., Wang, Y., Walker, D.L., Dong, H., Conley, C., Johansen, J., and Johansen, K.M. (1999). JIL-1: a novel chromosomal tandem kinase implicated in transcriptional regulation in *Drosophila*. *Molecular Cell* *4*, 129–135.
- Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology* *339*, 225–229.
- Kahn, T.G., Schwartz, Y.B., Dellino, G.I., and Pirrotta, V. (2006). Polycomb complexes and the propagation of the methylation mark at the *Drosophila* *ubx* gene. *The Journal of Biological Chemistry* *281*, 29064–29075.
- Kalverda, B., and Fornerod, M. (2010). Characterization of genome-nucleoporin interactions in *Drosophila* links chromatin insulators to the nuclear pore complex. *Cell Cycle (Georgetown, Tex.)* *9*, 4812–4817.
- Karch, F., Galloni, M., Sipos, L., Gausz, J., Gyurkovics, H., and Schedl, P. (1994). Mcp and Fab-7: molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of *Drosophila melanogaster*. *Nucleic Acids Research* *22*, 3138–3146.
- Karch, F., Weiffenbach, B., Peifer, M., Bender, W., Duncan, I., Celniker, S., Crosby, M., and Lewis, E.B. (1985). The abdominal region of the bithorax complex. *Cell* *43*, 81–96.
- Kellner, W. a, Ramos, E., Van Bortle, K., Takenaka, N., and Corces, V.G. (2012). Genome-wide phosphoacetylation of histone H3 at *Drosophila* enhancers and promoters. *Genome Research* *22*, 1081–1088.
- Kellogg, D.R., and Alberts, B.M. (1992). Purification of a multiprotein complex containing centrosomal proteins from the *Drosophila* embryo by chromatography with low-affinity polyclonal antibodies. *Molecular Biology of the Cell* *3*, 1–11.
- Kellogg, D.R., Oegema, K., Raff, J., Schneider, K., and Alberts, B.M. (1995). CP60: a microtubule-associated protein that is localized to the centrosome in a cell cycle-specific manner. *Molecular Biology of the Cell* *6*, 1673–1684.
- Kennison, J.A. (1995). The Polycomb and trithorax group proteins of *Drosophila*: trans-regulators of homeotic gene function. *Annual Review of Genetics* *29*, 289–303.
- Keogh, M.-C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., et al. (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* *123*, 593–605.
- Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenko, V. V, and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* *128*, 1231–1245.
- Kornberg, R.D. (2007). The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 12955–12961.
- Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* *98*, 285–294.

- Koshland, D., and Strunnikov, A. (1996). Mitotic chromosome condensation. *Annual Review of Cell and Developmental Biology* 12, 305–333.
- Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. *Molecular Cell* 44, 410–423.
- Kugler, S.J., and Nagel, A.C. (2010). A novel PzG-NURF complex regulates Notch target gene activity. *Molecular Biology of the Cell* 21, 3443–3448.
- Kutach, A.K., and Kadonaga, J.T. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Molecular and Cellular Biology* 20, 4754–4764.
- Kyrchanova, O., Ivlieva, T., Toshchakov, S., Parshikov, A., Maksimenko, O., and Georgiev, P. (2011). Selective interactions of boundaries with upstream region of Abd-B promoter in *Drosophila* bithorax complex and role of dCTCF in this process. *Nucleic Acids Research* 39, 3042–3052.
- Kyrchanova, O., Toshchakov, S., Podstreshnaya, Y., Parshikov, A., and Georgiev, P. (2008). Functional interaction between the Fab-7 and Fab-8 boundaries and the upstream promoter region in the *Drosophila* Abd-B gene. *Molecular and Cellular Biology* 28, 4188–4195.
- Lacoste, N., Utley, R.T., Hunter, J.M., Poirier, G.G., and Côte, J. (2002). Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. *The Journal of Biological Chemistry* 277, 30421–30424.
- Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & Development* 12, 34–44.
- Lanzuolo, C., Roue, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nature Cell Biology* 9, 1167–1174.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)* 23, 2947–2948.
- Lewis, B.A., Sims, R.J., Lane, W.S., and Reinberg, D. (2005). Functional characterization of core promoter elements: DPE-specific transcription requires the protein kinase CK2 and the PC4 coactivator. *Molecular Cell* 18, 471–481.
- Lewis, E.B. (1982). Control of body segment differentiation in *Drosophila* by the bithorax gene complex. *Progress in Clinical and Biological Research* 85 Pt A, 269–288.
- Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. *Current Opinion in Genetics & Development* 21, 175–186.
- Li, H., and Müller, M. (2011). Insulators, Not Polycomb Response Elements, Are Required for Long-Range Interactions between Polycomb Targets in *Drosophila melanogaster*. ... and *Cellular Biology* 31, 616–625.
- Li, M., Belozero, V.E., and Cai, H.N. (2010). Modulation of chromatin boundary activities by nucleosome-remodeling activities in *Drosophila melanogaster*. *Molecular and Cellular Biology* 30, 1067–1076.
- Lin, Q., Lin, L., and Zhou, J. (2010). Chromatin insulator and the promoter targeting sequence modulate the timing of long-range enhancer-promoter interactions in the *Drosophila* embryo. *Developmental Biology* 339, 329–337.
- Lis, J.T., Mason, P., Peng, J., Price, D.H., and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. *Genes & Development* 14, 792–803.
- Lobanenkov, V. V., Nicolas, R.H., Adler, V. V., Paterson, H., Klenova, E.M., Polotskaja, A. V., and Goodwin, G.H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* 5, 1743–1753.
- Lorch, Y., Zhang, M., and Kornberg, R.D. (2001). RSC unravels the nucleosome. *Molecular Cell* 7, 89–95.
- Lusser, A., and Kadonaga, J.T. (2003). Chromatin remodeling by ATP-dependent molecular machines. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 25, 1192–1200.
- Längst, G., and Becker, P.B. (2004). Nucleosome remodeling: one mechanism, many phenomena? *Biochimica Et Biophysica Acta* 1677, 58–63.

- Maeda, R.K., and Karch, F. (2006). The ABC of the BX-C: the bithorax complex explained. *Development (Cambridge, England)* *133*, 1413–1422.
- Martin, C.H., Mayeda, C.A., Davis, C.A., Ericsson, C.L., Knafels, J.D., Mathog, D.R., Celniker, S.E., Lewis, E.B., and Palazzolo, M.J. (1995). Complete sequence of the bithorax complex of *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* *92*, 8398–8402.
- Martin, D.I. (2001). Transcriptional enhancers--on/off gene regulation as an adaptation to silencing in higher eukaryotic nuclei. *Trends in Genetics : TIG* *17*, 444–448.
- Mazo, A., Hodgson, J.W., Petruk, S., Sedkov, Y., and Brock, H.W. (2007). Transcriptional interference: an unexpected layer of complexity in gene regulation. *Journal of Cell Science* *120*, 2755–2761.
- McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proceedings of the National Academy of Sciences of the United States of America* *101*, 1525–1530.
- Mihaly, J., Barges, S., Sipos, L., Maeda, R., Cléard, F., Hogga, I., Bender, W., Gyurkovics, H., and Karch, F. (2006). Dissecting the regulatory landscape of the Abd-B gene of the bithorax complex. *Development (Cambridge, England)* *133*, 2983–2993.
- Mihaly, J., Hogga, I., Barges, S., Galloni, M., Mishra, R.K., Hagstrom, K., Müller, M., Schedl, P., Sipos, L., Gausz, J., et al. (1998). Chromatin domain boundaries in the Bithorax complex. *Cellular and Molecular Life Sciences : CMLS* *54*, 60–70.
- Mishra, R., Mihaly, J., and Barges, S. (2001). The *iab-7* polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. ... and *Cellular Biology* *21*, 1311–1318.
- Mohan, M., Bartkuhn, M., Herold, M., Philippen, A., Heinel, N., Bardenhagen, I., Leers, J., White, R.A.H., Renkawitz-Pohl, R., Saumweber, H., et al. (2007). The *Drosophila* insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *The European Molecular Biology Organization Journal* *26*, 4203–4214.
- Mokhonov, V. V., Theendakara, V.P., Gribanova, Y.E., Ahmedli, N.B., and Farber, D.B. (2012). Sequence-specific binding of recombinant Zbed4 to DNA: insights into Zbed4 participation in gene transcription and its association with other proteins. *PLoS One* *7*, e35317.
- Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S.T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R., et al. (2005). CTCF is conserved from *Drosophila* to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Reports* *6*, 165–170.
- Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. *Trends in Biochemical Sciences* *35*, 323–332.
- Moshkovich, N., Nisha, P., Boyle, P.J., Thompson, B.A., Dale, R.K., and Lei, E.P. (2011). RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. *Genes & Development* *25*, 1686–1701.
- Muller, M., Hagstrom, K., Gyurkovics, H., Pirrotta, V., and Schedl, P. (1999). The *mcp* element from the *Drosophila melanogaster* bithorax complex mediates long-distance regulatory interactions. *Genetics* *153*, 1333–1356.
- Neff, T., Shotkoski, F., and Stamatoyannopoulos, G. (1997). Stem cell gene therapy, position effects and chromatin insulators. *Stem Cells (Dayton, Ohio)* *15 Suppl 1*, 265–271.
- Nègre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C. a, Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R. a H., et al. (2010). A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genetics* *6*, e1000814.
- Oegema, K., Marshall, W.F., Sedat, J.W., and Alberts, B.M. (1997). Two proteins that cycle asynchronously between centrosomes and nuclear structures: *Drosophila* CP60 and CP190. *Journal of Cell Science* *110 (Pt 1)*, 1573–1583.
- Oegema, K., Whitfield, W.G., and Alberts, B. (1995). The cell cycle-dependent localization of the CP190 centrosomal protein is determined by the coordinate action of two separable domains. *The Journal of Cell Biology* *131*, 1261–1273.
- Ohshima, N., Takahashi, M., and Hirose, F. (2003). Identification of a human homologue of the DREF transcription factor with a potential role in regulation of the histone H1 gene. *The Journal of Biological Chemistry* *278*, 22928–22938.
- Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E.M., and Müller, J. (2008). Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in *Drosophila*. *Developmental Cell* *15*, 877–889.

- Oliver, D., Sheehan, B., South, H., Akbari, O., and Pai, C.-Y. (2010). The chromosomal association/dissociation of the chromatin insulator protein Cp190 of *Drosophila melanogaster* is mediated by the BTB/POZ domain and two acidic regions. *BMC Cell Biology* *11*, 101.
- Pai, C., Lei, E.P., Ghosh, D., and Corces, V.G. (2004). The Centrosomal Protein CP190 Is a Component of the gypsy Chromatin Insulator. *Molecular Cell* *16*, 737–748.
- Painter, T.S., and Griffen, A.B. (1937). The Structure and the Development of the Salivary Gland Chromosomes of *Simulium*. *Genetics* *22*, 612–633.
- Pal, S., Gupta, R., Kim, H., Wickramasinghe, P., Baubet, V., Showe, L.C., Dahmane, N., and Davuluri, R. V (2011). Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. *Genome Research* *21*, 1260–1272.
- Palstra, R.-J., and Grosveld, F. (2012). Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. *Frontiers in Genetics* *3*, 195.
- Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes & Development* *20*, 2041–2054.
- Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* *132*, 422–433.
- Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Molecular and Cellular Biology* *27*, 3769–3779.
- Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W., and Mazo, A. (2006). Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* *127*, 1209–1221.
- Planet, E., Stephan-Otto, C., Reina, O., Flores, O., and Rossell, D. (2012). htSeqTools: Quality control, Visualization and Processing for High-throughput sequencing data.
- Prince, V.E., and Pickett, F.B. (2002). Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews. Genetics* *3*, 827–837.
- Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. *Cell* *108*, 501–512.
- Ptashne, M., and Gann, A. (2002). *Genes and Signals* (CSHL Press).
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods (San Diego, Calif.)* *24*, 218–229.
- Pérez-Lluch, S., Cuartero, S., Azorín, F., and Espinàs, M.L. (2008). Characterization of new regulatory elements within the *Drosophila* bithorax complex. *Nucleic Acids Research* *36*, 6926–6933.
- Ramos, E., Ghosh, D., Baxter, E., and Corces, V.G. (2006). Genomic organization of gypsy chromatin insulators in *Drosophila melanogaster*. *Genetics* *172*, 2337–2349.
- Rank, G., Prestel, M., and Paro, R. (2002). Transcription through intergenic chromosomal memory elements of the *Drosophila* bithorax complex correlates with an epigenetic switch. *Molecular and Cellular Biology* *22*, 8026–8034.
- Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proceedings of the National Academy of Sciences of the United States of America* *99*, 6883–6888.
- Regnard, C., Straub, T., Mitterweger, A., Dahlsveen, I.K., Fabian, V., and Becker, P.B. (2011). Global Analysis of the Relationship between JIL-1 Kinase and Transcription. *PLoS Genetics* *7*, 17.
- Reina, O., and Rossell, D. (2012). chroGPS: navigating through the epigenome.
- Richter, C., Oktaba, K., Steinmann, J., Müller, J., and Knoblich, J.A. (2011). The tumour suppressor L(3)mbt inhibits neuroepithelial proliferation and acts on insulator elements. *Nature Cell Biology* *13*, 1029–1039.
- Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development (Cambridge, England)* *134*, 223–232.
- Ringrose, L., Rehmsmeier, M., Dura, J.-M., and Paro, R. (2003). Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*. *Developmental Cell* *5*, 759–771.

- Roseman, R.R., Pirrotta, V., and Geyer, P.K. (1993). The su(Hw) protein insulates expression of the *Drosophila melanogaster* white gene from chromosomal position-effects. *The EMBO Journal* *12*, 435–442.
- Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteché, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J. a, and Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 8309–8314.
- Sadava, D., Heller, H., Hillis, D., and Berenbaum, M. (2009). *Life: The science of Biology* (WH Freeman).
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (2001). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press).
- Santaguida, S., and Musacchio, A. (2009). The life and miracles of kinetochores. *The EMBO Journal* *28*, 2511–2531.
- Sarma, K., and Reinberg, D. (2005). Histone variants meet their match. *Nature Reviews. Molecular Cell Biology* *6*, 139–149.
- Sawado, T., Halow, J., Bender, M.A., and Groudine, M. (2003). The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes & Development* *17*, 1009–1018.
- Schibler, U., and Sierra, F. (1987). Alternative promoters in developmental gene expression. *Annual Review of Genetics* *21*, 237–257.
- Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. *Genes & Development* *19*, 697–708.
- Schneider, J., and Shilatifard, A. (2006). Histone demethylation by hydroxylation: chemistry in action. *ACS Chemical Biology* *1*, 75–81.
- Schoborg, T. a, and Labrador, M. (2010). The phylogenetic distribution of non-CTCF insulator proteins is limited to insects and reveals that BEAF-32 is *Drosophila* lineage specific. *Journal of Molecular Evolution* *70*, 74–84.
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell* *128*, 735–745.
- Schwartz, B.E., and Ahmad, K. (2005). Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes & Development* *19*, 804–814.
- Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nature Genetics* *38*, 700–705.
- Schwartz, Y.B., Linder-Basso, D., Kharchenko, P. V, Tolstorukov, M.Y., Kim, M., Li, H.-B., Gorchakov, A. a, Minoda, A., Shanower, G., Alekseyenko, A. a, et al. (2012). Nature and function of insulator protein binding sites in the *Drosophila* genome. *Genome Research*.
- Schweinsberg, S., Hagstrom, K., Gohl, D., Schedl, P., Kumar, R.P., Mishra, R., and Karch, F. (2004). The enhancer-blocking activity of the Fab-7 boundary from the *Drosophila* bithorax complex requires GAGA-factor-binding sites. *Genetics* *168*, 1371–1384.
- Scott, K.C., Taubman, A.D., and Geyer, P.K. (1999). Enhancer blocking by the *Drosophila* gypsy insulator depends upon insulator anatomy and enhancer strength. *Genetics* *153*, 787–798.
- Scott, K.S., and Geyer, P.K. (1995). Effects of the su(Hw) insulator protein on the expression of the divergently transcribed *Drosophila* yolk protein genes. *The EMBO Journal* *14*, 6258–6267.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* *148*, 458–472.
- Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annual Review of Biochemistry* *76*, 75–100.
- She, W., Lin, W., Zhu, Y., Chen, Y., Jin, W., Yang, Y., Han, N., Bian, H., Zhu, M., and Wang, J. (2010). The gypsy insulator of *Drosophila melanogaster*, together with its binding protein suppressor of Hairy-wing, facilitate high and precise expression of transgenes in *Arabidopsis thaliana*. *Genetics* *185*, 1141–1150.
- Sheng, Y., and Previt, C. (2011). Genomic features and computational identification of human microRNAs under long-range developmental regulation. *BMC Genomics* *12*, 270.
- Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annual Review of Biochemistry* *75*, 243–269.

- Sipos, L., and Gyurkovics, H. (2005). Long-distance interactions between enhancers and promoters. *The FEBS Journal* *272*, 3253–3259.
- Smith, S.T., Wickramasinghe, P., Olson, A., Loukinov, D., Lin, L., Deng, J., Xiong, Y., Rux, J., Sachidanandam, R., Sun, H., et al. (2009). Genome wide ChIP-chip analyses reveal important roles for CTCF in Drosophila genome organization. *Developmental Biology* *328*, 518–528.
- Spana, C., Harrison, D.A., and Corces, V.G. (1988). The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. *Genes & Development* *2*, 1414–1423.
- Stamatoyannopoulos, J. a. (2012). What does our genome encode? *Genome Research* *22*, 1602–1611.
- Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* *403*, 41–45.
- Strutt, H., Cavalli, G., and Paro, R. (1997). Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *The EMBO Journal* *16*, 3621–3632.
- Swinton, J. (2011). Vennable: Venn and Euler area-proportional diagrams.
- Sánchez-Herrero, E., and Akam, M. (1989). Spatially ordered transcription of regulatory DNA in the bithorax complex of Drosophila. *Development (Cambridge, England)* *107*, 321–329.
- Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. *Nature Reviews. Molecular Cell Biology* *11*, 264–275.
- Taniue, K., Nishida, A., Hamada, F., Sugie, A., Oda, T., Ui-Tei, K., Tabata, T., and Akiyama, T. (2010). Sunspot, a link between Wingless signaling and endoreplication in Drosophila. *Development (Cambridge, England)* *137*, 1755–1764.
- Team, R.D.C. (2011). R: A language and environment for statistical computing.
- Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular Cell* *10*, 1453–1465.
- Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* *128*, 651–654.
- Udvardy, A., Maine, E., and Schedl, P. (1985). The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *Journal of Molecular Biology* *185*, 341–358.
- Vazquez, J., and Schedl, P. (1994). Sequences required for enhancer blocking activity of scs are located within two nuclease-hypersensitive regions. *The European Molecular Biology Organization Journal* *13*, 5984–5993.
- Vogelmann, J., Valeri, A., and Guillou, E. (2011). Roles of chromatin insulator proteins in higher-order chromatin organization and transcription regulation. *Nucleus* *2*, 358–369.
- Voncken, J.W., Schweizer, D., Aagaard, L., Sattler, L., Jantsch, M.F., and Van Lohuizen, M. (1999). Chromatin-association of the Polycomb group protein BMI1 is cell cycle-regulated and correlates with its phosphorylation status. *Journal of Cell Science* *112 (Pt 2)*, 4627–4639.
- Wallace, J.A., and Felsenfeld, G. (2007). We gather together: insulators and genome organization. *Current Opinion in Genetics & Development* *17*, 400–407.
- Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* *474*, 390–394.
- Wang, Y., Zhang, W., Jin, Y., Johansen, J., and Johansen, K.M. (2001). The JIL-1 tandem kinase mediates histone H3 phosphorylation and is required for maintenance of chromatin structure in Drosophila. *Cell* *105*, 433–443.
- Waterhouse, R.M., Zdobnov, E.M., Tegenfeldt, F., Li, J., and Kriventseva, E. V (2011). OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. *Nucleic Acids Research* *39*, D283–8.
- Weis, L., and Reinberg, D. (1997). Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Molecular and Cellular Biology* *17*, 2973–2984.
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* *451*, 796–801.
- West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. *Genes & Development* *16*, 271–288.

- Whitfield, W.G., Chaplin, M. a, Oegema, K., Parry, H., and Glover, D.M. (1995). The 190 kDa centrosome-associated protein of *Drosophila melanogaster* contains four zinc finger motifs and binds to specific sites on polytene chromosomes. *Journal of Cell Science* *108* (Pt 1, 3377–3387.
- Wood, A.M., Van Bortle, K., Ramos, E., Takenaka, N., Rohrbaugh, M., Jones, B.C., Jones, K.C., and Corces, V.G. (2011). Regulation of chromatin organization and inducible gene expression by a *Drosophila* insulator. *Molecular Cell* *44*, 29–38.
- Yang, J., and Corces, V.G. (2011). Chromatin insulators: a role in nuclear organization and gene expression. *Advances in Cancer Research* *110*, 43–76.
- Yang, J., Ramos, E., and Corces, V.G. (2012). The BEAF-32 insulator coordinates genome organization and function during the evolution of *Drosophila* species. *Genome Research* 1–20.
- Yang, L., and Yu, J. (2009). A comparative analysis of divergently-paired genes (DPGs) among *Drosophila* and vertebrate genomes. *BMC Evolutionary Biology* *9*, 55.
- Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nature Genetics* *39*, 1512–1516.
- Zhang, W., Deng, H., Bao, X., Lerach, S., Girton, J., Johansen, J., and Johansen, K.M. (2006). The JIL-1 histone H3S10 kinase regulates dimethyl H3K9 modifications and heterochromatic spreading in *Drosophila*. *Development (Cambridge, England)* *133*, 229–235.
- Zhang, W., Jin, Y., Ji, Y., Girton, J., Johansen, J., and Johansen, K.M. (2003). Genetic and phenotypic analysis of alleles of the *Drosophila* chromosomal JIL-1 kinase reveals a functional requirement at multiple developmental stages. *Genetics* *165*, 1341–1354.
- Zhao, K., Hart, C.M., and Laemmli, U.K. (1995). Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell* *81*, 879–889.
- Zhimulev, I.F., Belyaeva, E.S., Vatolina, T.Y., and Demakov, S.A. (2012). Banding patterns in *Drosophila melanogaster* polytene chromosomes correlate with DNA-binding protein occupancy. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* *34*, 498–508.
- Zhou, J., Barolo, S., Szymanski, P., and Levine, M. (1996). The Fab-7 element of the bithorax complex attenuates enhancer-promoter interactions in the *Drosophila* embryo. *Genes & Development* *10*, 3195–3201.
- Zhou, J., and Levine, M. (1999). A novel cis-regulatory element, the PTS, mediates an anti-insulator activity in the *Drosophila* embryo. *Cell* *99*, 567–575.
- Zhu, L., Pages, H., Gazin, C., Lawson, N., Ou, J., Lin, S., Lapointe, D., and Green, M. (2012). ChIPpeakanno.
- Zhu, Q., and Halfon, M.S. (2009). Complex organizational structure of the genome revealed by genome-wide analysis of single and alternative promoters in *Drosophila melanogaster*. *BMC Genomics* *10*, 9.

