

El transcriptoma d'*ash2*: dianes i funció



Sergi Beltran i Agulló

**Departament de Genètica
Universitat de Barcelona**

INTRODUCCIÓ

1 - *Drosophila melanogaster*

Durant el darrer segle, la ciència s'ha servit de diversos organismes model per poder estudiar al laboratori problemes biològics amb més facilitat. Els fonaments de la majoria de processos biològics són compartits per una gran varietat d'espècies i fins i tot les bases genètiques són extrapolables entre ells. Segurament, un dels processos que es troba més conservat al llarg de l'evolució és el relacionat amb els gens homeòtics (Hox), encarregats d'establir el patró bàsic del cos dels individus (revisat a Veraksa, 2000). Aquests gens han estat àmpliament estudiats a la mosca del vinagre o *Drosophila melanogaster* que, gràcies a les seves característiques, és un dels organismes preferits per la comunitat científica. El temps de generació curt, la seva abundant descendència, el fàcil manteniment degut a les seves reduïdes dimensions i la facilitat per introduir i combinar mutacions en el seu genoma per tal d'inferir la funció dels gens a partir de l'observació dels fenotips generats, són alguns d'aquests avantatges.

1.1 - *Drosophila* com a organisme model

A principis del segle XX, Thomas Hunt Morgan va assentar les bases que permetrien establir *Drosophila* com a organisme model. En el 1910 va demostrar que la teoria de l'herència a través dels cromosomes era certa (Morgan, 1910) i en el 1913 un estudiant del grup, H. Sturtevant, va crear els primers mapes genètics de l'espècie (Sturtevant, 1913). Aquesta feina va ser reconeguda internacionalment amb l'atorgament a Thomas H. Morgan del Premi Nobel de Fisiologia o Medicina l'any 1933.

Des del punt de vista de la Biologia del Desenvolupament, *Drosophila* presenta algunes característiques concretes que li confereixen un atractiu especial. Aquest dípter presenta un cicle vital molt curt amb estadis larvals i adults separats per una etapa pupal (Figura I1), durant la qual té lloc una metamorfosi completa. És durant aquest procés que els discs imaginals són externalitzats, mitjançant un procés d'eversió, per donar lloc a les estructures de l'adult com les ales, els ulls o les potes (Figura I2). Els discs imaginals es determinen ja a l'epiteli embrionari a partir de cèl·lules precursoras que invaginaran com a sacs epitelials per donar lloc als discs, formats inicialment per entre 10 i 40 cèl·lules organitzades en una monocapa (Bate and Arias, 1991; Cohen et al., 1993). Al llarg dels estadis larvals aniran augmentant de mida mitjançant la divisió cel·lular a la vegada que s'anirà establint el seu patró, tot determinant-se, per exemple, quines cèl·lules seran anteriors i quines posteriors. Durant la metamorfosi, els discs imaginals sofriran processos de reorganització i

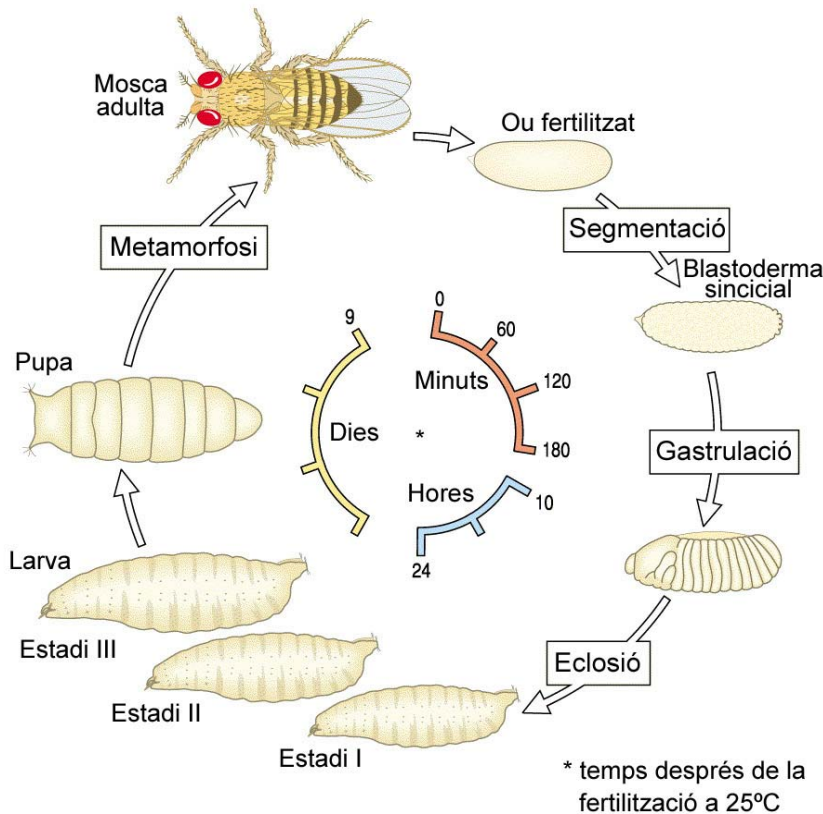


Figura I1 | Cicle vital de *Drosophila melanogaster*. Reproduït i traduït de Wolpert, L. et al. *Principles of Development (2nd Edition)*. 2001

les seves cèl·lules es diferenciaran per tal de donar lloc a l'estructura adulta per la qual havien estat determinades. Així doncs, els discs imaginals són uns teixits òptims per analitzar processos de determinació, diferenciació, proliferació i manteniment de la identitat cel·lulars, entre d'altres.

En general, *Drosophila* és un organisme model ideal per estudiar les bases dels processos biològics, ja que aquestes acostumen a ser extrapolables a d'altres organismes més complexos. Tant és així, que la comparació dels genomes de *Drosophila* i humans ha permès establir que dos terços dels gens relacionats amb malalties humanes tenen un homòleg a la mosca del vinagre (Reiter et al., 2001; Rubin et al., 2000). Segurament, la característica més rellevant que ha portat *Drosophila* a ser més estudiada que d'altres organismes, és la facilitat per realitzar anàlisi genètica ja que el seu genoma conté poc ADN repetitiu i, a diferència que en els mamífers, generalment no s'hi troba més d'un gen codificant per proteïnes redundants en estructura i/o funció. Aquesta eina tant poderosa permet inferir la funció dels gens a través de l'observació dels fenotips produïts per mutacions en els mateixos. A més, la comparació de fenotips generats per diferents mutacions i la combinació d'aquestes, permet establir relacions funcionals entre els gens i intuir si participen o no en els mateixos processos. Tanmateix, moltes de les mutacions aïllades acostumen a ser letals en

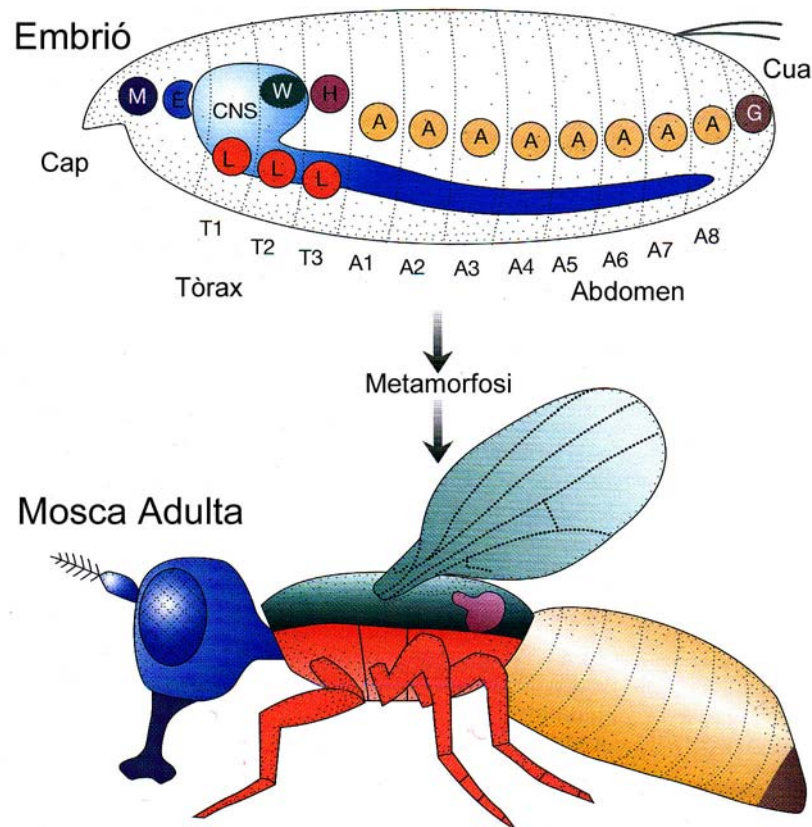


Figura I2 | Muntatge d'una mosca adulta a partir dels discs imaginals. Reproduït i traduït de Bier, E. *The coiled spring: how life begins*. 2000.

homozigosi, impossibilitant l'estudi del fenotip adult. Per tal d'evitar aquest problema es va desenvolupar una tècnica basada en la recombinació mitòtica que permet generar mosaics genètics, és a dir teixits heterozigots en els quals podem identificar, mitjançant marcadors, uns clons (o grups) de cèl·lules homozigotes per l'al·lel mutant i uns clons de cèl·lules homozigotes per l'al·lel salvatge (*wild-type* o *wt*). Aquesta tècnica, que es pot realitzar irradiant amb raigs X (Stern, 1936) o utilitzant unes seqüències (FRT) reconegudes per una recombinasa induïble (FLP; Golic and Lindquist, 1989; Xu and Rubin, 1993), ens permet comparar en un mateix teixit els efectes produïts per una mutació en heterozigosi i en homozigosi respecte el *wt* (Figura I3). Podem examinar, per exemple, quins fenotips produeix una mutació letal recessiva en l'ala adulta o testar si aquesta mateixa mutació provoca la reducció dels nivells d'una determinada proteïna. A més, l'ús de la tècnica *Minute* (Morata and Ripoll, 1975) permet estudiar mutacions que afecten la viabilitat cel·lular perquè dona avantatge proliferatiu a les cèl·lules homozigotes per la mutació al fer que les cèl·lules que haurien de ser *wt* no es desenvolupin ja que porten una mutació en un gen ribosomal.

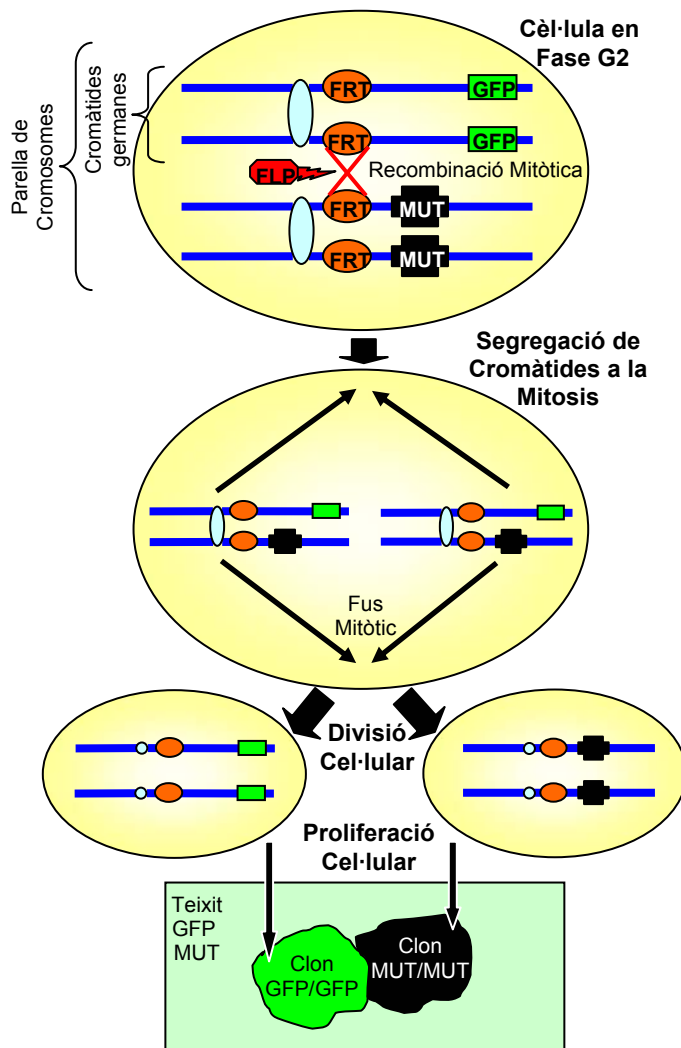


Figura 13 | Generació de clons per recombinació mitòtica. La Flipasa (FLP) s'expressa per l'activació d'un promotor induïble per xoc tèrmic (*heat shock* o *hs*) que reconeix les *Flipase Recombination Targets* (FRT o Diances de Recombinació per Flipasa), situades en els cromosomes complementaris, provocant la recombinació mitòtica. Es generen dues cèl·lules diferents, una homozigòtica per un marcador (en aquest cas *Green Fluorescent Protein*: GFP o Proteïna Verda Fluorescent) que serveix de control, i l'altra homozigòtica per una mutació (MUT). Les subseqüents divisions d'aquestes cèl·lules donaran lloc a dos clons diferents emmarcats en un fons heterozigot. El clon amb el marcador s'acostuma a anomenar *twin clone* (clon bessó).

L'extensiu estudi del qual ha estat objecte la mosca del vinagre ha permès emmagatzemar en bases de dades gran quantitat d'informació i desenvolupar multitud de tècniques que en faciliten encara més l'estudi. Entre les bases de dades són imprescindibles *Interactive Fly* (Brody, 1999) a <http://flybase.bio.indiana.edu/allied-data/lk/interactive-fly/aimain/1aahome.htm> i *Flybase* (Drysdale and Crosby, 2005) a <http://www.flybase.org>. La primera resumeix la bibliografia més destacada sobre els gens més estudiats de *Drosophila*, mentre que la segona presenta pràcticament tota la informació disponible sobre tots els gens descrits de l'organisme, des dels al·lells aïllats fins a les característiques de les proteïnes resultants, passant per la seva seqüència gènica i la seva funció. Precisament la publicació de la seqüència del genoma de *Drosophila* l'any 2000 (Adams et al., 2000) va marcar un punt d'inflexió, ja que ha permès la localització exacta de tots els gens, tant coneguts com predits, i

desenvolupar tècniques com els microarrays d'ADN o de cromatina, que permeten aproximacions a nivell genòmic. El genoma de *Drosophila* consta d'uns 165 milions de bases i conté 14.816 gens (segons la distribució 4.3 de l'anotació del genoma del 3 de Març de 2006 a <http://flybase.net/annot/>) distribuïts al llarg de quatre parells de cromosomes: els cromosomes sexuals X i Y i els autosomes 2, 3 i 4. El 2 i el 3 consten de braç dret (2R i 3R) i braç esquerre (2L i 3L), mentre que el 4 és de dimensions reduïdes. Una característica rellevant d'aquest organisme són els cromosomes politènics, els quals es divideixen moltes vegades a les glàndules salivals per poder produir gran quantitat de transcrits. Com que posteriorment les cadenes no se separen ni es produeix una divisió cel·lular, aquests cromosomes acaben essent molt gruixuts i són fàcilment observables a través d'un microscopi convencional. A més, quan es tenyeixen presenten un patró de bandes invariable (Figura I4). Per aquesta raó, els cromosomes estan dividits en 102 bandes (1 al 20 pel X, 21-60 pel 2, 61-100 pel 3 i 101-102 pel 4) dividides en 6 sub-bandes (A a F) subdividides al seu torn en fins a 13 porcions.

Si bé l'anàlisi genètica és essencial per relacionar gens i funció, es necessiten altres tècniques per entendre els processos a nivell molecular. La Genètica Molecular, la Biologia Cel·lular, la Bioquímica i més recentment la Bioinformàtica, proveeixen solucions que permeten abordar aquestes altres qüestions per tal d'assolir el repte d'entendre la Biologia des de tots els angles possibles a partir de l'anàlisi exhaustiva d'organismes model. *Drosophila* ocupa un lloc preferencial entre aquests, i l'estudi de la formació d'una ala adulta a partir d'unes poques cèl·lules embrionàries és un dels aspectes que més interès ha suscitat donat que necessita de l'acció coordinada de múltiples gens i processos, les bases dels quals són extrapolables a d'altres qüestions biològiques.



Figura I4 | Cromosomes politènics de glàndules salivals tenyits amb DAPI. En general, les bandes fosques corresponen a cromatina més condensada. Imatge cedida per M. Angulo.

1.2 - L'ala de *Drosophila melanogaster*

El disc imaginal d'ala prové de poc més de 20 cèl·lules invaginades durant l'embriogènesi a partir de l'ectoderm lateral del segment mesotoràcic (Bate and Arias, 1991; Cohen et al., 1993). Durant els estadis larvals, el disc proliferarà fins a generar una monocapa epitelial de desenes de milers de cèl·lules que anirant adquirint identitats diferenciades. Després de la metamorfosi, aquestes cèl·lules donaran lloc a les diferents estructures de l'ala adulta i el nòtum. Així doncs, el disc imaginal d'ala presenta durant els estadis larvals una alta activitat proliferativa finament regulada per gens d'establiment del patró, que determinaran la forma, la grandària, els compartiments i, en última instància, la seva funció. En aquests processos hi juguen un paper important les interaccions cèl·lula a cèl·lula i un ampli ventall de vies de transducció del senyal que seran claus per la construcció de l'òrgan adult. Totes aquestes característiques fan que el disc imaginal d'ala sigui un model extraordinari per estudiar una gran varietat de processos.

Un dels primers passos clau per la formació d'una ala és la compartimentalització del seu disc imaginal al llarg dels eixos pròximo-distal (P-D), dorso-ventral (D-V) i antero-posterior (A-P). Per entendre els compartiments i els principis de formació dels eixos ens podem fixar en l'eix A-P, objecte de nombrosos estudis que han permès conèixer amb força detall tant els gens involucrats com els mecanismes que estableixen el patró al llarg d'aquest eix.

Ja en els estadis embrionaris, el disc imaginal d'ala es subdivideix en compartiments anterior i posterior (A-P). Això es deu a que les cèl·lules que seran posteriors hereten l'expressió dels gens selectors *engrailed* (*en*) i *invected* (*inv*) al provenir de la regió anterior del parasegment (hi ha un desfàs respecte els segments) del qual invaginen les cèl·lules que formaran el disc (revisat a Blair, 2003 i Brook, 1996). En el compartiment posterior del disc imaginal d'ala, *Inv* provoquen l'expressió de *hedgehog* (*hh*) i inhibeixen l'activació de *cubitus interruptus* (*ci*), necessari per la transducció del senyal de Hh. Les cèl·lules de la frontera entre els dos compartiments, és a dir, les anteriors adjacents al compartiment posterior, reben la proteïna secretada Hh i com que no tenen *ci* reprimint hi poden respondre expressant, entre d'altres, *knot* (*kn*) i l'homòleg de TGF- β , *decapentaplegic* (*dpp*), que determinaran l'expressió d'altres gens per tal de conferir identitats cel·lulars diferencials i refinar el patró A-P de la regió central (*wing pouch*) del disc imaginal (Figura I5; revisat a De Celis, 2003). Cada tipus cel·lular donarà lloc posteriorment a una zona diferent de la fulla de l'ala adulta (*wing blade*) ja que aquesta presenta, al llarg de l'eix A-P, quatre venes longitudinals completes (L2-L5) en sentit P-D, dues venes longitudinals incompletes (L1 i L6) i una vena longitudinal marginal que ressegueix la vora anterior de l'ala (*wing margin*). A més, l'ala disposa de dues venes transversals (*cross-veins*) entre les venes L3-L4 i L4-L5 i d'estructures sensorials com les quetes (*bristles*; Figura I5). El procés de refinament del patró que ha de sofrir el

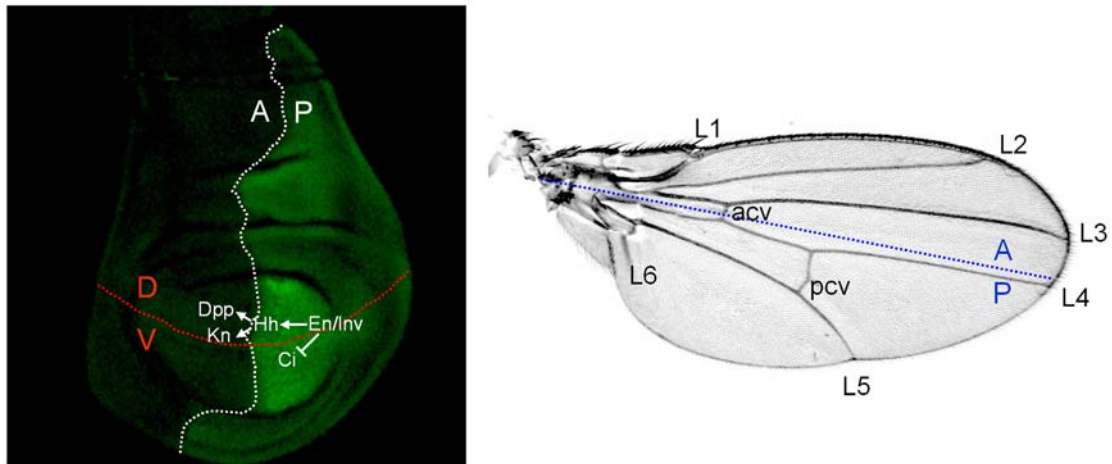


Figura 15 | Disc imaginal d'ala (esquerra) i ala adulta (dreta) de *D. melanogaster*. En el disc imaginal es mostren els compartiments anterior (A) i posterior (P) a través de la detecció d'En (en verd) en el P. Es mostra esquematitzada la relació entre algunes de les proteïnes implicades en l'establiment dels compartiments (veure text), dividits per la frontera, marcada amb una línia blanca. La línia vermella marca la frontera entre els compartiments dorsal (D) i ventral (V). A l'ala s'enumeren les venes longitudinals i les transversals anterior (acv) i posterior (pcv). La línia blava delimita els compartiments anterior (A) i posterior (P). Imatges del disc i de l'ala obtingudes amb microscopi confocal i de camp clar respectivament.

disc imaginal per tal de donar lloc a aquestes estructures ha d'estar molt ben orquestrat, i les cèl·lules han de "recordar" d'alguna manera el que són a mesura que es van dividint i especificant encara més. Així doncs, cada cèl·lula tindrà la seva pròpia identitat, i per tal de mantenir-la requerirà del que s'ha anomenat memòria cel·lular.

2 - La memòria cel·lular

Durant els anys 60 i 70, Ernst Hadorn va proposar que l'estat de determinació cel·lular durant el desenvolupament estaria controlat genèticament. Per arribar a aquesta conclusió va cultivar discs imaginals de *Drosophila* a la hemolimfa de l'abdomen de femelles adultes, per tal de que proliferessin sense entrar en metamorfosi. A continuació trasplantava els discs a larves que estaven entrant en metamorfosi. Tot i haver proliferat més del compte i en un ambient estrany, els discs imaginals donaven lloc a l'estructura adulta per la qual havien estat programats inicialment (Hadorn, 1963). Donat que les cèl·lules no perden la seva identitat, hi ha d'haver un mecanisme que la mantingui, que rep el nom de memòria cel·lular.

La pèrdua de la memòria cel·lular pot donar lloc a mort cel·lular no programada, a defectes en el desenvolupament o a transformacions. Un dels exemples clàssics són les transformacions homeòtiques, aberracions tals com l'aparició d'ales enlloc d'halteris o de potes en comptes d'antenes, que són degudes a mutacions en els gens homeòtics (gens Hox) o, en el cas de la memòria cel·lular, a la seva incorrecta expressió.

2.1 - Manteniment dels gens homeòtics durant l'establiment del patró inicial

Després de la fecundació, un conjunt de proteïnes maternes distribuïdes diferencialment al llarg de l'eix antero-posterior de l'embrió de *Drosophila*, activen unes cascades d'expressió gènica on hi participen els gens "gap", "pair-rule" i "segment-polarity". L'acció d'aquests delimitarà l'organització antero-posterior de l'embrió i activarà de manera diferencial els gens homeòtics (Hox), encarregats de determinar la identitat de cada segment. En estadis més avançats del desenvolupament, molts dels factors de transcripció que han provocat l'activació dels Hox ja no es troben presents, però aquests hauran de seguir essent expressats per tal de mantenir la identitat dels segments. Precisament, és la informació posicional de cada segment la que indueix la formació dels discs imaginals corresponents, fent que, per exemple, en el segment toràcic T2 aparegui el disc imaginal d'ala i en el T3 el d'halteris (Figura I2). Aquests discs imaginals proliferaran durant l'estadi larval mantenint la seva identitat tot i l'absència dels gens iniciadors del procés de determinació. Mutacions en els gens Hox o en els mecanismes de manteniment de la seva expressió condueixen a una incorrecta especificació dels segments, provocant alteracions en el posicionament dels discs imaginals. Aquestes mutacions provoquen les anomenades transformacions homeòtiques, en les quals alguna de les estructures adultes

derivades dels discs imaginals manca, es troba repetida, i/o en posició incorrecta. Per exemple, el gen homeòtic *Antennapedia* (*Antp*) especifica el segon segment toràctic (T2), d'on sorgiran els discs imaginals d'ala i els del primer parell de potes. L'expressió del gen *Antp* en el segment del cap provoca la transformació parcial del disc imaginal d'antena en un de pota de T2. Després de la metamorfosi, l'adult tindrà un fenotip aberrant, presentant potes enlloc d'antenes en el cap (Figura I6; Turner and Mahowald, 1979).

La cerca de transformacions homeòtiques a *Drosophila* mitjançant experiments de cribratge genètic, ha conduït a la identificació d'uns gens diferents que els homeòtics però que presenten transformacions com les d'aquests. Aquests gens s'han anat classificat en dos grans grups antagònics: els *Polycomb* (PcG) i els *trithorax* (*trxG*).



Figura I6 | Transformació homeòtica. El cap de l'esquerra és d'un organisme *wt* i el de la dreta correspon a un organisme amb una mutació del gen *Antp*, que li provoca una formació de potes enlloc d'antenes. Imatges de microscòpia electrònica realitzades per R. Turner.

2.2 - Els gens dels grups *Polycomb* (PcG) i *trithorax* (*trxG*)

El manteniment de la identitat dels discs imaginals és degut, fins a cert punt, a processos de regulació creuada (*cross-regulation*; Ali and Bender, 2004; Hafen et al., 1984; Krasnow et al., 1989) i d'auto-regulació (Bienz and Tremml, 1988; Carroll et al., 1986; Kuziora and McGinnis, 1988) entre els gens homeòtics. Són però els gens dels grups *Polycomb* (PcG) i *trithorax* (*trxG*) els que tenen el paper més important en el manteniment d'aquesta identitat durant la fase proliferativa. Mutacions de pèrdua de funció en gens PcG tenen com a conseqüència l'expressió ectòpica de gens homeòtics. Durant el tercer estadi larval, el gen *Ultrabithorax* (*Ubx*), per exemple, s'expressa en el disc imaginal d'halterí però no

en el d'ala. Mutacions de pèrdua de funció de gens PcG desencadenen l'expressió d'*Ubx* en el disc imaginal d'ala (Jones and Gelbart, 1990) conduint a transformacions d'ala a halteri en l'adult. Aquestes transformacions són similars a les provocades per *Contrabithorax* (*Cbx*), un al·lel de guany de funció d'*Ubx* (Lindsey and Zimm, 1992). D'altra banda, mutacions de pèrdua de funció dels gens *trxG* causen la pèrdua d'expressió dels gens *Hox* (Breen and Harte, 1991; Breen and Harte, 1993; Sedkov et al., 1994; Tamkun et al., 1992). Per exemple, mutacions en un dels gens d'aquest grup, *ash1*, provoca la pèrdua d'expressió d'*Ubx* en el disc imaginal d'halteri (LaJeunesse and Shearn, 1995), que desencadena la transformació de l'halteri a ala (Shearn, 1989; Shearn et al., 1987). Aquesta aberració és similar a la provocada per mutacions de pèrdua de funció d'*Ubx* com *bx* (Lindsey and Zimm, 1992).

Així doncs, mentre que els PcG són necessaris per mantenir la repressió gènica, els *trxG* tenen com a funció mantenir l'estat activat. Aquests gens, que acostumen a presentar una expressió ubiqua, es troben conservats en l'escala evolutiva des de llevats a mamífers (Otte and Kwaks, 2003). A més, el fet que cADNs (ADNs complementaris) de gens PcG o *trxG* de ratolí o humans rescatin parcialment el fenotip provocat per mutants del gen homòleg de *Drosophila*, demostra que existeix una conservació funcional entre les proteïnes (Atchison et al., 2003; Muller et al., 1995; Muyrers-Chen et al., 2004).

2.2.1 - Els gens PcG

Durant els anys quaranta es van identificar les primeres mutacions dels gens PcG: *extra sex combs* (*esc*) i *Polycomb* (*Pc*; Lindsey and Zimm, 1992; Slifer, 1942). Fins ara s'han identificat divuit gens PcG, que es classifiquen en tres grans grups: els que pertanyen a un, a l'altre o a cap dels dos complexos PcG aïllats fins al moment. La Taula I1 presenta una relació de gens PcG, indicant les seves característiques rellevants i alguns homòlegs de vertebrats. Només set de les divuit proteïnes PcG s'han trobat formant part de manera estable d'un d'aquests dos complexos, les subunitats dels quals varien lleugerament en funció de la metodologia i del material biològic utilitzats per aïllar-los. Tanmateix, s'ha demostrat que algunes de les altres proteïnes PcG poden interaccionar amb aquests dos complexos en alguns casos.

Donat que la funció d' Extra Sex Combs (ESC) és necessària molt aviat en el desenvolupament (Ingham, 1983; Struhl and Akam, 1985), es considera que el primer complexe necessari en la repressió per PcG és el Polycomb Repressive Complex 2 (PRC2 o ESC/E(Z)). S'han aïllat diferents variants d'aquest complexe amb el denominador comú de la presència de les proteïnes ESC i Enhancer of Zeste (E(Z); Czermin et al., 2002; Furuyama et al., 2003; Muller et al., 2002; Ng et al., 2000; Tie et al., 2001; Tie et al., 2003). Altres proteïnes que s'hi han trobat associades són les desacetilases d'histones Rpd3 (homòloga a la Histone

Deacetylase 1 (HDAC1) i Sir2, la proteïna d'unió a histones p55, Supressor of Zeste 12 (Su(Z)12) i Polycomb-Like (PcL; Czermin et al., 2002; Furuyama et al., 2004; Tie et al., 2001; Tie et al., 2003) . El fet que en diferents teixits i estadis de desenvolupament de l'organisme s'hagin trobat diferents proteïnes unides als complexos ESC/E(Z), apunta a que la composició d'aquests és dinàmica al llarg del desenvolupament (Furuyama et al., 2004; Furuyama et al., 2003; Otte and Kwaks, 2003). Pel que respecta al mecanisme d'acció, s'ha proposat que el reclutament dels complexos ESC/E(Z) es donaria a través de la unió de Pleiohomeotic (Pho) i Pleiohomeotic-Like (PhoL) als Elements de Resposta a PcG (PcG Response Elements o PREs), unes seqüències de la cromatina a les quals s'hi poden unir algunes de les proteïnes PcG i trxB permetent l'acoblament de complexos (Ringrose and Paro, 2004; Wang et al., 2004b). A continuació, les activitats desacetiladores d'algunes subunitats podrien modificar diferents substrats com histones o membres del mateix complexe (Chang et al., 2001; Furuyama et al., 2004) propiciant l'activitat metiltransferasa dels complexos ESC/E(Z). Aquests poden metilar les lisines 9 i/o 27 de la histona H3 (H3K9 i H3K27) a través del domini SET de E(Z), una modificació essencial per la repressió dels gens Hox (Czermin et al., 2002; Muller et al., 2002). Tot i així, no és de menysprear l'aportació de ESC, Su(Z)12 i p55, ja que tot i no disposar d'activitat catalítica *per se*, la seva associació amb E(Z) incrementa l'activitat metiladora sobre H3K27 en més de 1000 vegades (Ketel et al., 2005). La metilació d'H3K27 podria propiciar la unió de Pc mitjançant el seu cromodomini (Fischle et al., 2003), i la interacció d'aquesta amb ESC facilitaria el reclutament del Polycomb Repressive Complex 1 (PRC1), el segon complexe dels PcG descrit a *Drosophila* (Figura 17-A; Cao et al., 2002; Czermin et al., 2002; Poux et al., 2001b; Wang et al., 2004b).

PRC1 inclou, entre d'altres, Polycomb (Pc), Polyhomeotic (Ph), Posterior Sex Combs (PSC), dRing i Sex Comb on Midleg (SCM; Saurin et al., 2001; Shao et al., 1999). Entre les proteïnes associades, cal destacar Zeste i diferents Factors associats a TBP (TAFs; Saurin et al., 2001). El nucli (*core*) del complexe (anomenat PCC) pot ser reconstituït mitjançant la co-expressió de les 4 primeres proteïnes en cèl·lules, i pot bloquejar *in vitro* la remodelació de la cromatina que efectua el complexe humà SWI/SNF del grup trxB d'activadors (Francis et al., 2001; Shao et al., 1999). Si també es co-expressa Zeste, el complexe purificat conté PCC i Zeste i augmenta la seva capacitat repressora, en especial sobre els motlles (*templates*) amb motius d'unió de Zeste, un dels motius identificats en els PREs (Mulholland et al., 2003). Les altres dues seqüències ben caracteritzades que es troben en els PREs són la reconeguda pel GAGA Factor (GAF; codificat pel gen *trithorax like; Trl*) i Pipsqueak (Psq; Huang et al., 2002), i la reconeguda per Pleiohomeotic (Pho) o Pleiohomeotic-Like (PhoL). GAF interacciona amb Pc (Horard et al., 2000) i potencia la capacitat repressora de PCC sobre motlles amb seqüències GAGA (Mulholland et al., 2003). Pel que fa a Psq, s'ha trobat en

PcG <i>Drosophila</i>			Vertebrats
Proteïna	Gen	Informació rellevant	Proteïna
COMPLEXE PRC1			
Polycomb (Pc)	<i>Pc</i>	Cromodomini; uneix H3K27Me3, H3K9Me3	HPC1, HPC2, HPC3
Polyhomeotic (Ph)	<i>ph</i>	Dits de Zinc, domini SPM (SAM); s'uneix a SCM	HPH1, HPH2
Posterior Sex Combs (PSC)	<i>Psc</i>	Dits de Zinc, domini HTH; interacciona amb Pc i Ph	BMI1, MEL-18
dRING / Sex Combs Extra	<i>Sce</i>		RING1
COMPLEXE PRC2			
Enhancer of Zeste (E(Z))	<i>E(z)</i>	Histona metiltransferasa: el domini SET metila H3K9 i H3K27	EZH1 (ENX2), EZH2 (ENX1)
Extra Sex Combs (ESC)	<i>esc</i>	Repeticions WD40	EED
Supressor of Zeste 12 (SU(Z)12)	<i>Su(z)12</i>	Dit de Zinc, caixa VEFS	SU(Z)12
ALTRES PROTEÏNES PcG			
Extra Sex Combs-Like (ESCL)	<i>escl</i>		
Pleiohomeotic (Pho)	<i>pho</i>	Dit de Zinc que s'uneix a ADN (GCCATHWY); interacciona amb Pc	YY1
Pleiohomeotic-Like (PhoL)	<i>phol</i>	Dit de Zinc que s'uneix a DNA (GCCATHWY)	YY1
Polycomblike (PcL)	<i>Pcl</i>	Domini PHD i RING	PCL1, PCL2
Sex Comb on Midleg (SCM)	<i>Scm</i>	Domini SPM (SAM) que permet la interacció amb Ph	SCMH1, SCML2
Additional Sex Combs (ASX)	<i>As</i>	Dit de Zinc; domini ric en Q	ASXH2, ASXL1
Enhancer of Polycomb (E(Pc))	<i>E(Pc)</i>	Domini ric en Q i domini ric en A	EPC1
Supressor of Zeste 2 (SU(Z)2)	<i>Su(z)2</i>	RING	BMI1, MEL-18
Corto	<i>corto</i>	Cromodomini; interacciona amb E(Z), ESC, GAF	
Lola Like (batman)	<i>lolal</i>	Domini BTB-POZ	
Pipsqueak (Psq)	<i>psq</i>	Domini BTB-POZ; domini Psq; uneix (GA) _n	
PROTEÏNES ASSOCIADES A PcG			
Heat-shock protein cognate 4 (HSC70-4)	<i>HSC70-4</i>	Xaperona	
Rpd3	<i>Rpd3</i>	Desacetilasa	HDAC-1
Cramped (CRM)	<i>crm</i>	Domini ric en A; interacciona genèticament amb PCNA	
Mi-2	<i>Mi-2</i>	Dits PHD; cromodominis; domini ATPasa; motiu HMG-like; motiu myb-like	
Dorsal Switch Protein 1 (DSP1)	<i>Dsp1</i>	Domini HMG box	

Taula I1 | Resum dels PcG descrits a *Drosophila* i els seus homòlegs a vertebrats. S'han ordenat en funció de si pertanyen o no de manera estable a algun dels 2 complexos PcG descrits (codi de colors). S'enumeren algunes característiques rellevants de les proteïnes i els seus homòlegs a vertebrats. També es llisten algunes de les proteïnes que s'han trobat associades a PcG.

cèl·lules Schneider de *Drosophila* formant part d'una variant del complexe PRC1 anomenada CHRASCH, que també inclou HDAC1 (Huang and Chang, 2004). D'altra banda, Pho pot unir Pc i Ph (Mohd-Sarip et al., 2002) permetent que Pho i PCC s'uneixin de manera cooperativa a PREs amb com a mínim dos motius d'unió per Pho, cosa que permetria la formació d'un *silenciosoma* a través de l'engalament de múltiples proteïnes mitjançant interaccions proteïna-proteïna i proteïna-ADN (Mohd-Sarip et al., 2005). Així doncs, després de l'acció dels complexos ESC/E(Z), la unió del PRC1 a l'ADN seria possible gràcies a proteïnes tals com GAF, Psq, Zeste o Pho, que reconeixerien motius determinats. El silenciament de la cromatina podria ser degut, per exemple, a ubiquitinitzacions (Wang et al., 2004a), activitats modificadores d'histones com la de HDAC1 (Chang et al., 2001), o a actuacions sobre la maquinària transcripcional (Dellino et al., 2004).

2.2.2 - Els gens *trxG*

Els primers gens del grup *trithorax* (*trxG*) es van començar a identificar com a supressors del fenotip provocat per mutacions del gen *extra sex combs* (*esc*) del grup dels Polycomb. Tanmateix, malgrat que tots els gens classificats com a *trxG* suprimeixen els fenotips generats per mutacions en els PcG, només al·lels dels gens *trithorax* (*trx*) i *absent, small or homeotic discs 1 i 2* (*ash1* i *ash2*) provoquen transformacions homeòtiques per si mateixos, que es veuen potenciades si s'utilitzen dobles heterozigots (Ingham, 1983; Shearn, 1989; Shearn et al., 1987) Així doncs, s'ha proposat que aquests tres gens (*trx*, *ash1* i *ash2*) serien antagonitzadors específics de la funció dels PcG però que els altres *trxG* caurien en la classe d'activadors o co-activadors transcripcionals generals (Dejardin and Cavalli, 2005). A la Taula I2 s'enumeren els gens de *Drosophila* classificats com a *trxG*, les seves característiques principals i alguns homòlegs de vertebrats. Les proteïnes *trxG* també formen diferents complexos proteics, que tal i com passa en el cas del PcG, presenten variants en funció de com i d'on s'hagin aïllat. Tot i que se n'han descrit sis de diferents, aquests no inclouen tots els gens classificats com a *trxG*.

El Trithorax Acetylation Complex 1 (TAC1) té un pes molecular d'1 MDa, va ser aïllat a partir d'embrions de *Drosophila* i conté Trx, la proteïna anti-fosfatasa SET Domain Binding Factor (SBF) i l' acetiltransferasa d'histones CREB-Binding Protein (dCBP), la qual cosa suggereix un rol de la hiperacetilació en el manteniment de la transcripció (Figura I7-A; Cavalli and Paro, 1999; Petruk et al., 2001). Els components de TAC1 són necessaris per l'activació *in vivo* del gen *Ubx* i poden unir-se a PREs (també anomenades Elements de Resposta a Trithorax o TREs; Petruk et al., 2001). A més, Trx presenta un domini SET, molt similar al de la proteïna Set1 del llevat *Sacharomyces cerevisiae*, que li confereix activitat

trxG <i>Drosophila</i>			Vertebrats
Proteïna	Gen	Informació rellevant	Proteïna
COMPLEXE TAC1			
Trithorax (Trx)	<i>trx</i>	Domini SET i PHD; Histona metiltransferasa: H3K4	MLL/ALL-1
COMPLEXE ASH1			
Absent, Small, or Homeotic discs 1 (ASH1)	<i>ash1</i>	Domini SET. Histona metiltransferasa: H3K9, H3K4; H4K20	ASH1L
COMPLEXE ASH2			
Absent, Small, or Homeotic discs 2 (ASH2)	<i>ash2</i>	Domini PHD i SPRY. Homòleg a Bre2 i Spp1 de llevat	ASH2L
COMPLEXES BRM			
Brahma (Brm)	<i>brm</i>	Subunitat ATPasa en complexe Brahma. Té bromodomini i domini helicasa. Homòleg de SWI2/SNF2 de llevat	BRG1, HBRM
Osa (Osa)	<i>osa</i>	Domini ARID d'unió a l'ADN	ELD/OSA1
Moira (Mor)	<i>mor</i>	Homòleg a SWI3 de llevat.	BAF170/BAF155
COMPLEXE FACT			
GAGA Factor (GAF)	<i>Trl</i>	Dit de Zinc; domini BTB-POZ; interacciona amb Trx; s'uneix a ADN: (GA) _n	
ALTRES PROTEÏNES PcG			
Zeste	<i>z</i>	Domini HTH i cremallera de Leucines; forma oligòmers	
Kismet (Kis)	<i>kis</i>	Cromodomini; domini SNF-2 related; Helicasa d'ADN, DEDXc, BRK; domini Myb-like	
Tonalli	<i>tna</i>	SP-RING, dit de Zinc	
Taranis	<i>tara</i>		
Kohtalo	<i>kto</i>	Complexe mediador	
PROTEÏNES ASSOCIADES A trxG			
CREB-binding protein	<i>dCBP</i>	Histona acetiltransferasa; interacciona amb ASH1 i Trx	
SBF (o SBF1)	<i>Sbf</i>	Fosfatasa de proteïnes	
SPT16	<i>dre4</i>	Desenrotllador d'ADN	
SSRP1	<i>Ssrp</i>	Domini HMG box	
Modifier of mdg4	<i>mod (mdg4)</i>	Domini BTB-POZ, participa en la funció insuladora de Gypsy	
Snf5 related 1	<i>Snr1</i>	Interacciona amb Trx i Brm	
Little imaginal discs	<i>lid</i>	Domini PHD i ARID	

Taula I2 | Resum dels trxG descrits a *Drosophila* i els seus homòlegs a vertebrats. S'han ordenat en funció de si pertanyen o no de manera estable a algun dels 6 complexos trxG descrits (codi de colors). S'enumeren algunes característiques rellevants de les proteïnes i els seus homòlegs a vertebrats. També es llisten algunes de les proteïnes que s'han trobat associades a trxG.

histona metiltransferasa, actuant principalment sobre la H3K4 (Smith et al., 2004). Precisament, la metilació de la H3K4, i més concretament la seva tri-metilació, és una modificació associada als gens transcripcionalment actius (Lachner et al., 2003; Santos-Rosa et al., 2002).

Els dos complexos Brahma (Brm) descrits a *Drosophila* són segurament els més ben caracteritzats dintre els trxG i estan molt conservats al llarg de l'escala evolutiva, ja que presenten moltes similituds amb els complexos homòlegs de llevats i humans. L'anomenat BAP va ser el primer en ser caracteritzat, té un pes molecular de 2 MDa i està relacionat amb els complexos SWI/SNF de llevat i BAF-SWI/SNF- α humà. S'han identificat diverses subunitats del nucli del complexe, que inclouen Brahma (Brm), Moira (Mor) i Osa (Osa) i com a mínim quatre subunitats accessòries com Snr1 (Collins et al., 1999; Mohrmann et al., 2004; Papoulas et al., 1998). El segon complexe, PBAP, es diferencia del primer perquè conté Polybromo i BAP170 però no Osa. Aquest complexe seria homòleg als complexos RSC i BAF-SWI/SNF- β de llevats i humans respectivament (Mohrmann et al., 2004). A més, tot i que aquests dos complexos es troben en alguns loci diferents de cromosomes politènics, tots dos es troben associats a cromatina hiperacetilada. Aquesta modificació es relaciona amb regions actives del genoma, i la proteïna Pc, antagonista de Brm, no s'hi troba associada (Mohrmann et al., 2004). Finalment, ja que diferents subunitats dels complexos Brm interaccionen amb Zeste i Pho, s'ha proposat que aquestes dues proteïnes, que tenen capacitat d'unir-se a l'ADN, serien uns dels nexes d'unió dels complexos amb la cromatina (Kal et al., 2000; Mohd-Sarip et al., 2002).

Per la seva banda, els complexos quart i cinquè tenen un pes molecular de 2 MDa i 500 kDa i contenen ASH1 i ASH2, respectivament. La proteïna Trx no va ser detectada en aquests complexos i malauradament, les altres subunitats no han pogut ser identificades (Papoulas et al., 1998). De totes maneres, treballs recents han demostrat que ASH1 interacciona físicament amb Trx (Rozovskaia et al., 1999) i amb la histona acetiltransferasa dCBP (Bantignies et al., 2000) i que té un domini SET amb activitat histona metiltransferasa (Beisel et al., 2002; Byrd and Shearn, 2003).

Finalment, el sisè complexe que inclou un component del trxG, concretament GAF, és l'anomenat FACT (Facilitates Chromatin Transcription), que facilita l'elongació transcripcional per l'ARN polimerasa II en mamífers i coopera amb GAF in vivo per la correcta activació dels gens Hox (Orphanides et al., 1998; Shimojima et al., 2003).

Curiosament, tot i que les proteïnes PcG i trxG que interaccionen directament amb l'ADN s'han classificat dins d'un o altre grup (veure Taules I1 i I2), algunes s'han trobat a vegades interaccionant amb proteïnes del grup oposat o el resultat de certs experiments ha estat invers a l'esperat. Per exemple, GAF va ser classificat inicialment com a trxG perquè les mutacions del gen que el codifica, *Trl*, produïen fenotips de pèrdua de funció homeòtica (Farkas et al., 1994). Que

GAF interaccioni amb Trx i que estimuli la transcripció, donen validesa a aquesta classificació (Mahmoudi et al., 2002; Poux et al., 2002). Tanmateix, altres dades apunten en el sentit contrari, ja que s'ha detectat GAF en complexos PcG i col·labora en la repressió mitjançada per PRC1 (Horard et al., 2000; Poux et al., 2002). Sembla ser que algunes d'aquestes proteïnes col·laborarien indistintament amb els PcG o trxG en funció del context, ja que la seva funció principal seria la de permetre l'accés a la cromatina i ser un nexa d'unió entre aquesta i altres proteïnes per tal de poder formar complexos (Hur et al., 2002; Poux et al., 2002). D'altra banda, la regulació creuada entre els gens homeòtics pot haver conduït a la classificació errònia d'alguns gens si aquesta classificació s'ha fet només a partir dels resultats d'interaccions gèniques (Kennison, 2004).

Vist doncs el coneixement actual de les proteïnes PcG i trxG i dels seus complexos, ens queda preguntar-nos com exerceixen la seva funció; és a dir, com reconeixen els gens diana sobre els que han d'actuar? I no menys important, com es recorda després de la divisió cel·lular que aquells gens han de continuar sent reprimits o activats?

2.2.3 - Mecanismes de funcionament de les proteïnes PcG i trxG

Per tal que els gens PcG i trxG puguin mantenir la identitat cel·lular necessiten identificar els gens diana abans de que els seus activadors o repressors desapareguin, tot reconeixent si el destí del gen és el d'estar activat o reprimat. A continuació han d'establir un mecanisme per mantenir l'estat transcripcional del gen, que ha de ser restablert després de les mitosis, la qual cosa implica algun sistema per transferir la informació més enllà de la divisió cel·lular.

a) Localització dels gens diana: els PRE

Tincions de cromosomes politènics de glàndules salivals amb anticossos contra proteïnes PcG (Franke et al., 1992; Rastelli et al., 1993), Trx i ASH1 (Chinwalla et al., 1995; Rozovskaia et al., 1999; Tripoulas et al., 1996) indiquen que com a mínim en aquests teixits els primers regulen uns 100 loci mentre que els segons exerceixen la seva funció en uns 80 addicionals a més de molts dels 100 primers. Aquests assaigs, però, tenen una resolució molt baixa i podria ser que en cada un dels loci hi hagués més d'una regió reconeguda pels gens PcG i trxG. Estudis funcionals han identificat "Elements de Resposta a PcG o trxG" (sovint s'anomenen PREs o TREs indistintament) en diferents loci, incloent els complexos bithorax (Bx-C) i Antennapedia (Ant-C; revisat a Ringrose, 2004). La caracterització d'aquests elements revela que hi ha un enriquiment de motius reconeguts per Pho i PhoL, GAF i Psq o Zeste i que acostuma a haver-hi més d'un PRE a la mateixa regió, cosa que podria crear diferents llocs d'unió pels gens PcG i trxG per tal que s'unissin formant un gran complex tot fent una baga (*loop*) a la cromatina (Figura 17-C). Alternativament, la diversitat de PREs podria

significar que cadascun tingués un rol en la regulació dels gens d'una zona de cromatina.

Ringrose i Paro (Ringrose et al., 2003) es van basar en aquests estudis per desenvolupar un algoritme que va predir, *in silico*, 167 PREs en el genoma de *Drosophila*. En el cas dels PREs amb millor puntuació, la majoria en tenien algun altre més a prop, i en molts casos un estava molt proper (<800 bp) o en el mateix promotor del gen putativament regulat. Per confirmar la validesa de l'anàlisi, uns quants d'aquests PREs van ser testats experimentalment. La comparació de la localització dels 167 PREs amb les tincions de cromosomes politènics va evidenciar que com a mínim la meitat dels PREs del genoma no van ser detectats. Aquesta dada posa de manifest la complexitat dels PREs i el desconeixement que encara es té dels mateixos. En aquest sentit, el fet que recentment s'hagin trobat en els PREs seqüències reconegudes per Dorsal Switch Protein 1 (DSP1; Dejardin et al., 2005) i Grainyhead (Grh; Blastyak et al., 2006) representa un pas endavant que podria propiciar una millor detecció dels PREs. Tanmateix, el treball de Ringrose i Paro, revela que els gens regulats per PREs s'engloben en multitud de classes diferents, cosa que indica que els gens PcG i trxG estan involucrats en la regulació de múltiples processos. S'inclouen, per exemple, gens de segmentació, gens involucrats en el desenvolupament de l'ull o de l'ovari, supressors de tumors i factors de transcripció semblants a p53.

b) Detecció de l'estat de la cromatina

Un cop localitzats els gens sobre els quals han d'actuar, els PcG i trxG han de detectar si el gen està activat o reprimat. Alguns autors postulen que en els PREs es reclutarien complexos silenciadors per defecte, ja que s'ha observat mitjançant mosques transgèniques que si es col·loquen PREs davant dels promotors i estimuladors (*enhancers*) d'un gen marcador, aquest no s'expressa. Aquest silenciament és mitjançat pels PcG a través dels PREs, ja que no s'observa en mutants per aquestes proteïnes ni si s'escindeix el PRE durant el desenvolupament mitjançant seqüències de recombinació (Sengupta et al., 2004). Tot i així, si els promotors dels gens estan activats, la presència de PREs no és suficient per silenciar-los (Poux et al., 2001a), i s'ha observat que ASH1 i Trx poden revertir l'estat silenciament (Klymenko and Muller, 2004). El fet que al llarg de l'embriogènesi s'observi un cert dinamisme de la unió dels complexos PcG i trxG als PREs, indicaria una unió transitòria d'aquestes proteïnes amb els PREs per detectar-ne l'estat (Orlando et al., 1998; Poux et al., 2001b). En aquest sentit, la interacció de Trx ocorre en el moment en el que s'està establint el patró d'expressió dels gens homeòtics mentre que la interacció de Pc és més tardana. Això encaixa en un model en el qual Trx detectaria d'alguna manera que el gen s'ha de mantenir activat i, a través de la seva activitat metiltransferasa d'histones (Smith et al., 2004), el marcaria amb una modificació típica de cromatina activa

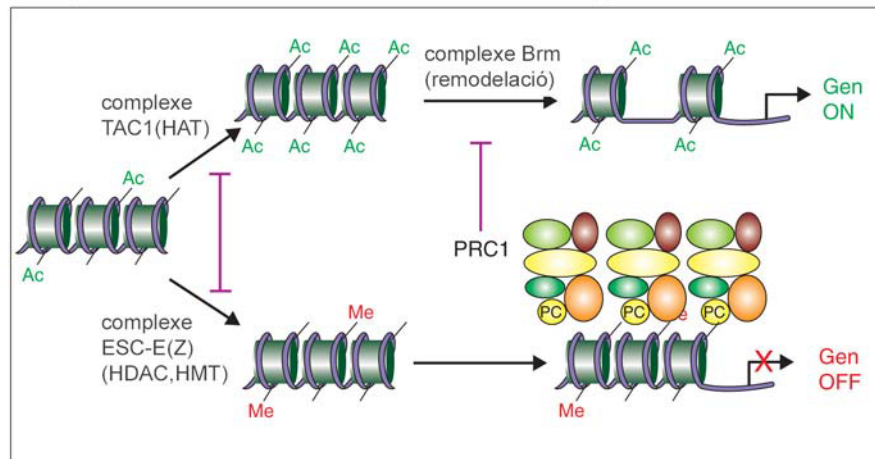
com H3K4. Aquesta marca evitaria que les proteïnes PcG el silenciessin, que seria el que hagués ocorregut per defecte (Ringrose and Paro, 2004).

La informació de que un promotor està activat podria venir donada per la transcripció dels propis TRE/PREs, ja que aquesta correlaciona amb la transcripció del gen que regulen i és necessària per mantenir un mecanisme d'anti-silenciament (Rank et al., 2002; Schmitt et al., 2005). El fet que la transcripció ectòpica dels TRE/PREs del BX-C en provoqui l'activació i l'aparició de transformacions homeòtiques donen suport a aquesta hipòtesi (Bender and Fitzgerald, 2002; Hogga and Karch, 2002). Més recentment s'ha demostrat que en el reclutament d'ASH1 a l'ADN hi juga un paper important la seva unió als tres TREs d'*Ubx*. Aquests es transcriuen juntament amb *Ubx* quan està activat i s'ha proposat que es quedarien hibridats amb l'ADN formant un híbrid estable ARN-ADN. Estructures secundàries formades per aquest ARN permetrien la unió d'ASH1, que no presenta motius coneguts d'unió a l'ADN (Sanchez-Elsner et al., 2006).

c) Manteniment de l'estat silenciament o activat

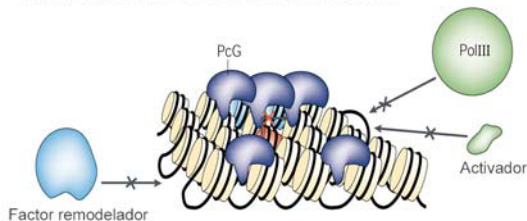
Hi ha diverses hipòtesis de com els PcG i trxG estableixen el silenciament o l'activació de la cromatina després de detectar-ne l'estat. En general no són excloents entre elles i podria ser que actuessin de forma cooperativa. Pel que fa als PcG, les evidències comentades anteriorment semblen indicar que els complexos ESC/E(Z) precedirien PRC1 tot facilitant la unió d'aquest a la cromatina. Ara bé, per explicar l'entrada en contacte entre el PRE i el promotor i el posterior silenciament de la cromatina s'han proposat mecanismes d'extensió (*spreading*) i de baga (*looping*) donat que els PREs del BX-C es troben molt allunyats dels promotors dels gens diana (Figura 17-B,C; revisat a Ringrose, 2004). En el primer dels casos el PRE seria un element nucleador a partir del qual es recobriria la cromatina fins al promotor per tal de que sigui inaccessible a d'altres factors. En el segon dels casos, el PRE i la zona promotora entrarien en contacte directe ja que es faria una baga (*loop*) a la cromatina per permetre-ho. Dades d'experiments d'immunoprecipitació posen en dubte el primer mecanisme i donen suport al segon, ja que ensenyen que tot i que en els PREs i els promotors s'hi troben força proteïnes PcG, en la cromatina situada entre aquests dos elements no n'hi ha tantes (Orlando et al., 1998; Strutt et al., 1997). A més, Pc i Ph tenen un temps de residència curt en els complexos (Ficz et al., 2005), els PcG no eviten que factors de transcripció generals s'uneixin als promotors de gens homeòtics (Breiling et al., 2001) i s'han trobat TAFs i TBPs en el PRC1 (Saurin et al., 2001). Així doncs, hi ha d'haver quelcom que eviti que els factors de transcripció presents en aquests gens portin a terme la seva funció. L'acció directa sobre la ARN polimerasa (Dellino et al., 2004), la compactació de la cromatina (Figura 17-B; Francis et al., 2004) i modificacions enzimàtiques com sumoilacions (Kagey et al., 2003) i ubiquitinitzacions (Wang et al., 2004a)

A) Model de funcionament dels complexos PcG i trxG

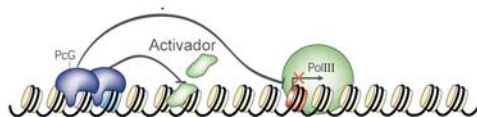


B) Mecanismes de repressió dels PcG

B1.- Estructura inaccessible de la cromatina

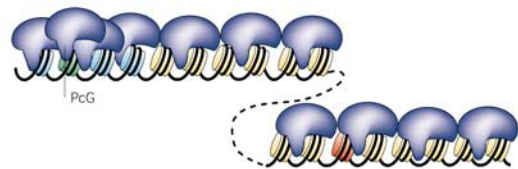


B2.- Inhibició directe de la maquinària transcripcional

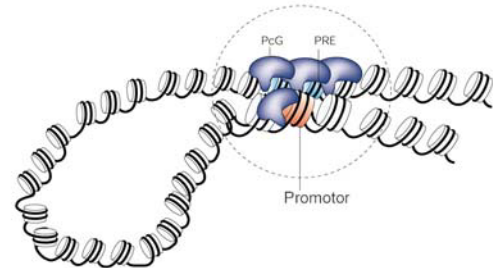


C) Model d'acció dels PcG

c1.- Mecanisme d'extensió (spreading)



c2.- Mecanisme de boga (looping)



c3.- Clusters de zones PRE per looping

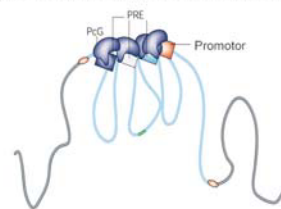


Figura 17 | Models de funcionament dels PcG i trxG. A) Model en el qual el complex TAC1 acetila (Ac) les histones dels nucleosomes, provocant un posterior reclutament del complex remodelador Brm, fet que propicia un manteniment de l'estat activat de l'expressió gènica. Per contra, el complex ESC-E(Z) crea un codi de metilacions (Me) que atrau al complex PRC1, el qual inhibeix la remodelació de la cromatina, facilitant així el manteniment de l'estat reprimat del gen. B) Esquemes de la repressió de la transcripció exercida pels PcG. Es crearia una estructura de la cromatina inaccessible per prevenir l'accés als factors de transcripció i factors remodeladors (B1), o bé els PcG interaccionarien de manera directa amb la maquinària transcripcional (B2). C) Models d'acció de les proteïnes PcG. Els complexos PcG s'estendrien (*coating; spreading*) per la cromatina des de les seqüències PRE fins al promotor (C1) o bé es formaria una boga (*looping*) que connectaria els complexos PcG situats als PRE amb la regió promotora del gen (C2). El tercer model implica interaccions entre els complexos PcG, que formarien *loops* a la cromatina i acostarien els PREs al promotor (C3). Els nucleosomes que estan en vermell representen la regió promotora. Adaptat de Francis, 2001 i Simon, 2002.

podrien ser claus en aquest punt, tot i que com s'ha vist anteriorment, les acetilacions i metilacions també hi jugarien un paper important.

Des del punt de vista dels *trxG*, i per tant de l'activació dels gens, les metilacions de les cues d'histones també serien importants. Estudis recents han posat de manifest que *Trx* metila H3K4 (Smith et al., 2004), que *ASH1* metila H3K4, H3K9 i H4K20 (Beisel et al., 2002; Byrd and Shearn, 2003) i que els seus dominis SET són essencials per desenvolupar aquestes activitats. Tot i així, el fet que tant la falta de funció de *Trx* com d'*ASH1* provoqui, en embrions i en discs imaginals, un silenciament de l'expressió d'*Ubx* deguda a l'acció dels PcG, indica que ni *Trx* pot fer la funció específica d'*ASH1* ni aquesta la de *Trx*. A més, en dobles mutants PcG/*trxG* es dona una expressió anòmala d'*Ubx*, cosa que suggereix que si bé *ASH1* i *Trx* són necessaris per mantenir l'estat activat, no ho són per provocar l'activació de la transcripció. Aquestes dades apuntarien a que *ASH1* i *Trx* serien necessaris durant tot el desenvolupament per contrarestar l'efecte basal de silenciament provocat pels PcG (Klymenko and Muller, 2004). D'altra banda, tot i que co-localitzen només en alguns loci de cromosomes politènics (Chinwalla et al., 1995; Tripoulas et al., 1996) i formen part de complexos diferents a l'embrió (Papoulas et al., 1998), en alguns gens podrien estar col·laborant ja que les dues co-immunoprecipiten en extractes d'embrió i la unió de *Trx* a cromosomes politènics de mutants d'*ash1* està disminuïda (Kuzin et al., 1994; Rozovskaia et al., 1999).

A banda de *Trx* i *ASH1*, *Zeste* i el complexe *Brm* també s'han trobat associats a PREs o a promotors de gens actius (Beisel et al., 2002; Dejardin and Cavalli, 2004). Tot i així, el fet de que *Brm* es trobi present en més loci de cromosomes politènics que no pas les altres tres, indicaria que desenvolupa un paper més ampli en l'activació transcripcional. Hi ha dades que revelen que en alguns casos *Zeste* seria el primer a unir-se a cromatina activa, cosa que implicaria el reclutament del complexe *Brm* (Dejardin and Cavalli, 2004; Kal et al., 2000). En d'altres, *ASH1* podria establir un patró de metilacions que facilités la unió dels complexos *Brm* (Beisel et al., 2002). Una altra possibilitat emana de l'observació que l'estat actiu ve acompanyat per una acetilació que du a terme la histona acetiltransferasa *dCBP*, una de les proteïnes aïllades amb el complexe *TAC1* (Cavalli and Paro, 1999; Petruk et al., 2001) i *ASH1* (Bantignies et al., 2000; Petruk et al., 2001). Aquesta modificació d'histones podria obrir la cromatina o servir de senyal per reclutar altres proteïnes com *Brm*, que la reconeixerien mitjançant el seu bromodomini (Hassan et al., 2002). Finalment, no es pot descartar l'involucrament d'altres mecanismes en el manteniment de l'estat activat. Uns exemples en serien l'elongació de la transcripció mitjançada pel complexe *FACT* o la transcripció de seqüències no codificants a partir dels PREs, ambdós comentats anteriorment.

d) Transmissió de la informació a través de la mitosi

Els mecanismes descrits poden explicar en part com es manté activa o silenciada la cromatina en un moment donat però, com s'estableix la memòria que permet reproduir aquest mecanisme després de la divisió cel·lular? El fet que s'hagin trobat quantitats ínfimes de proteïnes PcG associades a la cromatina mitòtica (Buchenau et al., 1998; Dietzel et al., 1999; Miyagishima et al., 2003), ha donat lloc a diferents idees de com es pot donar aquest procés (revisat a Brock, 2005). Una de les més acceptades actualment proposa que la metilació de les lisines de les cues de les histones seria la marca que es propagaria a través de la replicació semi-conservativa de l'ADN, permetent recordar l'estat de la cromatina després de la mitosi. Aquesta opció sembla plausible ja que aquesta modificació és molt estable, no es coneix cap histona desmetilasa (revisat a Lachner, 2003) i, com s'ha comentat anteriorment, diverses proteïnes PcG i trxG disposen d'activitat histona metiltransferasa (HMTasa).

Hem vist doncs, algunes de les idees actuals de com les proteïnes PcG i trxG mantenen la identitat cel·lular formant complexos dinàmics en el temps i en l'espai per tal de mantenir la cromatina silenciada o activada. Tal i com ha quedat de manifest, els gens PcG han estat molt més estudiats que no pas els trxG. Dintre d'aquest grup, gens que ja van ser identificats fa més de trenta anys com *ash1* o *ash2* han estat objecte de pocs estudis per la comunitat científica. Tot i així, les noves tècniques moleculars aparegudes recentment han provocat una explosió d'estudis en aquest camp i diversos grups han intentat explicar el funcionament tant genètic com molecular d'aquests dos gens.

3 - El gen *ash2*

Un treball del 1971 va descriure 63 mutants que eren letals entre els estadis larvals i l'adult i que tenien els discs imaginals afectats (Shearn et al., 1971). Els discs eren disseccionats de les larves originals i injectats en larves *wt*. Deu dies després, els adults que havien nascut eren disseccionats en busca dels implants, que eren examinats per veure si havien desenvolupat estructures adultes. Tres dels mutants aïllats presentaven transformacions homeòtiques, bàsicament d'halteri a ala, d'antena a pota i de genitals a pota i antena.

3.1 - Anàlisi genètica dels gens *ash2* i *ash1*

En el 1974 es van mapar aquestes mutacions en 2 loci diferents (Shearn and Garen, 1974), i en el 1987 Shearn i col·laboradors van aïllar nous mutants d'aquests dos loci que, degut als efectes que produeixen les seves mutacions, van anomenar *ash1* (*absent, small or homeotic discs 1*) i *ash2* (*absent, small or homeotic discs 2*; Shearn et al., 1987). Van detectar que els dos loci eren essencials per la viabilitat larval i que, tot i que les transformacions homeòtiques que presentaven les mutacions dels dos loci eren semblants, hi havia algunes diferències. Alguns al·lels d'*ash1* en homozigosi, per exemple, tenien transformacions exclusives de posterior a anterior de l'ala i del segon parell de potes. D'altra banda, mentre que les transformacions d'antena a pota eren molt freqüents en al·lels d'*ash2* en homozigosi, en al·lels d'*ash1* ho eren molt poc (veure Taula I3). Això podria ser degut a que els al·lels estudiats fins al moment no fossin totalment nuls, permetent uns certs nivells de transcripció del gen. Van

Transformació	<i>ash1</i>	<i>ash2</i>
Proboscis a pota i/o antena	L	L
Antena a pota	L	F
Protórax dorsal a ala	L	L
Pota 1 a pota 2	F	L
Ala posterior a ala anterior	M	
Pota 2 posterior a pota 2 anterior	L	
Halteri a ala	F	F
Pota 3 a pota 2	F	L
Genitals a pota i/o antena	L	F

Llegenda: L - penetrància < 0.33
M - 0.33 < penetrància < 0.66
F - penetrància > 0.66

**Taula I3 | Transformacions homeòtiques dels mutants d'*ash2* i *ash1*.
(Segons Shearn, 1989)**

produir clons de cèl·lules mutants en els discs imaginals i van determinar que les transformacions eren autònomes cel·lulars (*cell autonomous*) ja que es trobaven només en les cèl·lules mutants. A més, van proposar també una relació funcional entre els dos loci, ja que els dobles mutants per *ash2* i *ash1* no presentaven discs imaginals en molts casos i tenien una viabilitat extremadament reduïda, morint inclús en el primer estadi larval. Els discs dels dobles mutants que van poder aïllar i injectar desenvolupaven defectes més forts que no pas els mutants senzills, posant de manifest una interacció entre els mateixos. Curiosament, i al contrari del que passa amb molts al·lells PcG, no van detectar alteracions del patró de segmentació de les larves ni en mutants senzills ni en dobles (Shearn et al., 1987). El fet de que mutacions del gen *trithorax* (*trx*) es comportessin de la mateixa manera en aquest aspecte i presentessin transformacions homeòtiques (revisat a Ingham, 1998) similars a les d'*ash1* i *ash2*, els va portar a agrupar aquests gens en l'anomenat "Grup de gens trithorax" (*trxG*). En el 1989, Allen Shearn va publicar que aquests tres estaven actuant en processos similars (Shearn, 1989). Les dades més rellevants foren que els transheterozigots dobles per alguns al·lells d'*ash1*, *ash2* i *trx* potenciaven els fenotips dels al·lells en heterozigosi simple i/o en donaven de nous, i que els al·lells dels tres gens podien suprimir totalment o parcial el fenotip de pintes tarsals extres (*extra sex combs*) en mascles adults provocat per una mutació de Pc en heterozigosi.

Uns anys més tard LaJeunesse i Shearn van suggerir que, tot i estar involucrats en processos semblants, *ash1* i *ash2* podrien no formar part d'un mateix complexe o compartir els mecanismes d'acció ja que al·lells mutants per aquests dos gens no provocaven exactament els mateixos canvis d'expressió de gens homeòtics involucrats en la identitat del tòrax adult (LaJeunesse and Shearn, 1995). Mentre que al·lells mutants d'*ash2* i d'*ash1* afectaven d'alguna manera l'expressió de *sex combs reduced* (*scr*) i *Ubx*, només els mutants d'*ash1* tenien conseqüències sobre l'expressió d'*Antp* o *en*. Els autors van proposar que *ash1* podria tenir un rol més general en la regulació de la transcripció de gens Hox mentre que *ash2* en podria tenir un de més específic tant a nivell de gens com de teixit. L'ala, per exemple, és una de les estructures que necessita d'*ash2* per desenvolupar-se correctament tal i com s'ha demostrat en dos estudis publicats amb dos al·lells aïllats en el nostre laboratori: *ash2*¹¹²⁴¹¹ i *ash2*¹ (Amoros et al., 2002; Angulo et al., 2004).

3.2 - *ash2* és necessari pel desenvolupament de l'ala: els al·lels *ash2*¹¹²⁴¹¹ i *ash2*¹

L'al·lel *ash2*¹¹²⁴¹¹ va ser aïllat a partir del cribratge d'una col·lecció de mutants, generada per la inserció d'un element *P-lacW* en el cromosoma 3 (Deak et al., 1997), en el qual es buscaven fenotips anòmals dels discs imaginals d'ala, pota i/o halteri. Tot i la letalitat de la mutació en heterozigosi, un 12% dels individus arribaven a la fase adulta i el fenotip més visible i penetrant (95%) era l'aparició d'entre una i cinc venes transversals ectòpiques en qualssevol dels compartiments d'intervena excepte en el comprès entre les venes longitudinals L3 i L4. Aquest al·lel presentava una única inserció de l'element *P-lacW* a l'intró 4 del gen *ash2* i aquest element era el causant del fenotip observat ja que la seva escissió precisa revertia el fenotip a *wt*. L'al·lel *ash2*¹¹²⁴¹¹ és recessiu i presenta transformacions homeòtiques, tal i com s'espera per un gen *trxG*. Entre elles cal destacar transformacions d'halteri a ala a una molt baixa freqüència i una reducció total o parcial de les pintes tarsals del primer parell de potes dels mascles, indicant una transformació del primer segment toràctic (T1) al segon (T2). Els fenotips detectats amb diversa penetrància a l'ala inclouen, a més de l'aparició de venes transversals ectòpiques, l'engruiximent de les venes L3 i L5, el traçat irregular de la vena L2, la pèrdua parcial del marge posterior de l'ala i la reducció de la mida total de l'ala, essent aquesta reducció més pronunciada en el compartiment posterior que en l'anterior (Figura I8).

A partir de l'escissió imprecisa de l'element *P-lacW* de la línia *ash2*¹¹²⁴¹¹, es va obtenir un al·lel més fort que es va anomenar *ash2*¹ (Amoros et al., 2002). Com veurem més endavant, la caracterització molecular d'aquest al·lel ha estat objecte d'estudi d'aquesta Tesi. Pel que fa a la viabilitat, *ash2*¹ és letal en homozigosi entre els estadis de larva 3 tardana i pupa primerenca i la combinació



Figura I8 | Alteracions de formació del patró a les ales de mutants d'*ash2*. Ales *wt* (esquerra), *ash2*¹¹²⁴¹¹/*ash2*¹¹²⁴¹¹ (centre) i amb clons de cèl·lules homozigotes per *ash2*¹ (dreta) delimitats per una línia blava. Ales no a escala. A l'ala *wt* s'enumeren les venes longitudinals, les transversals anterior (acv) i posterior (pcv) i els territoris d'intervena (A-E). Imatge adaptada d'Angulo et al., 2004.

heteroal·lèlica *ash2¹/ash2¹¹²⁴¹¹* presenta només un 0.37% d'individus viables amb defectes similars als descrits (Amoros, 2001; Amoros et al., 2002). Mosaics genètics generats per recombinació mitòtica amb l'al·lel *ash2¹* sobre un fons *Minute*, van permetre descriure alteracions en els clons de cèl·lules homozigotes per la mutació. Entre aquestes cal destacar la reducció del territori d'intervena i l'ampliament del de vena, tant de venes transversals com longitudinals, amb l'excepció de la vena L4 (Figura I8). Aquesta observació i l'anterior respecte a l'absència de venes transversals a la regió d'intervena entre L3 i L4, va fer proposar que *ash2* no era necessari pel desenvolupament d'aquesta regió (Amoros et al., 2002). En un estudi més detallat de la implicació d'ASH2 a l'ala de *Drosophila*, es va demostrar que el gen *ash2* és necessari per mantenir els nivells de *net* i Blistered (Bs), essencials pel desenvolupament de les regions d'intervena. D'altra banda, interaccions gèniques van posar de manifest que la via del Receptor de l'Egf (EGFR), implicada en l'especificació de venes, es trobava sobre-activada en els mutants d'*ash2* (Angulo et al., 2004). A més a més, l'expressió de *rhomboid* (*rho*; Diaz-Benjumea and Hafen, 1994), un activador de la via de l'EGFR, estava eixamplada en les zones de vena en ales pupals homozigotes per *ash2¹¹²⁴¹¹*. Donat que *net* funciona com antagonista de *rho* (Brentrup et al., 2000), la disminució del primer deguda a la manca d'ASH2 podria explicar l'augment en l'activitat de la via de l'EGFR (Angulo et al., 2004).

3.3 - Caracterització molecular del gen *ash2*

Adamson i Shearn (1996), van realitzar unes primeres aproximacions moleculars per caracteritzar aquest locus. Van situar el gen en el braç dret del cromosoma 3 (3R) i van descriure un transcrit de 2 kb que donaria lloc a una proteïna de 572 aminoàcids. La seqüenciació del genoma de *Drosophila* (Adams et al., 2000) va permetre situar el gen amb exactitud a la banda 96A13 i, com veurem més endavant, els treballs d'aquesta Tesi van portar a la descripció d'un transcrit addicional més petit fruit d'un *splicing* diferencial.

El pes molecular de la proteïna ASH2 resultant del transcrit de 2 kb seria de 64.8 kDa però diferents treballs que han utilitzat el mateix anticòs (Adamson and Shearn, 1996) han identificat bandes corresponents a ASH2 de diferent pes molecular. En un primer moment es van detectar dues proteïnes, una de 70 kDa en estadis larvals i una de 53 kDa en estadis pupals. Es va proposar que la forma gran podria ser deguda a modificacions post-traduccionals o a una incorrecta migració de la proteïna en el gel d'acrilamida, mentre que la forma petita podria venir donada per un truncament (Adamson and Shearn, 1996). Més recentment, en extractes d'embrions s'ha detectat una banda de 94 kDa i una de 48 kDa que també es detecta des de larves a adults juntament amb una de 65 kDa (Cheng and Shearn, 2004; Papoulas et al., 1998).

En tot cas, la distribució d'ASH2 és homogènia en els discs imaginals i en els nuclis de les glàndules salivals i de les cèl·lules del cos adipós (Adamson and Shearn, 1996). Aquesta proteïna tindria un Motiu de Localització Nuclear Bipartit (Bipartite Nuclear Localization Motif or Signal: BNLS), un domini PHD, una seqüència PEST, (Adamson and Shearn, 1996) i un domini SPRY (Figura I9; Ponting et al., 1997). La descripció dels dominis d'ASH2 podria ajudar a entendre millor el seu funcionament, però malauradament no se'n coneix gairebé res dels mateixos. El domini BNLS és típic de les proteïnes que són transportades a nucli, cosa essencial per una proteïna reguladora de la transcripció com ASH2. El domini PHD (Plant HomeoDomain), per la seva banda, és un doble dit de zinc ric en cisteïnes que presenta una histidina conservada i residus aromàtics. Aquest domini forma part, des de llevats a humans, de diverses proteïnes reguladores de la transcripció, entre les quals cal destacar PcL entre els PcG i Trx i ASH1 entre els trxG (Aasland et al., 1995). En altres proteïnes que contenen PHD, s'ha demostrat que aquest domini està involucrat en interaccions entre proteïnes i que podria estar-ho en el reconeixement de modificacions de les cues d'histones, ja que pot unir-se a nucleosomes (Aasland et al., 1995; Bienz, 2006; Ragvin et al., 2004). Per la seva banda, les seqüències PEST contenen regions riques en prolina (P), àcid glutàmic (E), serina (S) i treonina (T), i són característiques de les proteïnes de vida curta (Rogers et al., 1986). Finalment, el domini SPRY es va identificar per primera vegada en la SP1A quinasa de *Dictyostellum* i en el receptor RYanodine de conill, té uns 140 aminoàcids i s'ha proposat que podria tenir un paper en les interaccions proteïna-proteïna i ARN-proteïna (Ponting et al., 1997).

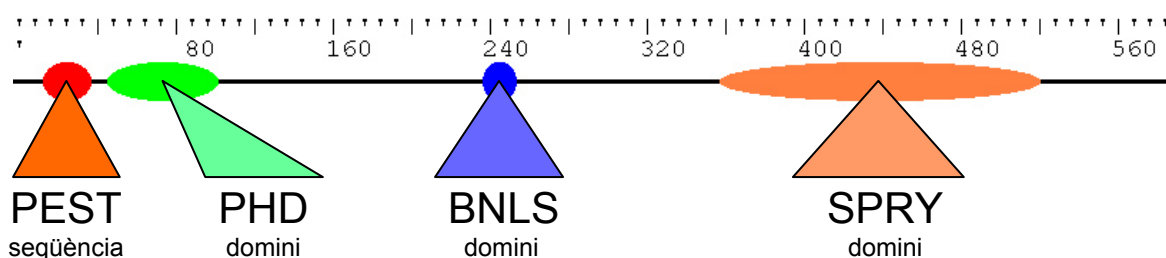


Figura I9 | Esquema de la proteïna ASH2 i els seus dominis. L'escala superior indica el número d'aminoàcid de la seqüència proteica (total de 572 aminoàcids).

3.4 - Els complexos d'ASH2 a *Drosophila*

Tot i que el coneixement dels dominis integrats a ASH2 és important per donar indicacions del seu funcionament, no ho és menys saber amb quines proteïnes es troba associada, ja que cap dels dominis descrits fa pensar que es pugui unir per ella mateixa a cromatina. Malauradament, la bibliografia de *Drosophila* que tracta aquest tema en particular és ben escassa. D'una banda se sap que ASH2 no formaria associacions estables amb Brm i que es trobaria com a mínim, en embrions, en un complex de 500 kDa (Papoulas et al., 1998). D'una altra, s'ha descrit que tant *in vitro* com *in vivo* la forma de 48 kDa (i la de 65 kDa *in vitro*) però no la de 94 kDa s'uneix amb Skittles (SKTL), una proteïna amb putativa activitat fosfatidilinositol 4-fosfat 5-quinasa (PIP5K; Cheng and Shearn, 2004). Aquests enzims fosforilen el fosfatidilinositol 4-fosfat per transformar-lo en fosfatidilinositol 4,5 bifosfat (PIP2). Segons estudis amb mutants sensibles a temperatura, ASH2 necessita unir-se als cromosomes per tal de fer la seva funció i SKTL requereix ASH2 per unir-se a cromosomes politènics però no a la inversa (Cheng and Shearn, 2004). El fet que mutants de *sktl* potenciïn les transformacions homeòtiques produïdes pels d'*ash2*, i que l'eucromatina dels dobles mutants *ash2-sktl* presentin una hiperfosforilació de la histona H1, implica que PIP2 podria jugar algun paper en el manteniment de la cromatina transcripcionalment activa (Cheng and Shearn, 2004). Finalment, el fet que els cromosomes politènics de mutants d'*ash2* no presentin ni una reducció en la dimetilació de H3K4 ni acumulació de la histona metiltransferasa ASH1, però si a la inversa, fa postular que ASH2 no estaria formant part del complexe metiltransferasa d'ASH1 però que el necessitaria per tal d'unir-se a l'ADN i fer la seva funció (Byrd and Shearn, 2003). Sortosament, tant en llevats com en mamífers s'han purificat diversos complexos que presenten els homòlegs d'ASH2. Donat que els complexos de Brm, contenen subunitats conservades en els complexos homòlegs de llevats i humans (Mohrmann et al., 2004; Papoulas et al., 1998), els complexos d'ASH2 a *Drosophila* podrien ser també molt similars als trobats en llevat i humans.

3.5 - ASH2 a llevats i mamífers

A *Saccharomyces cerevisiae* s'han trobat dues proteïnes amb homologia a ASH2 associades a diferents variants d'un complexe que s'ha anomenat Set1C (Nagy et al., 2002; Roguev et al., 2001) i COMPASS (Miller et al., 2001) ja que conté una proteïna amb domini SET, Set1, i pot metilar específicament H3K4 (Nagy et al., 2002; Roguev et al., 2001). Tot i que Set1 presenta homologia amb Trx, el seu equivalent a *Drosophila* seria CG17396 (Roguev et al., 2001). Una de les dues proteïnes amb homologia a ASH2 és Bre2/Cps60 que presenta, com ASH2, un

domini SPRY. Cal destacar però, que la homologia entre elles va més enllà d'aquest domini (Nagy et al., 2002; Roguev et al., 2001). L'altra proteïna és Spp1/Saf41p/Cps40 i presenta un domini PHD, cosa que ha fet postular que el complexe Set1C incorpora un anàleg d'ASH2 mitjançant la unió de Bre2 i Spp1 (Nagy et al., 2002; Roguev et al., 2001). Respecte a la funció específica d'aquestes dues proteïnes, se sap que no són essencials per mono- o di-metil·lar H3K4 però sí per tri-metil·lar-la (Schneider et al., 2005), un estat associat amb gens transcripcionalment actius (Santos-Rosa et al., 2002), donant suport a una funció moduladora d'ASH2. Vistes aquestes dades, alguns autors van predir que el complexe d'ASH2 en eucariotes més evolucionats inclouria un homòleg de Set1 i que tindria activitat H3K4 metiltransferasa (Roguev et al., 2001). I efectivament, des que l'ASH2 humà va ser clonat i caracteritzat juntament amb el de ratolí (Ikegawa et al., 1999), s'ha trobat en altres complexos amb activitat HMTasa. Un d'ells és anàleg al COMPASS de llevat i inclou ASH2, Set1 i un homòleg de Spp1 (CFP1). Les dues segones co-localitzen en zones del nucli riques en eucromatina (*nuclear speckles*), i el fet de que hi hagi CFP1, que té dos dominis PHD, a banda d'ASH2 indicaria que aporten funcions diferents (Lee and Skalnik, 2005). Un altre complexe, anomenat ASCOM, conté més proteïnes trxG (dues formes de MLL2 i una de MLL3) amb domini SET i l'Activating Signal Cointegrator 2 (ASC-2), un co-activador transcripcional sobre-expressat en càncers humans (Goo et al., 2003). MLL2, ASH2 i altres homòlegs del Set1C de llevat es troben també en un complexe que conté el supressor de tumors menin (MEN1; Hughes et al., 2004), que s'ha trobat associat a un altre complexe relacionat amb Set1C que inclou MLL i Host Cell Factor 1 (HCF-1; Yokoyama et al., 2004). Precisament, HCF-1 pot unir, a més d'un complexe HMTasa que inclou Set1 i ASH2, un altre que conté Sin3A i les desacetilases d'histones HDAC1 i HDAC2 (Figural10; Wysocka et al., 2003). És sorprenent trobar junts dos complexos amb activitats que tenen conseqüències generalment oposades sobre la transcripció, però sembla ser que no és excepcional ja que Sin3A, HDAC1 i HDAC2 també s'han trobat en un supercomplexe que conté MLL i BRM (Nakamura et al., 2002). La diversitat de complexos en els quals s'ha trobat ASH2 podria ser deguda o bé a limitacions de les tècniques de descripció dels mateixos o bé a que aquesta proteïna estigui involucrada en la regulació transcripcional de molts tipus de gens diferents en funció del teixit, el moment i de les subunitats que l'acompanyin en el complexe.

Finalment, cal destacar que el que s'hagi trobat ASH2 associat amb el supressor de tumors MEN1 i el proto-oncogen MLL és un indicador de que pot tenir algun paper en l'aparició de càncers i leucèmies. De fet, després d'observar una baixada dels nivells d'ASH2 en línies cel·lulars de leucèmia i en la diferenciació megacariocítica induïda, s'ha hipotetitzat que ASH2 podria jugar algun paper en l'hematopoesi i que podria estar involucrada en leucèmies (Wang et al., 2001). Nous experiments seran necessaris per desxifrar quin tipus de relació hi ha entre ASH2 i aquestes malalties.

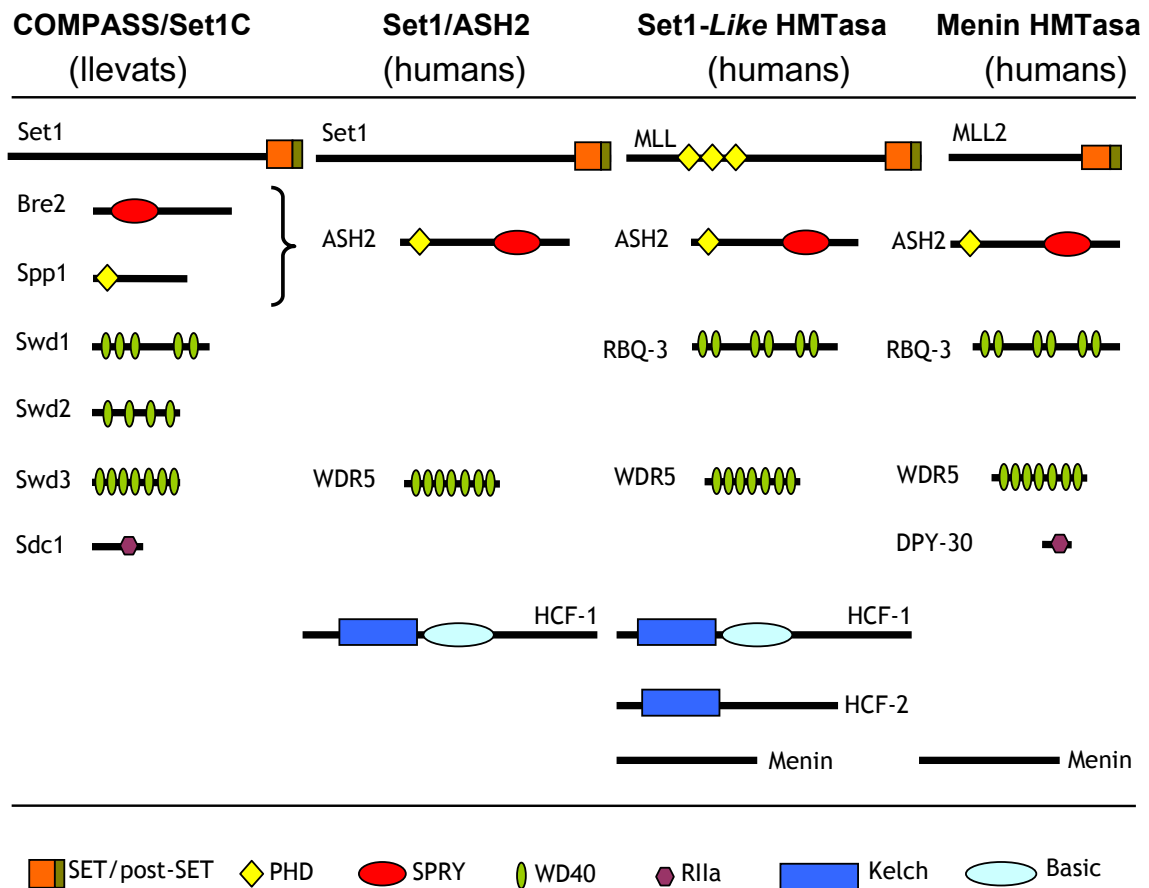


Figura I10 | Complexes d'ASH2 a llevats i humans. Es mostren alguns dels components dels complexos COMPASS/Set1C de llevats (Miller et al., 2001; Nagy et al., 2002; Roguev et al., 2001) i els seus homòlegs en els complexos Set1/ASH2 HMTasa (Wysocka et al., 2003), Set1-Like HMTasa (Yokoyama et al., 2004), i Menin HMTasa (Hughes et al., 2004) d'humans. En llevats, Bre2 i Spp1 podrien formar una associació funcional equivalent a ASH2. A sota de la figura s'indica quin domini representa cada símbol de color.

Vist que la identitat cel·lular ve donada pel patró d'expressió gènica de la cèl·lula i que en aquest procés hi estan involucrats els gens PcG i trxG, queda clar que una bona aproximació experimental per ampliar el coneixement en aquest camp és estudiar els transcriptomes (conjunts de gens expressats o reprimits) de mutants per aquests gens comparant-los amb la condició *wt*. Per tal de dur a terme aquest objectiu, ens hem servit d'una tècnica desenvolupada recentment que permet fer una "fotografia" del nivell d'expressió de tots els gens d'un organisme, teixit o cèl·lula en un moment donat: els microarrays d'ADN.

4 - La tecnologia dels microarrays

Segurament, la tecnologia que més ha revolucionat la Biologia en els últims anys és la dels microarrays. Des que es va publicar el primer treball on els utilitzaven (Schena et al., 1995), l'evolució i la implementació de la tècnica han crescut a ritmes trepidants, fins al punt que han aparegut multitud de plataformes específiques per una gran diversitat d'organismes. Els microarrays d'expressió gènica en concret, s'han utilitzat per estudiar una àmplia varietat de processos biològics, des de l'expressió diferencial de gens en l' esporulació de llevat (Chu et al., 1998) fins als tumors de mama humans (Perou et al., 2000). En el cas de *Drosophila*, es van utilitzar inicialment per estudiar la metamorfosi (White et al., 1999) i més recentment s'han emprat també en treballs que han tingut el discs imaginals com a protagonistes (Butler et al., 2003; Klebes et al., 2002; Klebes et al., 2005; Ren et al., 2005).

Aquesta tecnologia ha rebut diferents noms, els més comuns dels quals són microarrays d'ADN o cADN (DNA o cDNA microarrays) i xips de gens (GeneChips). Tot i que no hi ha definicions estàndard, a vegades es fan distincions entre les diferents nomenclatures per associar-les a variants de la tecnologia. Els noms donen per si mateixos una idea de quines són les premisses de la tècnica. Xip o micro fa referència a les dimensions reduïdes del suport utilitzat; *array* (matriu) indica que en aquest suport hi ha una disposició ordenada d'alguna cosa, que en aquest cas és ADN equivalent a gens, tal i com també se'n desprèn del nom. Els tres tipus de microarrays d'ADN que s'han utilitzat en aquesta Tesi són concretament seqüències d'ADN, que corresponen a gens, immobilitzades en una matriu ordenada sobre un suport de vidre o de plàstic de la mida d'un portaobjectes (*slide*). Cada punt (*spot* o *feature*) de la matriu equival a un gen conegut de *Drosophila* i, mitjançant una aproximació basada en la hibridació tradicional com la del *Northern blot*, podem quantificar l'expressió absoluta o relativa del gen quan dipositem tota la població d'ARNs missatgers (mARNs) d'una mostra marcats amb un fluorocrom (Figura I11). Malgrat que la definició donada és vàlida pels tres tipus de microarrays d'ADN emprats aquí, aquests presenten algunes característiques diferencials en funció de com són les seqüències, de com s'han dipositat i immobilitzat i, no menys important, de com s'han marcat i hibridat les poblacions d'ARN, que en el context dels microarrays sol anomenar-se sonda marcada (*labelled probe*). Per tal de facilitar la descripció de les variants utilitzades en aquest Tesi, microarrays d'ADN farà referència als dos tipus no comercials i GeneChip al de la casa Affymetrix. La Figura I11 mostra un disseny experimental possible utilitzant microarrays d'ADN d'hibridació competitiva o de dos colors.

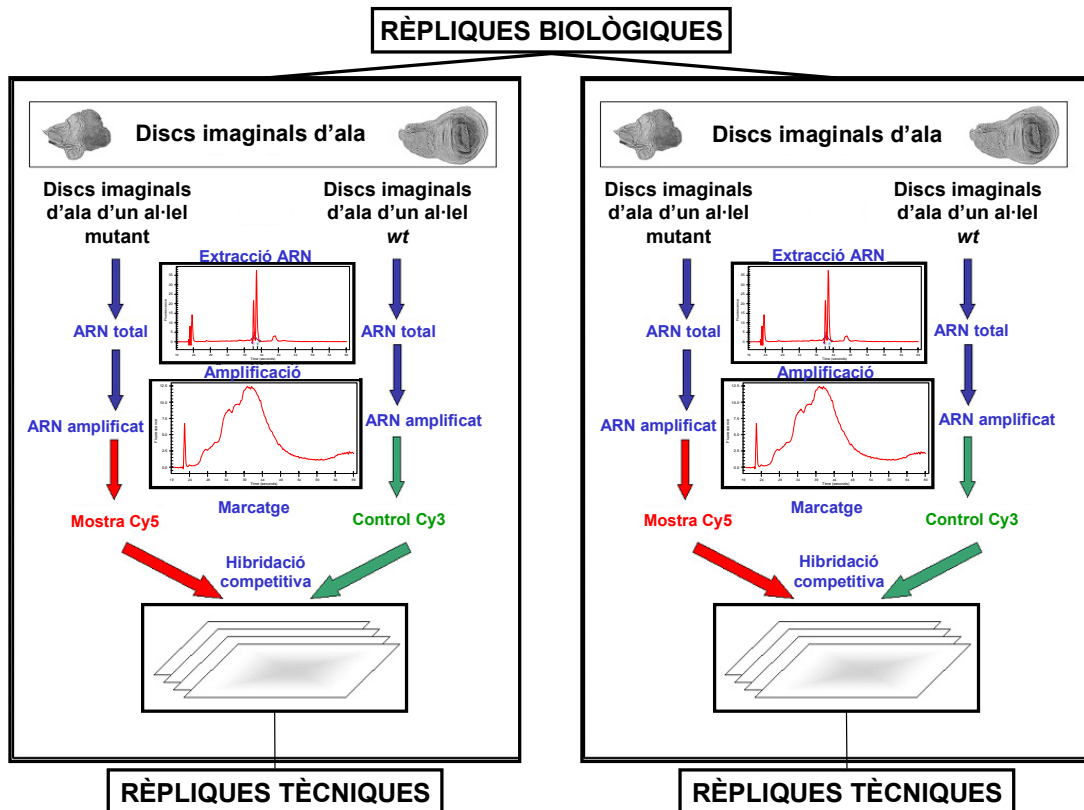


Figura I11 | Exemple d'un experiment de microarrays d'ADN on es compara el patró d'expressió gènica de discs imaginals d'una línia *wt* amb els d'una mutant. Es parteix de quatre extraccions d'ARN total, dues de discs *wt* i dues de discs mutants les quals serviran per fer rèpliques biològiques (veure text). La qualitat de l'ARN total es comprova amb l'Agilent 2100 Bioanalyzer, que produeix un electroferograma (els dos pics definits d'ARN ribosomal indiquen que no hi ha degradació). El pas marcat com amplificació es fa independentment per cadascuna de les quatre mostres i inclou de manera seqüencial: la retrotranscripció de l'ARN total a cADN, la degradació de l'ARN, la síntesi d'una cadena d'ADN complementària a la cADN (ccADN) i la síntesi a partir de la transcripció de la ccADN de grans quantitats d'ARN (ARN amplificat o aARN) amb una base modificada (aminoallyl-dUTP). Cada aARN es marca amb un fluorocrom diferent (Cy3 o Cy5) i es procedeix a la hibridació de dos microarrays, cadascun amb una mostra mutant (mostra) i una mostra *wt* (control). De cada rèplica biològica es poden fer diverses rèpliques tècniques, incloent *dye-swaps*.

4.1 - Tipus de microarrays

En el primer microarray d'ADN que vàrem utilitzar les seqüències eren dipositades mecànicament (impreses) a sobre del vidre mitjançant un braç robot. L'ADN provenia d'amplificar per PCR (*Polymerase Chain Reaction*) una llibreria de cADNs (*complementary DNAs*) coneguts clonats en plàsmids. Cada clon es troba en un pouet determinat d'una placa i, si tots els cADNs estan clonats en un mateix plàsmid, amb només un parell d'encebadors (*primers*) complementaris al plàsmid podem amplificar tota la llibreria. Els principals avantatges són que no necessitem

conèixer la seqüència del genoma de l'organisme, que les PCRs solen tenir èxit i que no hi ha amplificació inespecífica. Malauradament, els desavantatges són molts. La quantitat de gens que podem imprimir ve limitada per la llibreria de cADNs, les condicions de les PCRs han de ser diferents en funció de la mida del cADN, la quantitat de producte obtingut és variable, s'ha de monitoritzar amb gels d'agarosa totes les PCRs, i les T_m d'anillament són diferents en cada cas. Una variant d'aquesta aproximació permet deixar enrera alguns d'aquests desavantatges si disposem de la seqüència de tots els gens de l'organisme. En aquest cas, podem dissenyar un parell d'encebadors específics per cadascun dels gens per obtenir productes de PCR (*amplicons*) de mides i T_m similars a partir d'ADN genòmic. Tot i guanyar en alguns aspectes, hem de fer les PCRs i monitoritzar-les, podem obtenir productes inespecífics i s'han de sintetitzar grans quantitats d'encebadors.

Per totes aquestes raons, el segon tipus de microarray d'ADN que hem utilitzat és el que més ha proliferat en els últims anys. En aquest cas el material es diposita de la mateixa manera que en l'anterior, però l'ADN són oligonucleòtids sintètics d'entre 25 i 80 bases. A partir de la seqüència dels gens, es dissenya mitjançant un algoritme informàtic un oligonucleòtid per cada un d'ells, tot fent que compleixi unes condicions de mida, T_m i situació respecte l'extrem 3' del transcrit. Els avantatges d'aquesta aproximació són evidents, i a més permet produir microarrays de transcrit, de manera que si dissenyem un oligonucleòtid diferencial per cada un d'ells podem saber exactament quin dels transcrits d'un gen és el que varia en expressió.

El tercer tipus de microarray que hem utilitzat és el GeneChip d'Affymetrix. Probablement és la metodologia comercial més utilitzada, i presenta força diferències en relació als tipus descrits fins ara. Pel que respecta a les seqüències immobilitzades, aquestes són de 25 nucleòtids i són sintetitzades *in situ* mitjançant un procés que combina la fotolitografia i la química combinatòria. Cadascun dels punts del xip està subdividit per acomodar entre 16 i 20 oligonucleòtids diferents complementaris al gen. Aquests poden ser perfectament complementaris (*perfect match*) o variar en una de les posicions (*mismatch*), cosa que permet estimar la hibridació inespecífica.

L'altra gran diferència entre els dos primers tipus de microarrays i els xips d'Affymetrix és la manera de realitzar la hibridació i presentar els resultats. En els primers (Figura I11) es parteix de dues mostres diferents d'ARN total o missatger, que es retrotranscriuen a cADN i es marquen, directa o indirectament, cadascuna amb un fluorocrom diferent (generalment Cy3 i Cy5). Els cADNs marcats es barregen i s'hibriden conjuntament (hibridació competitiva) en un mateix microarray. L'excitació de cada fluorocrom es produeix mitjançant làsers a longituds d'ona diferents (532 i 635nm respectivament), i la intensitat de l'emissió en cada canal és directament proporcional a les quantitats relatives de cada ARN que hi havia a cadascuna de les mostres. Per tant, els resultats es presenten en

forma de quocients (*ratios*) de la fluorescència d'una mostra respecte l'altra per cadascun dels gens. En el cas dels GeneChips, cada portaobjectes s'hibrida només amb una mostra marcada generalment amb biotina-streptavidina, que és excitada a 570 nm. Els valors indiquen la quantitat absoluta de cada transcrit que hi havia a la mostra inicial. L'inconvenient és que per tal de comparar dues mostres s'han d'hibridar dos xips diferents, però això es converteix en un avantatge si el que es vol és comparar directament moltes mostres diferents.

4.2 - La variabilitat dels microarrays

La varietat de plataformes existents és un dels handicaps de la tecnologia dels microarrays, ja que és una immensa font de variabilitat i dificulta la comparació entre experiments. En aquest sentit, la creació d'un consorci internacional per produir un microarray estàndard per la comunitat científica de *Drosophila* és un primer pas per solucionar aquest problema. Aquesta organització, anomenada INDAC (INternational *Drosophila* Consortium; www.indac.org), ha dissenyat i sintetitzat una col·lecció d'oligonucleòtids que cobreixen tots els gens coneguts i predits del genoma. Cada laboratori integrant del consorci és lliure de fabricar els seus propis microarrays a partir d'aquesta col·lecció. Malauradament, les fonts de variabilitat dels microarrays inclouen molts altres aspectes, tant biològics com tècnics. Entre els primers són d'important rellevància les diferències de sexe, estadi, fons genètic (*genetic background*) entre les mostres o simplement les diferències entre individus. Entre els segons hi trobem, a tall d'exemple, els protocols d'extracció d'ARN, marcatge i hibridació de mostres, tipus de portaobjectes, fluorocroms, escàner utilitzat i el propi executor de l'experiment. Part d'aquesta variabilitat pot ser contrarestada agrupant material de diferents individus (*pooling*), fent un disseny experimental complex amb un bon anàlisi de dades i/o realitzant rèpliques biològiques i tècniques. Les més utilitzades entre les tècniques són les d'intercanvi de fluorocrom (*dye-swap*), en les quals s'hibriden dos microarrays per comparar una mostra i un control o referència: si en una de les hibridacions marquem la mostra amb Cy3 i la referència amb Cy5, en l'altra ho farem a la inversa (Figura 111). Això permet tenir dades replicades alhora que s'adreça el problema de l'esbiaix de fluorocroms, produït per les diferències en la incorporació dels mateixos i en les seves propietats.

Aquesta variabilitat intrínseca a la tècnica, ha despertat escepticisme i inquietud respecte a la reproducibilitat de resultats produïts en diferents laboratoris i/o utilitzant diferents plataformes i/o protocols, fins al punt que diferents estudis han detectat baixa correlació entre resultats de diferents laboratoris i/o plataformes (Kuo et al., 2002; Tan et al., 2003). D'altres treballs que han adreçat aquest problema a través d'estudis multi-laboratori i multi-plataforma han detectat bones correlacions i solapament (*overlap*) de resultats quan

s'utilitzen protocols i mètodes d'anàlisi estandaritzats (Bammler et al., 2005; Irizarry et al., 2005; Larkin et al., 2005; Yauk et al., 2004). Així doncs, tant o més important que l'adopció de bones pràctiques alhora de fer aquest tipus d'experiments, ho és el processament de les dades obtingudes. Donat que la tecnologia de la qual provenen en gran part els resultats d'aquesta Tesi és la dels microarrays d'ADN de "dos colors" (*two-color*) o d'hibridació competitiva, les consideracions que s'exposaran a continuació respecte a l'obtenció de dades i a l'anàlisi de les mateixes es restringeixen a aquest tipus de microarrays.

4.3 - Processament de dades

4.3.1 - Anàlisi de la imatge

Un cop s'ha hibridat un microarray, aquest s'ha d'escanejar per obtenir dues imatges independents, que corresponen a la senyal emesa per cadascun dels fluorocroms al excitar-los amb els làsers. Les imatges obtingudes han de ser convertides a informació numèrica vàlida per ser analitzada. Cada vegada hi ha més programes, tant comercials com gratuïts, que permeten fer aquest pas. Entre els primers el més utilitzat és el GenePix d'Axon, mentre que entre els segons és de destacar el Spottfinder del paquet d'anàlisi TM4 (Saeed et al., 2003). Tots dos funcionen d'una manera similar. En primer lloc es fa coincidir amb la imatge una graella que conté la informació de la situació dels punts en el microarray. Tot seguit s'executa un algoritme per tal que els cercles de la graella s'ajustin amb els punts del microarray, que solen ser més o menys circulars. Els primers algoritmes que van sortir ajustaven el cercle de la graella fent-lo simplement més petit o més gran, però recentment han sortit mètodes més elaborats que modifiquen el cercle per tal que ressegueixi el punt, aconseguint un ajust més exacte; la Figura I12 mostra un exemple dels dos casos. Un cop localitzats tots els punts es procedeix a extreure la informació de la imatge, obtenint de cada punt mesures com la intensitat de cada canal a dintre (senyal o *foreground*) i a fora del punt (soroll de fons o *background*), el número de píxels del punt, el percentatge de píxels saturats, la circularitat, etc. La intensitat d'un punt acostuma a calcular-se amb la mediana de la intensitat de tots els píxels, ja que aquesta mesura es veu menys afectada que la mitjana pels valors extrems d'algun píxel deguts, per exemple, a un gra de pols.

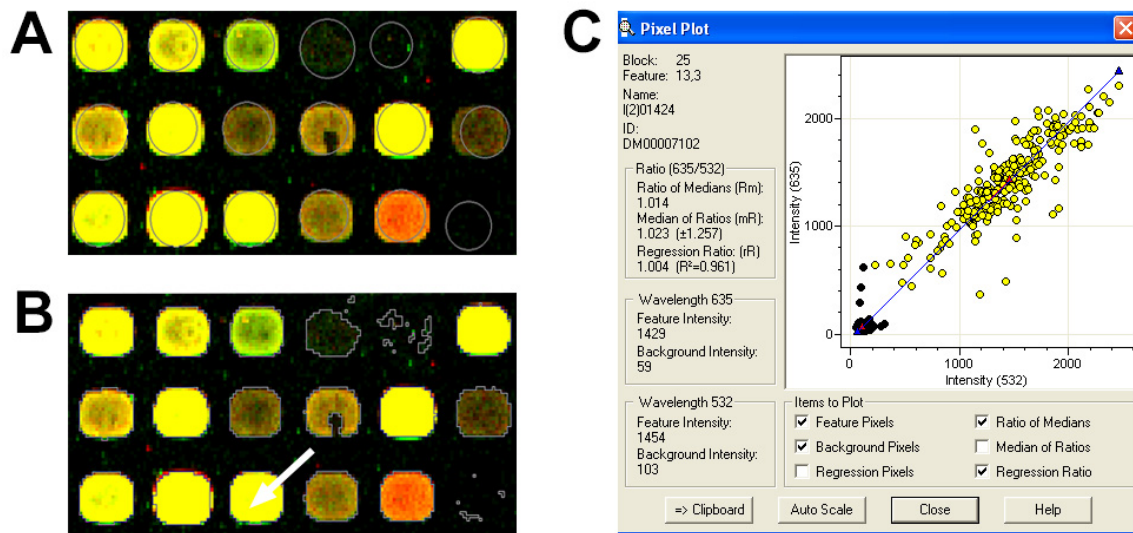


Figura 112 | Anàlisi de la imatge i filtratge de dades. Imatges obtingudes amb GenePix Pro 6.0 d'un microarray de *Drosophila* produït i hibridat a la Plataforma de Transcriptòmica SCT-PCB-UB. A) Punts detectats amb algoritme circular. B) Punts detectats amb algoritme irregular. Observar el millor ajust de la línia grisa al voltant dels punts. C) Panell corresponent al punt indicat amb una fletxa. Els punts grocs de la gràfica corresponen als píxels de la senyal, mentre que els negres corresponen al soroll de fons. Observar les diferències d'intensitat (gràfiques i numèriques) entre senyal i soroll, així com la similitud dels *ratios* calculats de diferents maneres. Aquest és un punt de bona qualitat.

4.3.2 - Filtratge de dades

Les mesures obtingudes per cada punt amb l'anàlisi de la imatge ens serveixen per filtrar els punts decidint quins són creïbles i quins no. Es pot ser més o menys astringent en aquest punt: si ho som massa perdrem molts punts que ens podrien donar una informació vàlida però si som massa laxes obtindrem molts falsos positius. El criteri més utilitzat és que el senyal del punt sigui com a mínim dues vegades més gran que el soroll de fons, però també s'empren altres filtres com que diferents maneres de calcular el quocient o *ratio* entre les dues intensitats del punt siguin similars. La Figura 112 mostra algunes de les mesures obtingudes per un punt.

4.3.3 - Transformació de dades

La transformació de dades és el pas necessari per obtenir dades analitzables tot eliminant la màxima variabilitat possible. Aquest és un dels punts que més atenció ha rebut i rep per part de matemàtics, estadístics i bioinformàtics i, tot i que s'han proposat multitud de mètodes, no n'hi ha cap que s'hagi erigit com a estàndard indiscutible. L'investigador pot avaluar els que cregui convenients i decidir-se pel que proporcioni resultats més fiables o creïbles en el seu cas. Per tal de poder fer una comparació amb objectivitat, són d'especial interès els controls, tant interns com externs. Els interns són aquells pels quals l'ARN de la mostra disposa de

transcrits; si distribuïm oligonucleòtids que els reconeguim per tot el microarray, podrem avaluar la reproducibilitat entre les rèpliques biològiques i estimar si la mesura de l'expressió es veu afectada per la situació del punt en el microarray. Els externs s'anomenen també controls *spike-in*, ja que són quantitats conegudes d'ARN exogen produït *in vitro* que s'afegeixen a la mostra perquè s'hibridin amb els seus punts complementaris dipositats al portaobjectes. Aquests punts poden també estar replicats al llarg del microarray i la mescla amb la mostra endògena es pot fer en diversos moments del protocol. Si els afegim al principi de tot, per exemple, podrem monitoritzar si tot el protocol de marcatge i hibridació ha funcionat bé. A més a més, si n'afegim de diversos i en diferents concentracions a la mostra i a la referència, podrem estudiar si els mètodes de transformació de dades ens donen els resultats esperats i conèixer quin és el rang linear d'intensitats que ens permet una detecció fiable dels valors d'expressió. A continuació s'expliquen breument, a tall d'exemple, cinc dels mètodes disponibles de transformació de dades. Les més usuals com la Normalització Global o la Lowess es troben implementades en la majoria de paquets comercials i gratuïts com GenePix i Acuity d'Axon, MIDAS de TM4 (Saeed et al., 2003) o Bioconductor (Gentleman et al., 2004). Els més innovadors acostumen a aparèixer primer en forma de paquet per a aplicacions de lliure distribució com Bioconductor, que poc a poc s'està convertint en una plataforma d'anàlisi de dades àmpliament utilitzada gràcies a la seva versatilitat i possibilitat d'analitzar experiments en massa. Aquesta versatilitat ve però acompanyada d'una dificultat afegida alhora d'utilitzar-lo, ja que manca d'interfície gràfica i funciona mitjançant línies de comandes.

a) Transformació logarítmica

Aquesta transformació és reversible i s'utilitza de manera generalitzada ja que facilita la visualització i el tractament de les dades. Com hem dit, el canvi en l'expressió d'un gen determinat vindrà donat pel quocient o *ratio* de la intensitat de la mostra (M) respecte a la referència (R). Si l'expressió del gen no està variada, M serà igual que R i el ratio M/R serà 1. En el cas que l'expressió del gen canviï, si $M > R$, el *ratio* M/R podrà estar comprès entre 1 i infinit, però si $M < R$, aleshores el *ratio* M/R només podrà tenir valors compresos entre 0 i 1, creant-se una asimetria entre els valors dels gens sobre- i sub-expressats. Una manera de solucionar el problema és que quan $M < R$ el *ratio* es calculi a la inversa (R/M) i es posi un signe negatiu davant del valor per entendre que està sub-expressat. Aquesta aproximació dificulta però les anàlisis matemàtiques ja que no hi ha valors compresos entre -1 i 1. Per aquestes raons, la transformació logarítmica en base 2 és globalment acceptada com la millor solució a aquest problema, en tant que transforma un gen que no varia (amb un *ratio*=1) en un $\log_2(\text{ratio})=0$ i tots els valors >0 correspondran a gens sobre-expressats i tots els valors <0 correspondran a gens sub-expressats. Per exemple, considerem dos gens (A i B) que presenten

un canvi d'expressió de dues vegades, però mentre que el A està dues vegades més present en la mostra que en la referència, el B ho està dues vegades menys.

	Gen A	Gen B
Valors d'intensitat	M=200 , R=100	M=100 , R=200
<i>Ratio</i>	M/R = 200/100= 2	M/R=100/200= 0.5
$\log_2(\text{ratio})$	$\log_2(M/R)=\log_2(2)=$ 1	$\log_2(M/R)=\log_2(0.5)=$ -1

b) Normalització Global

Abans de poder-nos creure els valors dels *ratios* hem de corregir les dades eliminant al màxim possible la variabilitat. Si partim de la premissa que entre la mostra i la referència que es comparen la gran majoria de gens no haurien d'estar canviats, i només hi haurà uns pocs que estaran sobre-regulats o sub-regulats, arribem a la conclusió que la suma de totes les intensitats de M ha de ser igual a la suma de totes les intensitats de R. Com que aquest no acostuma a ser el cas degut a la variabilitat introduïda pels fluorocroms i l'escaneig del microarray, podem calcular un factor de correcció (F) dividint la suma de totes les M entre la suma de totes les R. Per tal d'obtenir els *ratios* o quocients normalitzats, només hem de dividir M entre R corregit (M/RxF).

c) Normalització Lowess (o Loess)

Tot i ser vàlida, la normalització anterior no corregeix el fet que sovint els punts amb intensitat més baixes es troben més desviats del $\log_2(\text{ratio})=0$ esperat dels gens que no canvien que no pas les intensitats més altes. Una aproximació que afronta aquesta observació permetent eliminar els efectes de fluorocrom específics d'intensitat és la Lowess o Loess, basada en la regressió lineal compensada localment (*Locally weighted scatterplot smooth*; Cleveland, 1979; Yang et al., 2002). En aquest cas també s'assumeix que la majoria de gens no canviaran, i s'aplica un funció que calcula un factor de correcció per cada punt tenint més en compte (donant-hi més pes) els punts d'intensitats més similars, de manera que si el conjunt de punts amb menys intensitats es desvien globalment més del $\log_2(\text{ratio})=0$ esperat que els de més intensitats, seran més severament modificats. Tot i que Lowess i Loess són sovint termes que s'utilitzen indistintament, es diferencien en el model que utilitzen en la regressió.

Una modificació d'aquesta normalització, la Print-Tip Lowess, permet eliminar també la variabilitat deguda a l'agulla d'impressió (*print-tip*). Com que generalment cada bloc del microarray prové d'una agulla d'impressió diferent, si apliquem Lowess independentment per cada bloc, podrem eliminar aquest efecte. Per tal de que aquesta opció sigui vàlida, necessitem que en cada bloc hi hagi un número suficient de punts i que aquests estiguin distribuïts a l'atzar de manera

que en un bloc no hi hagi molts gens que han d'estar sobre-regulats i en un altre molts de sub-regulats.

d) Normalització OLIN

D'altres algorismes com OLIN (*Optimised Local Intensity-dependent Normalisation* o Normalització Optimitzada depenent de Localització i Intensitat; Futschik and Crompton, 2005), van encara un pas més enllà intentant corregir una varietat d'efectes espacials del microarray, incloent l'agulla, variabilitat de la superfície del microarray i diferències locals degudes a les condicions d'hibridació. OLIN utilitza el mètode LOCFIT, similar al Lowess, per fer la regressió local, però no hi ha paràmetres ajustables per l'usuari com, per exemple, el que determina quants han de ser els punts propers, sinó que l'algorisme compara els resultats derivats de donar diferents valors a aquests paràmetres i escull els més adients. Després es van fer iteracions fins que l'algorisme detecta que la transformació de dades ja no és millorable.

4.3.4 - Anàlisi de dades

Un cop tenim les dades normalitzades ja podem procedir a l'anàlisi de les mateixes per tal de llistar, per exemple, quins són els gens que estan significativament sobre- o sub-expressats en un experiment donat o trobar relacions entre diversos experiments. La complexitat de l'anàlisi augmentarà proporcionalment a la del disseny experimental: no és el mateix comparar unes cèl·lules tractades amb un fàrmac amb unes sense tractar a un temps determinat, que comparar entre elles cèl·lules tractades amb diversos fàrmacs i combinacions dels mateixos al llarg del temps. A banda d'aquesta complexitat inherent, en funció de quins siguin els objectius finals s'utilitzaran uns mètodes o uns altres.

La situació més senzilla és aquella en la qual només es disposa d'un microarray. Com que no es pot fer estadística, s'escull subjectivament un valor de canvi d'expressió a partir del qual direm que un gen està sobre- o sub-expressat. Quan disposem de rèpliques, que és el recomanable, podem utilitzar el mateix mètode fent la mitjana dels *ratios* de les rèpliques de cada gen o utilitzar mètodes que ens donin un nivell de credibilitat per cada gen; és a dir, que ens diguin si la reproducibilitat del *ratio* és estadísticament significativa entre les rèpliques. Entre aquests mètodes hi trobem, per exemple, el *Significance Analysis of Microarrays* (SAM; Tusher et al., 2001) o el Limma (Smyth, 2004), que permeten fer des de estudis molt senzills en els quals només es comparen dues condicions experimentals fins a d'altres de més complexos en els quals es comparen diferents condicions inclús al llarg del temps tot detectant quins són els gens significativament canviats. Altres tipus d'anàlisi com el de Components Principals (*Principal Component Analysis* o PCA) o Agrupament Jeràrquic (*Hierarchical*

Clustering o HCL) tenen per objectiu trobar grups de gens variats de la mateixa manera al llarg de diferents condicions experimentals.

Òbviament, a part de buscar patrons d'expressió i realitzar tests estadístics, hom pot dibuixar el seu propi itinerari i utilitzar les dades amb multitud d'objectius. Amb els gens diferencialment expressats podem, per exemple, veure si hi ha sobre-representació d'alguns tipus de gens en particular anotant-los funcionalment amb el Gene Ontology (Ashburner et al., 2000), situar-los en els cromosomes per identificar si hi ha regions on s'activen o es reprimeixen més gens, buscar conjunts (*clusters*) de gens co-regulats, analitzar les seves regions promotores per identificar seqüències compartides o comparar-los amb altres estudis de microarrays que ens puguin aportar alguna dada complementària.

OBJECTIUS

L'objectiu general d'aquesta Tesi és entendre la funció del gen *ash2* durant el desenvolupament a partir d'una aproximació genòmica mitjançant microarrays. Per tal de dur a terme aquesta tasca es proposen els següents objectius:

1.- Caracteritzar genèticament i molecular els al·lels *ash2*¹, *ash2*¹¹²⁴¹¹ i *ash2*¹. Els objectius concrets d'aquest punt són:

- Classificar la severitat dels al·lels mitjançant combinacions heteroal·lèliques.
- Seqüenciar l'al·lel *ash2*¹ per identificar la seva mutació.
- Analitzar l'expressió d'*ash2* en individus *wt* i en els diferents al·lels mutants.

2.- Desenvolupar i optimitzar la tecnologia dels microarrays de *Drosophila*. Els objectius concrets d'aquest punt són:

- Dissenyar i produir els microarrays i un conjunt de controls externs.
- Optimitzar els protocols d'extracció d'ARN de mostres petites, amplificació, marcatge i hibridació de mostres.
- Optimitzar les tècniques d'anàlisi de dades i comparar mètodes de normalització.

3.- Identificar els gens i processos regulats per ASH2. Els objectius concrets d'aquest punt són:

- Obtenir els perfils d'expressió gènica (transcriptomes) de diferents mutants pel gen *ash2*.
- Anotar i classificar els gens alterats a partir del Gene Ontology i de la integració de dades d'abast genòmic provinents de diverses fonts.
- Comparar el transcriptoma dels mutants del gen *ash2* amb els dels dos membres *trxG* funcionalment relacionats amb ell, cosa que implica obtenir els perfils d'expressió gènica de mutants d'*ash1* i *trx*.

