

UNIVERSIDAD MIGUEL HERNÁNDEZ
FACULTAD DE CIENCIAS EXPERIMENTALES
TRABAJO FIN DE GRADO EN BIOTECNOLOGÍA



Título: *Screening* de genes que cooperan con *pipsqueak* y/o *escargot* en los procesos de invasión celular y metástasis tumoral.

AUTOR: VALENZUELA MIRALLES, ANTONIO

Nº expediente: 931.

TUTOR: Domínguez Castellano, María

COTUTOR: Caparrós Cayuela, Esther

Departamento y Área: Mecanismos de control del crecimiento y cáncer en *Drosophila*. Neurobiología del desarrollo, Instituto de Neurociencias (CSIC-UMH).

Curso académico: 2017-2018.

Convocatoria de febrero.

Resumen:

Bien se ha constatado el papel activo de los genes *pipsqueak* (*psq*) y *escargot* (*esg*) en la enfermedad del cáncer en general, y en los procesos de invasión celular y metástasis tumoral en concreto. En el presente trabajo, se ha realizado una búsqueda de genes que puedan cooperar con *psq* y *esg* en dichos procesos, favoreciendo actuaciones sinérgicas que nos ayuden a desvelar un poco más los mecanismos moleculares intrínsecos en esta enfermedad. Entre los genes encontrados, destacamos *dL(3)mbt*, un gen supresor de tumores, cuya alteración, provoca la desregulación de un gran número de genes, muchos de ellos, con gran relevancia en la línea germinal. Mediante los resultados obtenidos en el laboratorio, y la realización de un exhaustivo análisis bioinformático, proponemos una hipótesis acerca de cómo este gen, siendo integrante de un complejo proteico, LINT, actúa reprimiendo estos genes en circunstancias normales, bajo la interacción física y funcional de SCM ("*sex comb on midleg*") y de PSQ-A (isoforma A del producto proteico de *pipsqueak*) en un mecanismo independiente de complejos proteicos pertenecientes al grupo Polycomb.

Abstract:

The active role of the *pipsqueak* (*psq*) and *escargot* (*esg*) genes in cancer disease in general, and in the processes of cellular invasion and tumor metastasis in particular, has been well established. In the present work, a search has been made of genes that can cooperate with *psq* and *esg* in these processes, favoring synergistic actions that help us to reveal a little more the intrinsic molecular mechanisms in this disease. Among the genes found, we highlight *dL(3)mbt*, a tumor suppressor gene whose alteration causes the deregulation of a large number of genes, many of them with great relevance in the germ line. Through the results obtained in the laboratory, and the performance of a comprehensive bioinformatic analysis, we propose a hypothesis about how this gene, being a member of a protein complex, LINT, acts by repressing these genes under normal circumstances, under the physical and functional interaction of SCM ("*sex comb on midleg*") and of PSQ-A (isoform A of the *pipsqueak* protein product) in a mechanism independent of protein complexes belonging to the Polycomb group.

Palabras clave:

Epigenética, LINT, *Screening*, SCM, *pipsqueak*.

Índice

Introducción.

Antecedentes.....	4
Transición epitelio-mesenquimal.....	9
<i>psq</i> y <i>esg</i> en cooperación.....	11

<u>Objetivos</u>	12
------------------------	----

Materiales y métodos

Cepas.....	12
Controles.....	15
Flujo de trabajo experimental.....	16
Análisis bioinformático: Un mundo movido por la curiosidad.....	18

<u>Resultados: Realización del <i>screening</i></u>	22
---	----

Discusión

<i>dL(3)mbt</i> , el aliado perfecto.....	25
LINT, un complejo capaz de unirse a SCM.....	26
Unión física y funcional entre SCM y PSQ.....	30
LINT se une cerca de los TTS.....	34
LINT reprime por un mecanismo independiente de Grupos Polycomb.....	34

Conclusiones

Conclusiones generales del trabajo.....	36
Conclusiones específicas del trabajo.....	36

<u>Perspectivas futuras</u>	36
-----------------------------------	----

<u>Bibliografía</u>	37
---------------------------	----

Introducción.

Antecedentes.

Aunque en el reino animal existe un amplio espectro de formas y una inmensa variedad de complejidades dentro de los sistemas aquí clasificados, los animales parecen poseer más características comunes de lo que a simple vista podríamos suponer. Un ejemplo de esto, lo podemos encontrar a nivel de procesos biológicos, los cuales, son similares en muchos organismos, sobre todo en lo referente a procesos celulares.

Por otra parte, continuamente se descubren nuevas especies, y por el contrario, la puesta a punto de ciertas técnicas conocidas sobre estas especies descubiertas recientemente puede ser muy costosa tanto en el plano económico como en el teórico-práctico. Por esta razón surge el concepto de “sistema modelo”, siendo el mismo, un organismo sujeto a un profundo e intenso estudio, con la idea de además de entender al propio sistema, encontrar esa potencialidad de extrapolación a otros organismos, estableciendo pues, un modelo general.

Dentro de estos “sistemas modelo”, encontramos el género *Drosophila*, siendo uno de los más utilizados. En este género, encontramos aproximadamente unas 2000 especies descubiertas, entre las cuáles destacamos *Drosophila melanogaster* como observamos en la siguiente imagen:

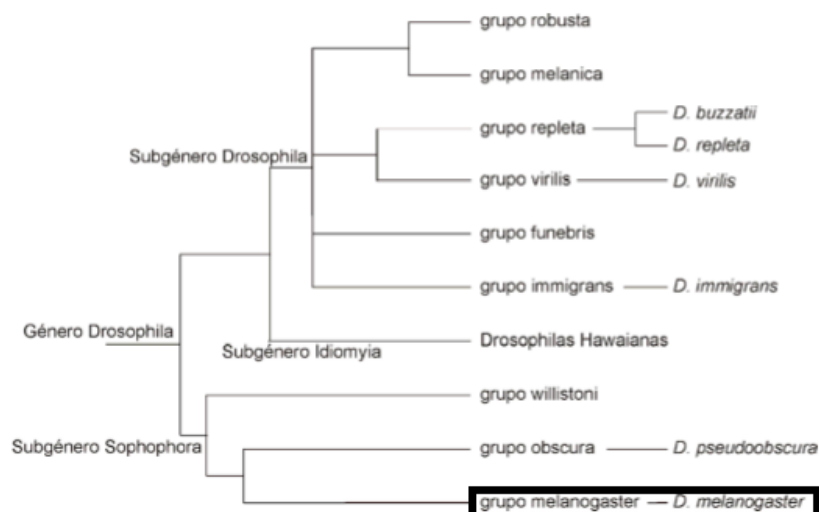


Figura 1. Árbol filogenético de los grupos principales en *Drosophila*. Destacado en negro, la especie *Melanogaster*, empleada en este trabajo. Fuente: Modificado de; Powell et al. (1997).

De este género, somos conocedores de comportamiento, ecología, desarrollo, biología molecular y un largo etcétera, sin olvidar que fue un modelo crucial para diversos estudios empíricos en la síntesis de la teoría genética y evolutiva (Powell et al. 1997). Algo crucial en este sistema, es la conservación de diversos procesos biológicos con respecto al ser humano, extendiendo de forma interesante una relación entre este díptero y la salud humana. De esta forma, cuando se identifica en la mosca un gen homólogo a un gen de humano o de mamífero (más generalizado) que juega algún rol con cierta importancia, todo aquel conjunto de técnicas sobre todo genéticas desarrolladas en *Drosophila* nos permite caracterizarlo y trabajar con él. Todo ello nos lleva al establecimiento y caracterización de una serie de redes génicas en la mosca que guardan homología en mamíferos en general o en humanos en particular, involucradas en aspectos potencialmente interesantes como puede ser el cáncer (Kornberg et al. 2000).

Durante los últimos años, tanto la codificación del genoma de *Drosophila*, como el desarrollo de diversas herramientas genéticas avanzadas para la edición del mismo, han contribuido a formar un estímulo muy importante con su consiguiente expansión en la utilización de este sistema modelo en cáncer. Estudios en la mosca de la fruta fueron fundamental en el descubrimiento de que el cáncer estaba causado por mutaciones en genes específicos (Gateff & Schneiderman, 1967, 1969; Gateff, 1978). El primer mutante capaz de indicar cáncer se describió en *Drosophila melanogaster* (Bridges, 1916) pero sólo 25 años después se obtuvo la evidencia de que la transformación neoplásica era causada por la alteración de un gen celular. Posteriormente, estudios genéticos en *Drosophila* establecieron el papel crucial de las vías de desarrollo como Notch, Hedgehog, y Wingless/WNT en cáncer (Artavanis-Tsakonas, 1999; Taipale et al. 2001). Más recientemente estudios genéticos en *Drosophila* han sugerido el papel de tres nuevos factores en la progresión tumoral, y la metástasis (Wang et al. 2014). El primer factor es el papel de la vía de señalización de Hippo, situándose en el control tanto del crecimiento como en la supervivencia celular, junto con las rutas de señalización Scrib/Dlg/Lgl, interviniendo la polaridad celular para la regulación de los tamaños de los órganos (Martin-Belmonte et al. 2012). El segundo factor trata el mecanismo de competición celular intrínseco para morfógenos durante la formación de epitelio, donde encontramos la eliminación de las células de crecimiento lento (Levayer & Moreno, 2013). Por último, el tercer factor, son las señales mitogénicas inducidas por el proceso de apoptosis en la proliferación compensatoria para reemplazar aquellos tejidos circundantes que poseen daños (Fan & Bergmann, 2008). Estas vías subrayan la importancia de la comunicación intercelular, así como la propia organización de los tejidos, cuya interrupción, puede facilitar el proceso de tumorigénesis y la metástasis.

La vía de comunicación celular mediada por los receptores Notch (NOTCH1-4 en el hombre) es fundamental para muchos tipos de tumores y en la formación de metástasis. La vía canónica de Notch, descrita por primera vez en *Drosophila*, está activada por dos tipos de ligandos, Delta (denominados DELTA-LIKE, DLL1, 3-4 en el hombre, y Dll1, 2-3 en ratón) y Serrate (JAGGED 1-2, en el hombre; y Jagged1,2 en ratón). La vía de Notch desempeña papeles fundamentales, conservados evolutivamente, en la determinación de destino celular, diferenciación, migración celular y transición epitelio-mesenquimal, formación de patrones, y en la proliferación celular y muerte celular programada (Previs, Coleman et al.

2014). Todos estos procesos en los que interviene Notch, son de vital importancia en la enfermedad del cáncer, moldeando el comportamiento del tumor y la respuesta a un hipotético tratamiento. Destacar pues, que la dependencia tanto del tejido como del contexto, hacen que alteraciones en Notch puedan derivar a mostrar fenotipos oncogénicos o supresor tumoral (Previs, Coleman et al. 2014).

Por otra parte, existe una amplia variación en la expresión de los cuatro receptores de Notch en tejidos adultos y en tejidos embrionarios, observando patrones de expresión superpuestos, pero con papeles únicos y definidos durante la formación de células madre hematopoyéticas, en el destino y desarrollo de linfocitos T y B y morfogénesis vascular. La desregulación de la vía de Notch esta implicada en una extensa variedad de neoplasias de carácter maligno tanto hematológicas como sólidas, viéndose establecida esta anormalidad por diversos mecanismos como puede ser la activación o inactivación mutacional, las modificaciones postraduccionales y el factor epigenético (Ntziachristos et al. 2014).

Como ocurre con otros oncogenes, la activación anómala de la vía de Notch, no parece ser condición suficiente para el desarrollo tumoral, conduciendo a la idea de la necesidad de la existencia de acción o acciones sinérgicas con otros genes para dar lugar a la patología, sin olvidar la herencia epigenética, implicada además de en cáncer, en el control de la proliferación celular durante la fase de desarrollo y la renovación de células madre (Ferres-Marco, Gutierrez-Garcia et al. 2006).

Bajo este marco referencial, el laboratorio de María Domínguez, en el Instituto de Neurociencias (Centro mixto CSIC-UMH), en Sant Joan d'Alacant, utilizando el cribado genético directo, y como modelo de estudio *Drosophila melanogaster*, ha sido pionera en el diseño de cribados de genes que en cooperación causa tumores in vivo (Ferres-Marco et al. 2006). El laboratorio está centrado en genes que cuando se activan (o se inactivan, genes supresor de tumor) facilitan la transformación tumoral por la vía de Notch. Estos estudios llevaron a desvelar por primera vez que vías como la de Notch cooperan con las vías de silenciamiento epigenéticos, que regulan la memoria celular, en la iniciación y progresión tumoral. Estos estudios abrieron el camino para el desarrollo de nuevas terapias antitumorales para cánceres asociados a la vía de Notch.

Para la identificación óptima y eficiente de genes, se utilizó la metodología Gene Search (GS) (Toba et al. 1996). Esta técnica se basa en la utilización de un vector capaz de transponerse (saltar) en el genoma y que contiene en orientación inversa copias de las secuencias potenciadoras UAS ("Upstream Activating Sequence") inducibles por el factor de levadura Gal4. El vector además contiene el gen *miniwhite* que permite trazar los eventos de transposición del vector en el genoma mediante el cambio de color en los ojos de la mosca.

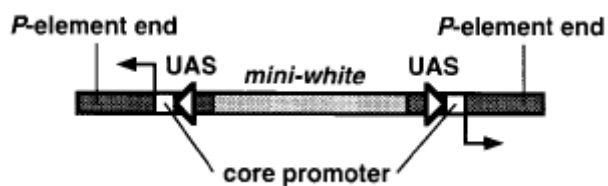


Figura 2. Vector utilizado en el método de búsqueda de genes Gene Search. Destacar que el gen *mini-white* funciona como un marcador. Fuente: Toba et al. 1996.

El procedimiento consta de tres etapas consiguientes (Brand & Perrimon, 1993; Toba et al. 1996; Ferres-Marco et al. 2006): En primer lugar, se parte de una línea de individuos poseedora de una inserción individual del vector, que permite la transcripción de la secuencia genómica próxima al sitio de inserción del vector cuando dicha línea se cruza con otra línea de moscas que produce la proteína Gal4. En segundo lugar, se generan nuevas inserciones mediante la transposición de la línea original utilizando una línea que produce la enzima transposasa. En tercer lugar se realiza un cribado de las nuevas líneas GS y se seleccionan las líneas que son capaces de producir tumor cuando se cruzan con moscas que expresan la proteína Gal4 y además el ligando de Notch, Delta (Ferres-Marcó et al. 2006). El último paso sería la identificación a nivel molecular de aquellas secuencias transcritas por las líneas seleccionadas.

Los cribados de nuevos oncogenes que cooperaban con la vía de Notch utilizaron la línea Gal4 que se expresa de forma específica en el tejido del ojo en proliferación (línea *eyeless (ey)-Gal4*). Esta línea se combinó con un transgen que expresa Delta y denominado por simplicidad (*ey>DI*). La sobreexpresión de Delta (ligando de Notch), con *ey-Gal4* produce sobre-crecimiento en el ojo dando lugar a moscas adultas con un ojo de un tamaño ~120% más grande que el control. El cribado de genes utilizó esta línea *ey>DI* y se analizaron las líneas GS, y por lo tanto, los genes próximos a la inserción de la línea GS, que bajo la acción de la proteína Gal4 convertía el sobre-crecimiento inducido por *Delta* en crecimiento tumoral. Una de las líneas GS identificadas en dicho cribado fue la línea *GS88A8*, denominada posteriormente *eyeful*, en la que se apreciaron el desarrollo de tumores oculares, y en un tercio de los individuos, crecimiento metastático (Ferres-Marco et al. 2006).

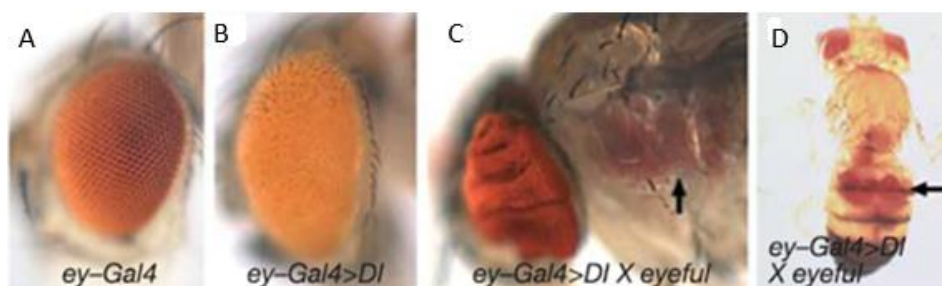


Figura 3. Comparación del crecimiento ocular y desarrollo tumoral.

A) Fenotipo de la línea *ey-Gal4*. B) Fenotipo de la línea mutante *ey-Gal4*. Observamos un crecimiento leve del ojo al sobre-expresar el ligando Delta, pasando de 800 omatidios a 1200 en su ojo compuesto (Domínguez & de Celis, 1998). C) Fenotipo tumoral de la línea mutante *ey-Gal4>DI X eyeful*. D) Crecimiento ocular secundario en el mutante *ey-Gal4>DI X eyeful*. Fuente: (Ferres-Marco, Gutierrez-Garcia et al. 2006).

A continuación, el equipo extrapoló los resultados al disco imaginal de ala. A nivel celular, el disco de ala crece y se forma a partir de un primordio de unas 50 células que tras nueve divisiones alcanza una extensión final de unas 32.000 células (Martín & Morata, 2006).

Al principio de la fase de desarrollo, encontramos el disco subdividido en compartimentos, especificados por una serie de genes selectores y originados por pequeños grupos celulares, denominados policlones (Montes Ruiz, A.J., tesis doctoral, 2016). Otro nivel en que podemos definir el disco imaginal de ala, es en el plano morfo-histológico. El disco adopta una morfología típica de un saco plano, diferenciando un lado imaginal (epitelio columnar, mucho más denso) y un lado peripodial (epitelio escamoso, menor densidad celular) encapsulando un lumen (García-Bellido et al. 2006). Del primero, además de desarrollar el ala, da lugar al nótum, es decir, la parte dorsal del tórax. (García-Bellido et al. 2006)

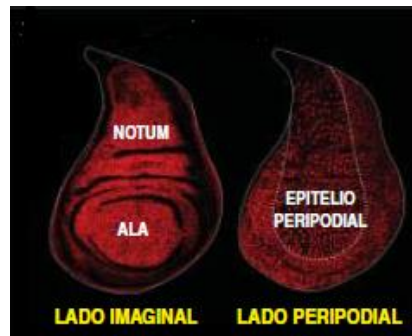


Figura 4. Disco imaginal de ala. Apréciase una mayor densidad celular en el lado imaginal (a la izquierda) en contraste con el epitelio peripodial. Fuente: García-Bellido et al. 2006.

Pero quizás, uno de los ámbitos de análisis más interesantes de los discos imaginales, en concreto para este trabajo los discos imaginales de ala, es a nivel morfo-genético. Éstos se subdividen como se ha comentado anteriormente y se van especificando sin que ocurran mezclas celulares debido a las diferentes afinidades existentes. (Montes Ruiz, A.J., tesis doctoral, 2016). Este modelado de las diferentes afinidades está conformado por la señalización de diversas vías y moléculas denominadas morfógenos. En el caso concreto del ala, encontramos cuatro a destacar. En primer lugar, encontramos Engrailed (*En*), el cual, se expresa en células del compartimento posterior (P), dotando de identidad al compartimento, y a su vez, estableciendo el límite anterior-posterior (A-P) (Beira & Paro, 2016). A continuación encontramos la molécula de corto alcance Hedgehog (*Hh*), pudiendo cruzar el límite A-P y promover la inducción de Decapentaplegic (*Dpp*), la cual, se expresa a lo largo del límite comentado, permitiendo su marcado, la señalización de largo alcance para dirigir el patrón de un fragmento de ancho del disco. Por último, encontramos Wingless (*Wg*), que se manifiesta en la zona del límite dorsal-ventral (D-V). (Beira & Paro, 2016).

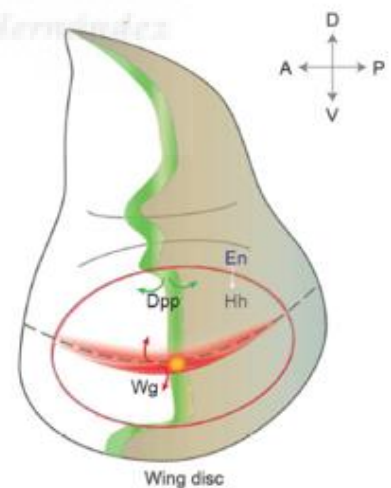


Figura 5. Patrón y dominios de expresión en discos de ala. Encontramos la actuación de los morfógenos *En*, *Hh*, *Dpp* y *Wg*, imprescindible en el desarrollo del disco imaginal. Fuente: Beira & Paro, 2016

Utilizando *dpp-Gal4* para dirigir la coexpresión de *Delta* y *eyeful* a lo largo del límite A-P, el equipo observó mediante técnicas de fluorescencia (expresión de GFP) que las células mutantes *eyeful* con *Delta* podían

diseminarse e invadir nuevo tejido, mientras que las wt (solo *eyeful*) expresaban el típico patrón de *dpp* (Ferres-Marco, Gutiérrez-García et al. 2006).

A nivel genético, *eyeful* se encuentra insertado en el segundo intrón del gen *lola* (*longitudinal lacking*), uno de los genes principales en la regulación de la guía axonal (codifica 25 RNA_m por splicing alternativo, generando 19 FT (Goeke et al. 2003). Debido a la bidireccionalidad del sistema GS (Toba et al. 1996), encontramos hacia el lado opuesto, la aparición del gen *psq* (*pipsqueak*). Este gen codifica nueve variantes mediante un mecanismo de splicing alternativo, dando lugar a cuatro diferentes isoformas proteicas con ciertas características comunes (Goeke et al. 2003). Tres de las isoformas de este gen poseen dominios POZ (POxvirus and Zinc finger) también denominados BTB (Broad-Complex, Tramtrack and Bric a brac) en posición N-terminal, cuya funcionalidad a nivel bioquímico reside en mediar procesos de homodimerización o heterodimerización con otras proteínas (Bardwell & Treisman, 1994). Por otra parte, dos de las isoformas que contienen dominios BTB y la isoforma que no lo posee, cuentan en su haber cuatro copias distribuidas en tándem de un motivo de unión a DNA, denominado motivo HTH (Helix-Turn-Helix) (Siegmund & Lehmann, 2002). Entre las funciones de este gen, encontramos la unión a la secuencia GAGA, presente en muchos genes homeóticos (Hox) imprescindibles para el desarrollo de las moscas de la fruta (Schwendemann & Lehmann, 2002), jugando un papel destacable en la determinación del destino de las células de la retina (Weber et al. 1995) y además, es esencial en la orientación específica de la secuencia para que se produzca la actividad histona deacetilasa (HDAC) del complejo PcG (Ringrose & Paro, 2004), proceso epigenético que, podría ayudar a desarrollar tumores.

Todos los experimentos que realizaron confirmaron que la transcripción de ambos genes (*psq* y *lola*), estaban influenciadas directamente por *eyeful* debido a la activación de la proteína Gal4, y que además, la desregulación de ambos genes estaban implicados y eran necesarios (por si solos no dan lugar a tumores) para dar lugar al fenotipo tumoral (Ferres-Marco, Gutiérrez-García et al. 2006). Aun siendo esto así, la desregulación de *psq* parece contribuir en mayor medida al fenotipo neoplásico, por su acción silenciadora a nivel epigenético, más profundamente, produce la pérdida de trimetilación de la histona H3 en K4 (H3K4me3), por lo cual, produce un silenciamiento aberrante de la cromatina, conduciendo en última instancia a la formación de tumores (Ferres-Marco, Gutiérrez-García et al. 2006).

Transición epitelio-mesenquimal.

Las células de la epidermis larvaria (y la de los discos imaginales), muestran una conformación y una organización epitelial, es decir, uno o más estratos celulares funcionalmente polarizados en el eje ápico-basal (ortogonal al plano del epitelio) (García-Bellido et al. 2006). Estas células no optan por funcionar de forma individual, sino que, forman una compleja red de comunicación celular entre las células adyacentes ya sea por adhesión intercelular como uniones adherentes, uniones estrechas y/o uniones gap, o por el contrario mediante la lámina basal, formada por matriz extracelular (MEC) (Acloque et al. 2009).

Uno de los procesos morfo-genéticos más importante que encontramos en la morfogénesis epitelial, es la transición epitelio-mesenquimal (EMT), siendo un proceso integral para la formación de muchos tejidos

y órganos durante diversos estadios de desarrollo, siguiendo una secuencia bien definida y ordenada de eventos. En primer lugar, la zona donde se producirá la transición, debe especificarse a través del patrón espacio-temporal de aquellas células que sufrirán la transformación, además de la reorganización morfológica del propio tejido epitelial, con la finalidad de trasladar dichas células al sitio de la EMT (Nisticò et al. 2012). Por otra parte, las comunicaciones entre las células y la membrana basal (MB), debe ser interrumpida, pudiendo ocurrir esto, mediante la liberación de aquellos contactos célula-MB, o bien, por la degradación proteolítica de la propia membrana basal (Nisticò et al. 2012). En tercer lugar, las células que se encuentran en transformación, como es lógico, deben separarse de la lámina epitelial mediante procesos que alteren lo menos posible la integridad del epitelio, generalmente por reordenamientos de actomiosina, seguido del rastreo de células epiteliales que se encuentren retenidas, con el objetivo de cerrar la brecha que se genera fruto de la egresión. Por último, ocurre la diferenciación fenotípica de las células transformadas, ocurriendo cambios en la organización del citoesqueleto, así como ciertos cambios a nivel metabólico (Nisticò et al. 2012).

Durante la progresión neoplásica, el establecimiento de la transición epitelio-mesenquimal permite que las células tumorales puedan adquirir mayor motilidad confiriendo capacidades migratorias, y por ende, invasivas (Acloque et al. 2009). Aun pareciendo que existe una relación directa entre EMT y la progresión tumoral, la realidad nos expone un fenómeno aún más complejo. Se ha demostrado que la inhibición de la EMT está relacionada con la curación del cáncer, y que además, la transición mesénquimal-epitelial (MET), la inversa a la EMT, está asociada con la capacidad de iniciación tumoral, requerida para la posterior colonización metastásica (Yang et al. 2017). Así que, estas evidencias contradictorias per sé, se pueden explicar de forma lógica si pensamos en un estado intermedio de EMT, donde es más probable que ocurran cambios dinámicos y/o plásticos entre ambos estados (haciendo surgir el concepto de células epitelio-mesenquimatosas híbridas) en lugar de un fenotipo celular fijo en el contexto clínico (Yang et al. 2017).

De una forma u otra, este proceso (EMT) parece albergar relaciones con la patología y esto se explica por un cúmulo de mecanismos moleculares que subyacen a la transición provocando cambios de importante relevancia como en el citoesqueleto (base de la ganancia de motilidad), cambios en la expresión génica, etc. Existen un número elevado de moléculas que intervienen, como fibronectinas, citoqueratinas, vitronectinas, N- y E-cadherinas, integrinas, etc., controladas por diversos factores de transcripción como factores ZEB (Zinc-finger E-box-Binding), factores bHLH (basic Helix-Loop-Helix) y factores del tipo SNAIL, los cuales, buscan reprimir aquellos genes de marcador epitelial, y activar los genes asociados al fenotipo mesenquimal (Iamouille et al. 2014).

La familia SNAIL se describió por primera vez en *Drosophila*. Este organismo contiene tres genes de esta familia denominados *snail*, *worniu* y *escargot* (*esg*), siendo este último el ancestro de la familia SNAIL. El gen *esg*, con una extensión de 6.273 kbp, codifica una proteína de 470 aminoácidos (con una masa de 51.963 Da), que consta de 5 motivos de dedos de zinc, cuatro de ellos del tipo C₂-H₂, y uno atípico aunque enmarcado en este tipo (Webster et al. 1991). En cuanto a la funcionalidad, se presenta como un factor

de transcripción, que aunque puede activar el proceso de transcripción, se asocia más como un represor. Entre sus funciones endógenas destacan, la mantención de la diploidia en células imaginales (al actuar como represor de ciertos genes que producen endoreplicación), también es importante en el desarrollo del disco genital, así como es un factor determinante intrínseco en el destino de las células del ala (Hayashi et al. 1993). Además, es capaz de regular la motilidad mediante la represión del gen *E-cadherin* uniéndose a la secuencia consenso 5'-A/GCAGGTG-3', estableciendo la iniciación del proceso de invasión (Peinado et al. 2007; Ferres-Marco, D., tesis doctoral, 2010).

Cooperación entre *psq* y *esg*.

La doctora M^ª Dolores Ferres aisló en el laboratorio una línea GS que se encontraba dentro del locus de *escargot*, la línea *GS(2)101M2*, la cual, con sobreexpresión de *Delta* mostraba un crecimiento del ojo. Esta línea era incapaz por sí misma dar lugar a formación de tumores ni invasividad, pero en combinación con sobreexpresión de *pipsqueak* y/o *lola* si puede dar lugar a invasión y metástasis.

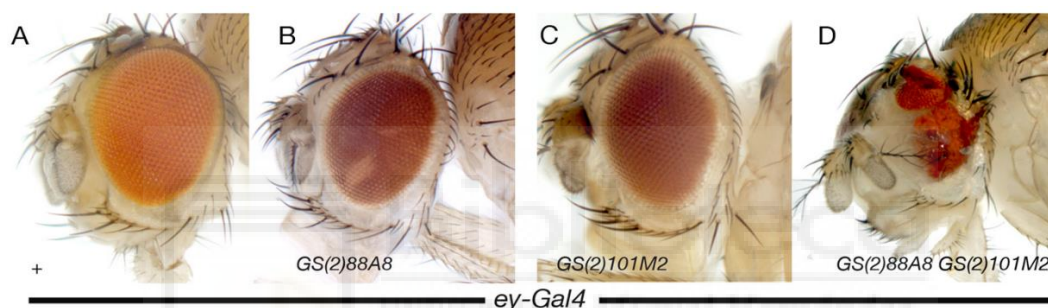


Figura 6. Inducción de tumores invasivos debido a la co-expresión de *GS(2)88A8* y *GS(2)101M2*. A) Foto tomada a una hembra "wildtype" pudiendo observar la morfología y tamaño de su ojo. B) Mutante *ey-Gal4>GS(2)88A8* (*eyeful*). No hay atisbos de invasión celular en el ojo. C) Mutante *ey-Gal4>GS(2)101M2* (*esg*). Observamos un crecimiento ocular característico. D) Mutante *ey-Gal4>GS(2)88A8 GS(2)101M2*. Se observa invasión en los ojos en más de la mitad de las moscas. Fuente: Ferres-Marco, D., tesis doctoral, 2010

Por otra parte, el fenotipo mostrado por la co-expresión de ambos FT fue una completa alteración de la onda de diferenciación (ordenación columnar de los omatidios que sigue un patrón espacio-temporal desde la región posterior hasta la región anterior del disco), del patrón antero-posterior, de la subdivisión cabeza-retina, así como la adquisición de una morfología mesenquimática o semejante (Ferres-Marco, D., tesis doctoral, 2010). Además de todo ello, estos discos presentaban altos niveles de muerte celular, idea que se deduce debido a la fuerte tinción con un anticuerpo específico contra la caspasa-3 activada (y mediante un mecanismo independiente de la vía JNK), mostrando muchas células la aparición de cuerpos pignóticos y citoplasmas fragmentados (Ferres-Marco, D., tesis doctoral, 2010).

Lo último comentado, está fuertemente ligado al concepto de apoptosis o muerte celular programada, inducción de la cuál esta mediada por diversos genes como *reaper* (*rpr*) (se une y reprime a las IAPs (Proteínas Inhibidoras de Apoptosis) (Abrams et al. 1999; Ferres-Marco, D., tesis doctoral, 2010)), cuyo

promotor posee secuencias de unión GAGA (Farkas et al. 1994; Ferres-Marco, D., tesis doctoral, 2010), a los que se puede unir *psq* (Schwendemann & Lehmann, 2002), lo que llevó a demostrar al equipo, que este gen se inducía consistentemente en aquellos discos que expresaban tanto *pipsqueak* como *escargot*, y aunque, *esg* no es un potente factor pro-apoptótico, si aumenta la capacidad de *psq* para inducir la expresión de *rpr* (Ferres-Marco, D., tesis doctoral, 2010).

Esto que a priori parece contradictorio con invasión y metástasis (ya que se supone la resistencia de las células tumorales al proceso de apoptosis) cobra sentido si hacemos una separación temporal de los sucesos, es decir, aquellas células que en estadios tempranos son capaces de resistir a procesos apoptóticos, son las que finalmente podrán migrar e invadir espacios adyacentes, haciendo un cribado por tanto el primer proceso y facilitando finalmente la invasión (Ferres-Marco, D., tesis doctoral, 2010).

Objetivos

El principal objetivo de este trabajo es profundizar en el mecanismo molecular que subyace la progresión tumoral hacia el crecimiento metastático. Para ello, tenemos como objetivos particulares:

1_ Identificar, mediante un cribado genético, genes que facilitan la formación de metástasis por los factores *Psq* y *Esg*.

2_ Definir las redes genéticas que facilitan el crecimiento invasivo y metástasis.

Debido a que las moscas *ey-Gal4>GS412ND(psq)>GS101M2(esg)* son poco viables, el cribado de nuevos genes se ha llevado a cabo utilizando el disco imaginal de ala en estadio larvario y la línea *dpp-Gal4*. Esta línea *dpp-Gal4* se expresa en el estadio larvario en una franja de células en el borde entre el compartimiento anterior y posterior del disco de ala. La expresión de los genes *psq* y *esg* en el dominio de *dpp* produce una franja más gruesa que el wt, debido a un pequeño incremento de la proliferación celular, y además de un grupo de células de este dominio es capaz de invadir y migrar desde el dominio '*dpp*' hacia otras partes del compartimiento anterior del ala e incluso invadir el compartimiento posterior, algo que nunca hacen las células normales. Este fenotipo es robusto y la estrategia de trabajo ha sido buscar genes que cuando se silencian mediante RNA interferente (RNAi) modifican (agravan) los defectos causado por la expresión de *Psq* y *Esg* en el dominio *Dpp*. Por último, trataremos de definir redes moleculares en las que se integren los nuevos factores identificados y que contribuyan a entender como se inicia el proceso metastático *in vivo*.

Materiales y métodos

Cepas:

GS(2)psq/lola GS(2)esg / CyO Gal80; dpp-Gal4 UAS GFP /TM2: Se utilizaron hembras vírgenes con este genotipo. Al establecer esta nomenclatura salta a relucir un nuevo concepto, cromosomas balanceadores o equilibradores. Este término derivado a herramienta genética, se utiliza para mantener mutaciones deletéreas (mutaciones homocigóticas letales o estériles) en poblaciones estables, además de evitar que

diversos alelos en un mismo cromosoma se escindan por recombinación meiótica. Además, estos cromosomas normalmente son letales o estériles en uno de los dos sexos, o bien, exhiben una viabilidad bastante reducida en el estado homocigótico con la finalidad de evitar el desplazamiento de sus homólogos en stocks equilibrados. Añadiendo, estos cromosomas se encuentran marcados de tal manera que su presencia en heterocigotos es reconocible, gracias normalmente a una combinación de alelos mutantes dominantes y recesivos. Los dos balanceadores aquí utilizados son “CyO” y “TM2”, siendo el primero posiblemente el balanceador más utilizado en el segundo cromosoma, y cuyos puntos de ruptura más distales los encontramos en los puntos 22D y 58B, por lo que resulta más eficiente su utilización en zonas próximas relativamente al centrómero. Por otro lado, “TM2”, lo encontramos balanceando al tercer cromosoma (la primera letra proviene de “Third”, la segunda de “Multiply inverted” y el número es para diferenciar de otros balanceadores de este cromosoma), cuya participación elimina el entrecruzamiento en este cromosoma a excepción del final del brazo derecho del mismo. Cabe destacar la posibilidad de efecto inter cromosómico, es decir, al evitar la recombinación meiótica en una región concreta del genoma, la propia célula reacciona con un aumento de la recombinación en otras regiones, por lo que, al coexistir varios balanceadores, puede ocurrir recombinación en ciertas regiones donde no esperaríamos cuando solo hay un equilibrador (*Bloomington Drosophila Stock Center*).

Por otra parte, debemos utilizar el sistema híbrido Gal4-Gal80 (Gal80 reprime a Gal4), como forma de auto-represión en circunstancias basales, y que por tanto, estas moscas per se, no expresen fluorescencia (debido a la no expresión de GFP), y que solo la expresen aquellos individuos de la generación filial que nos interese fruto de los cruces realizados con las líneas de RNAi.

En situaciones de ausencia de galactosa, Gal4 se encuentra en un estado de inactivación debido a la unión de la proteína represora Gal80, evitando por tanto la interacción entre el dominio de activación de la primera, con la maquinaria transcripcional (Traven, Jelcic et al. 2006), repercutiendo a escalas tales que, se reprima la acción de la proteína de unión a la secuencia TATA (“TATA Binding Protein” o TBP) (Wu et al. 1996), o el bloqueo de la interacción del dominio de activación de Gal4 con Spt-Ada-Gcn5 acetiltransferasa (SAGA) (Carrozza et al. 2002) por citar algunos ejemplos.

En contraste, la inhibición de esta represión se ve facilitada por la unión de una tercera proteína de esta familia, Gal3, siendo un factor crucial la presencia de galactosa para que Gal3 y Gal80 interaccionen como reportan los estudios bioquímicos (Zenke et al. 1996). Aun no sabiendo con exactitud como la formación del complejo Gal3-Gal80 es capaz de aliviar la represión de Gal4, se hipotetizó la idea de que Gal3 interactúa con Gal80 en el propio núcleo celular, provocando cambios conformacionales que desestructurarían y romperían el complejo Gal80-Gal4 (Leuther & Johnston, 1992). Existiendo diversas evidencias que favorecen la hipótesis, como por ejemplo estudios de microscopía de transferencia de energía por resonancia de fluorescencia (FRET), que confirman la permanencia de la asociación entre Gal4 y Gal80 en presencia de galactosa (Bhaumik et al. 2004), lo cierto es que diversos estudios de localización afirman con bastante rotundidad, que Gal3 es un factor estricta y exclusivamente citoplasmático, conduciéndonos a la idea de que la unión entre Gal80 y Gal3 ocurre en zonas de transición entre el

citoplasma y el núcleo, además de afirmar que si ocurre disociación entre Gal4 y Gal80 (Peng & Hopper, 2000).

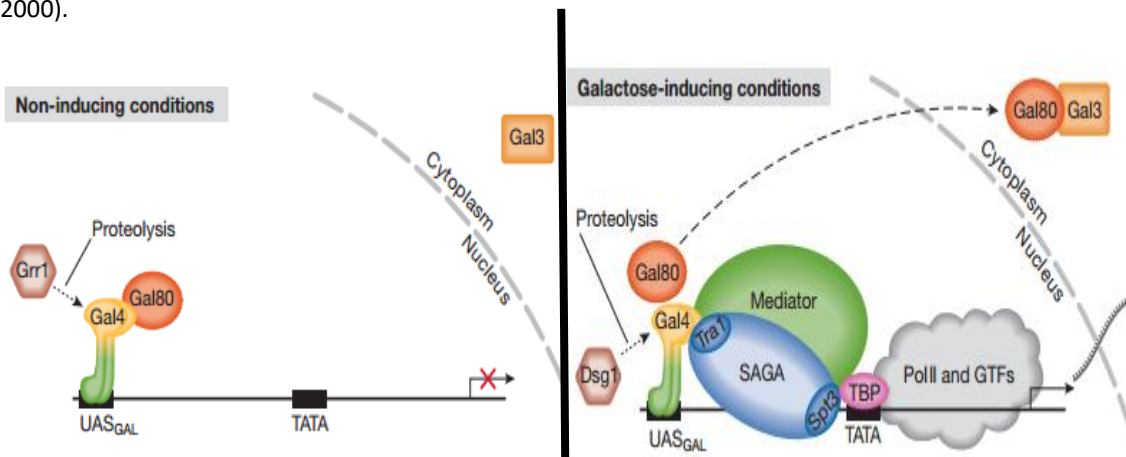


Figura 7. Activación transcripcional por Gal4. Podemos ver que en condiciones de no inducción (a la izquierda), Gal4 se encuentra inhibida gracias a su interacción con Gal80, en el dominio de activación. Por otra parte, la proteína de la F-Box, Grr1, regula los niveles de Gal4 mediante un mecanismo proteolítico. A la derecha, observamos las condiciones de inducción por galactosa, escindiéndose Gal80 del dominio de activación de Gal4, uniéndose a Gal3, y por tanto dejando en un estado activo a Gal4. La encargada de mantener los niveles de Gal4 en estas condiciones es Dsg1, mediante proteólisis. Fuente: Traven et al. 2006.

Líneas RNAi: De estas líneas utilizamos solo los machos. Trabajamos con líneas de varias librerías, tanto de VDRC (Vienna Drosophila Resource Center) como de BDSC (Bloomington Drosophila Stock Center). Dentro de las librerías de Viena, trabajamos con dos, GD (librería con elemento P) y KK (librería $\Phi C31$), dirigiéndose entre las dos al 91% de los genes codificantes (12.671 genes) de *Drosophila*. Aun teniendo mayor cobertura la primera librería (1,2 líneas/gen frente 0,7 líneas/gen en KK), debido a la aleatoriedad de inserción y la localización imprecisa de las secuencias UAS-IR a lo largo de los cromosomas X, 2 y 3 (Dietzl et al. 2007), la librería GD, la utilizamos en menor medida que la librería KK, la cual, presenta inserciones en el cromosoma 2, en la posición 22019296 (5' a CR33987) (Horn & Boutros, 2010)

En lo referente al stock de Bloomington, las líneas utilizadas han sido generadas por TRiP ("Transgenic RNAi Project"), en alguno de los cinco vectores tipo VALIUM (VALIUM1, VALIUM10, VALIUM20, VALIUM21 Y VALIUM22). Éstas líneas utilizan el sistema Gal4/UAS (Brand & Perrimon, 1993), con el objetivo de inducir específicamente una estructura en forma de horquilla, capaz de silenciar la expresión del gen diana. Ahora bien, estos transgenes contenidos de la horquilla, son insertados mediante recombinación específica de sitio (Groth et al. 2004) en loci óptimos, haciendo de esta metodología una herramienta competitiva a la par que simple y rápida, ofreciendo entre otras ventajas, el control tanto espacial como temporal de la construcción.

Cruces resultantes: La generación filial resultante de los cruces debe poseer el siguiente genotipo: 412 101M2; *dpp-Gal4 UAS GFP/ línea RNAi*, siendo "412 101M2" sinónimo de "*GS(2)psq/lola GS(2)esg*", de lo contrario, las moscas no mostrarán fluorescencia por estar reprimidas debido al sistema Gal4-Gal80.

Controles:

RNA_i/+; dpp-Gal UAS GFP: Éste sería nuestro control negativo. Estamos expresando la línea RNA_i sobre el *wildtype*, así que esperaríamos que el RNA_i per sé no produjera un cambio en el fenotipo, pues de lo contrario, esto nos daría a conocer que no existe una interacción entre *pipsqueak* y/o *escargot* con el gen inhibido, sino que el propio gen es capaz de inducir cambios, cuestión interesante, pero no en el desarrollo del presente trabajo.

GS(2)psq GS(2)esg/RNA_i; dpp-Gal4 UAS GFP/+: Utilizado como control positivo. Expresando tanto *pipsqueak* como *escargot* sobre la línea RNA_i, esperaríamos una modificación del fenotipo basal del patrón de dpp. Esto nos daría a conocer, que existe una interacción entre el gen inhibido por el RNA de interferencia y *pipsqueak* y/o *escargot*. Ahora bien, este control no nos permite discriminar con que gen está asociado la línea (con *psq* o *esg*), o si por el contrario, es una acción conjunta fruto de una interacción múltiple con ambos genes.

GS(2)101; dpp-Gal4 UAS GFP/ RNA_i: Control positivo. El genotipo resultante surge del cruzamiento (♀) *GS(2)101M2/CyO Gal80; dpp-Gal4 UAS GFP/ TM2* x (♂) línea RNA_i. El control aquí descrito, nos permite conocer si la acción sinérgica de la línea RNA_i es debida a la interacción con *escargot*, mostrando en caso afirmativo, el cambio oportuno de fenotipo, ya sea como acción represora (normaliza el patrón de dpp), o como acción potenciadora de invasión y metástasis, agravando el patrón visual normal del dominio.

GS(2)412; dpp-Gal4 UAS GFP/RNA_i: Control positivo. Como en el caso anterior, este tipo de control, nos permite dilucidar que gen es el que mantiene cooperación con la línea RNA_i, en este caso con *pipsqueak* (también se expresa *lola*).

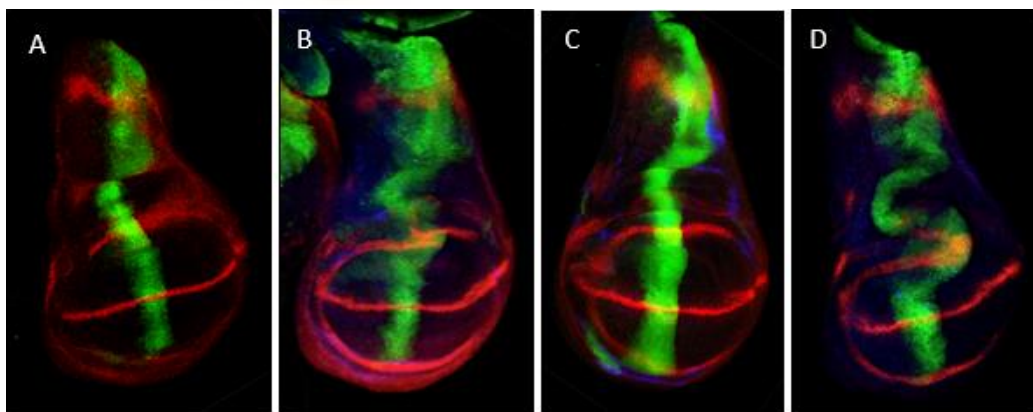


Figura 8. Visualización de los diferentes controles con microscopía confocal.

A) Control negativo. Observamos un dominio dpp de amplitud normal, con bordes bien definidos. En el límite dorsal-ventral hallamos en rojo al dominio Wg marcado mediante técnicas de inmunohistoquímica. B) Control positivo con el gen *esg*. Apreciamos un ensanchamiento de la banda transversal de GFP, así como bordes difuminados en la misma. En azul, podemos observar indicios de actividad apoptótica debido a la presencia de caspasa 3, marcada con anticuerpos específicos, lo que nos da a entender la existencia de células pre-invasoras. C) Control positivo con el gen *psq* (también *lola*). A priori podemos contemplar un ensanchamiento del dominio dpp, aunque en menor medida que en el caso anterior, y con bordes bien definidos. D) Control positivo general con ambos genes. Como es lógico, advertimos ensanchamiento del dominio dpp y bordes difuminados.

Flujo de trabajo experimental:

El trabajo comienza con la selección de hembras vírgenes. Esto lo podemos hacer debido a que las hembras no copulan hasta pasadas unas 8 horas desde su emergencia de la pupa. Para poder discriminar las hembras de los machos, debemos tener en cuenta una serie de características, como por ejemplo la presencia de peines sexuales (filas de cerdas negras y quitinosas en el primer segmento tarsiano del primer par de patas) en los machos, o la presencia del meconio (mancha oscura en el abdomen) en las hembras (King & Reiss, 2001). En cuanto a la pigmentación, los machos presentan 5 bandas oscuras de forma transversal y muy unidas en el extremo posterior del abdomen al contrario que las hembras (7 bandas espaciadas), pero debido a la falta de coloración en el momento de la separación, no lo barajamos como criterio de discriminación (King & Reiss, 2001). Esto lo hacemos bajo la lupa óptica Leica MX7.

Una vez obtenidas las hembras, seleccionamos los machos de las líneas RNAi y los ponemos a cruzar en tubos de poliestireno de 2,4 cm x 7,5 cm (Ø ext. x h), con un número total de 18-20 individuos (9-10 ♂ y 9-10) por tubo, con el soporte nutricional (*Iberian fly food*), suplementado con levadura en gránulos. Al cabo de 5 días, y manteniendo los tubos en un incubador (*aqua lytic*) a una temperatura de 26,7±0,3°C, nos encontramos larvas en un estadio larvario III, con un tamaño de unos 4 mm de longitud.

Llegados a este punto, depositamos las larvas en un pocillo previamente llenado con PBS (*Phosphate-Buffered Saline*) en una dilución 1x, y bajo la lupa de fluorescencia (Leica modelo MZ16 F) seleccionamos solo las que tengan el genotipo correcto, es decir, *412 101M2; dpp-Gal4 UAS GFP/línea RNAi*, gracias a la manifestación visual de fluorescencia verde (GFP) en las glándulas salivales de dichos individuos.

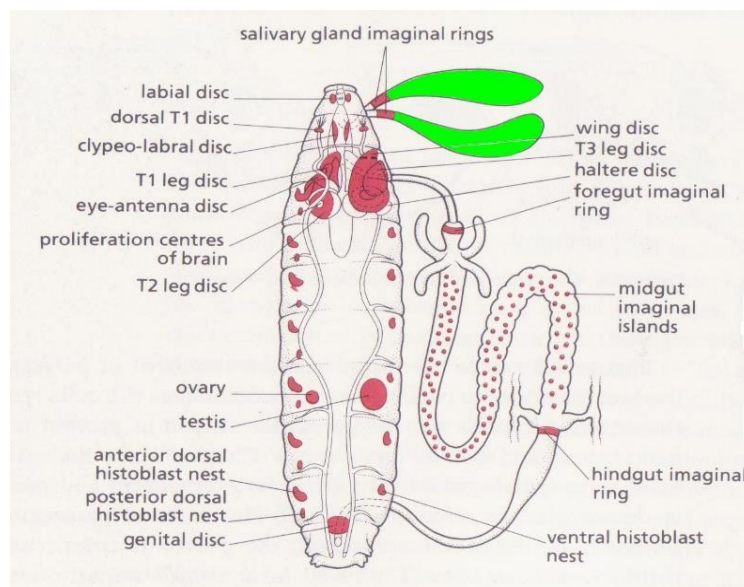


Figura 9. Discos y tejidos imaginales en *Drosophila melanogaster*. Apreciamos en verde las glándulas salivales (órganos pares situados cerca del extremo anterior), utilizadas en este trabajo como criterio de selección de las larvas con el genotipo adecuado. En rojo, podemos examinar los diversos discos/tejidos imaginales que darán lugar a tejidos u órganos específicos en adultos. Modificado de Lawrence (1992).

El siguiente paso del proceso es la disección de los discos imaginales de ala de las larvas, que realizaremos en placas de Petri de 90 mm con el elastómero Sylgard 184. Utilizando un par de pinzas de disección (marca FST, tamaño 55), obtenemos el tejido deseado mediante tracción de la mandíbula, desprendiéndose de esta forma el resto del cuerpo (destacar que debemos mantener trozos de cutícula, otros discos, etc, con el objetivo de no perder los discos de ala en el tubo eppendorf debido a su tamaño). Los discos por tanto, se ponen a fijar en un tubo eppendorf de 0,5 ml, con 100 μ l de paraformaldehído al 16% y 300 μ l de buffer PBS, en agitación continua con un agitador (Red Rotor Hofer) “overnight”. Al día siguiente, se realizan tres lavados consecutivos de la muestra con buffer (unos 300 μ l de buffer), dejando el tejido listo para el montaje.

El último paso en cuanto a la preparación de los discos sería el montaje de los mismos. Utilizando un portaobjetos, vertemos con las pinzas de disección, una gota del medio de montaje Fluoromount-G™, con DAPI (“4’,6-Diamidino-2’-phenylindole dihydrochloride”) (es un medio líquido claro que contiene una tinción nuclear fluorescente para su uso en el montaje de portaobjetos después de la tinción inmunofluorescente, utilizando anticuerpos por ejemplo contra la α -caspasa 3, así como la tinción de DNA bicatenario, con una emisión de fluorescencia máxima en 455 nm), con el código 17984-24 (info@ebioscience.com).

De esta forma pues, se sumergen los discos con ciertos restos que no nos interesan (estos restos ayudan a que los discos se depositen en el fondo del tubo por gravedad) previa estancia en hielo, en la gota de preparado. Ahora, solo resta separar los discos de ala de cutículas, discos de antena-ojo, restos de la mandíbula, etc, desechar dichos restos, y finalmente cubrir la preparación con un cubreobjetos de 24 x 36 mm (alternativamente 24 x 40 mm).

Para la visualización de los discos, utilizamos el microscopio de fluorescencia (Nikon Eclipse 90i). Este microscopio posee diversos canales de fluorescencia, encontrando en el 3^{er} canal, la longitud idónea para el marcaje con DAPI (455 nm) y en el 4^o, el canal GFP. El primero lo utilizamos a bajos aumentos (4x), para encontrar los discos en la preparación, debido a su fuerte marcaje azulado de los núcleos y el contorno celular. Una vez hallados, a más aumentos (10x, 20x, 40x), inspeccionamos la fluorescencia por la acción de la GFP, en el dominio Dpp, buscando alteraciones en el mismo, debido a acciones génicas supresoras o potenciadoras.

En caso de existir alteraciones en el fenotipo, realizamos de nuevo el cruce para comprobar que no fue fruto de un cúmulo de casualidades, y a partir de ahí, con los respectivos controles, obtenemos imágenes de mayor resolución con el microscopio confocal Leica DM IRE2.

Añadiendo un paso al proceso, una vez comprobada la existencia real de alteraciones en el patrón Dpp, realizamos ensayos de inmunohistoquímica, ya sea para apreciar otros dominios como Wg, o bien, para observar actividad apoptótica mediante el marcaje de caspasa-3 activada. Para ello, después de la fijación

con paraformaldehído, realizamos tres lavados consecutivos con PT (amortiguador fosfato de sodio, Triton X-100 0,3%) durante una extensión temporal de 1 hora. Hecho esto, los discos se incuban durante toda la noche a 4°C con el anticuerpo primario. Hemos utilizado dos tipos de anticuerpos primarios. a-Wg (un anticuerpo específico contra Wingless, en nuestro caso, localizado bajo la entrada 4D4, en DSHB ("*Developmental Studies Hybridoma Bank*", con un peso de unos 48 kDa) a una dilución 1:100. Por otro lado hemos utilizado a-cas3 (procedente de *Cell Signaling Technology*, bajo la entrada 9662, de ratón) a una dilución 1:500. Al día siguiente, realizamos tres lavados con PT durante 10 minutos, y ponemos a incubar el anticuerpo secundario durante 2 horas a temperatura ambiente. Los anticuerpos secundarios utilizados fueron: "*AffiniPure Donkey Anti-Rabbit IgG (H+L)*" de *Jackson ImmunoResearch*, bajo el código 711-005-152, procedente de mono. Por otro lado, también se utilizó "*AffiniPure Donkey Anti-Mouse IgG (H+L)*" bajo el código 715-005-151. Ambos a una dilución 1:200.

Al término de este tiempo, se realiza un último lavado durante 1 hora pudiéndose a continuación proceder al montaje de la misma forma que se ha comentado con anterioridad (Sullivan et al. 2000).

Análisis bioinformático: Un mundo movido por la curiosidad.

La identificación de nuevos genes que cooperan con *psq* y/o con *esg* en el proceso de invasión y metástasis es sólo una de las finalidades de este trabajo, pero realmente esto deja en el aire ciertos aspectos que una mente inquieta no puede dejar de indagar. Por ello, el próximo punto de exploración viene establecido por el modo de interacción/activación que presentan los genes que dieron positivo en nuestro cribado experimental. Para ello, un factor crucial y determinante de este análisis es la estructura tridimensional de las proteínas codificante por los genes positivos en nuestro estudio. Esta estructura puede dar claves sobre la función biológica en la que están implicadas, y también para entender en este caso, el modo de conexión con las proteínas Pipsqueak (PSQ) y/o Escargot (Esg).

Actualmente, los proyectos de secuenciación masiva y metagenómica entre otros, nos han dado a conocer vastas cantidades de secuencias nucleotídicas, surgiendo a la luz gran cantidad de genes que producen productos proteicos. Sin embargo, los avances en estos campos de análisis han superado con creces a la capacidad que tenemos a día de hoy de obtener la estructura a nivel tridimensional de todas las proteínas halladas de forma experimental, ya sea con técnicas de resonancia magnética nuclear (RMN), o por cristalografía de rayos X, entre muchas otras. Aun así, el ser humano está siendo capaz de superar este diferencial mediante el empleo de métodos computacionales cada vez más potentes con el transcurso de los años, para poder desenmascarar la estructura de estos productos. Estas metodologías se pueden diferenciar por factores como la complejidad computacional, el rango de tamaños de las proteínas a modelar, la precisión, la fiabilidad de los resultados, etc. (Baker et al. 2001). En este caso, se ha optado por el uso del modelado por homología, pudiendo establecer la estructura teórica de nuestra proteína (las que no están ya en las bases de datos), utilizando como referencia o plantilla la estructura de una proteína homóloga adecuada que guarda cierto porcentaje de identidad con nuestro modelo (Fiser, 2010).

Concretamente, hemos utilizado el programa SWISS-MODEL, en su versión más reciente, el servidor Workspace, el cual, incluye los avances más significativos en cuanto a algoritmos de modelado y tecnologías web (Bordoli et al. 2012). Constituye una herramienta con un modo totalmente automatizado, asumiendo los cuatro pasos indispensables del modelado (selección de la plantilla, alineación plantilla-proteína de interés, construcción del modelo y evaluación de la calidad del mismo), y además eliminando el sesgo humano, asegurando desde un punto de vista teórico, la reproducibilidad del modelo obtenido, siendo éste un requisito previo a su evaluación y comparación de la fiabilidad y precisión (Battey et al. 2007; Koh et al. 2003)

En cuanto a la forma de proceder del software, SWISS-MODEL genera las coordenadas estructurales del modelo, basándose en el mapeo entre los residuos de la proteína a formar y los aminoácidos de la plantilla/s utilizada/s. Como se ha comentado anteriormente, la plantilla guarda un porcentaje de identidad, por lo que habrá zonas, que la misma no nos pueda proporcionar la información estructural que deseamos para nuestro modelo, así que para solventar esta situación, se construyen estas regiones mediante la utilización de bibliotecas de fragmentos de estructura conocida (Zhang & Skolnick, 2005) o en su defecto, mediante la reconstrucción del espacio de restricción de estos segmentos estructurales (Peitsch, 1995). A continuación, la geometría local subóptima del modelo, por ejemplo, impedimentos estéricos ocasionados por la combinación imperfecta de fragmentos, o enlaces distorsionados, se eliminan o regularizan por minimización de energía limitada, mediante el empleo de un campo de fuerza (Van Gunsteren et al. 1995). El último paso, lo constituye la validación de la calidad del modelo general, utilizando herramientas de estimación como QMEAN (explicado en la discusión) (Benkert et al. 2009). Este último paso en el procedimiento es sin duda alguna, un aspecto central en el modelado de proteínas, ya que ello nos determinará su utilidad para aplicaciones de carácter bioquímico y biomédico.

Continuando con nuestro análisis y una vez obtenida nuestra proteína teórica (en caso de no hallarla en bases de datos), el siguiente punto que debemos abordar es como se establece esa interacción entre *pipsqueak* y *escargot* con los productos (o genes) positivos en nuestro trabajo. Cuando nos referimos a interacción, no solo incluimos una unión física entre los productos proteicos, sino que ampliamos este rasgo conceptual, a efectos de potenciación (*enhancers*) y represión (*repressors*) génica. Todo esto, nos conduciría a la utilización de unos recursos temporales muy elevados para intentar entender estas conexiones, ya que aunque existen amplias redes de interacciones biológicas documentadas y publicadas, encontramos modelos de red que no intercambian esta información de forma muy eficaz y/o fácil. La solución a esta situación, la encontramos en la herramienta EsyN, un software bionfórmico, escrito casi en su totalidad en lenguaje javascript, por la universidad de Cambridge, cuya finalidad reside en facilitar el intercambio de datos de red y agilizar el proceso de colaboración de los mismos, pudiendo crear y publicar una red de interacción biológica cualquiera, teniendo al alcance información que puede ser importada por el usuario. (Bean et al. 2014)

Con esta herramienta, introdujimos aquellos genes de interés, y el software nos puso de manifiesto que la realidad, a veces enigmática, escondía una versión más compleja de la que nosotros teníamos en mente

a priori. El programa nos devolvió una amplia red de interacciones, en forma de gráficos, en los que se presentaban nodos haciendo alusión a ciertas entidades biológicas (en nuestro caso genes) conectados con bordes de diverso tipo, cada uno con un significado concreto, pudiendo equivaler a potenciaciones, represiones o interacciones físicas (Bean et al. 2014). Añadiendo, entendimos la causa mediante la cual, algunos de los genes, manifestaban un cambio en el fenotipo, sin colaborar con *pipsqueak* o *escargot*, (ni tampoco per sé) o al menos de forma directa, ya que nos dimos cuenta de la existencia de interacciones indirectas entre estos “falsos positivos” y los genes sobreexpresados. Llegados a este punto de la investigación, teníamos un marco de referencia, a partir del cual, desenmascarar estas interacciones, que aunque en muchas ocasiones ya éramos conocedores de su existencia debido a su publicación, no tenían por qué establecer el “cómo” de dicha conexión.

Para ello, nos centramos primariamente en las interacciones físicas entre los productos de los genes, mediante *docking* proteína-proteína. En cuanto a programas para realizar *docking*, existen muchos, con diferentes algoritmos, y cada uno puede distar mucho de otro en cuanto al método de transformación hexadimensional del espacio, así como de la evaluación del complejo proteína-proteína resuelto. En este caso, hemos optado por la utilización del programa PatchDock, cuyo algoritmo fue probado con éxito en las tres últimas rondas de la Crítica de Evaluación de la Importancia de las Interacciones (CAPRI) (Inbar et al. 2005). PatchDock es un algoritmo de acoplamiento molecular basado en geometría, por tanto, está dirigido a hallar aquellas transformaciones de acoplamiento que forman una óptima complementariedad molecular. Al aplicar estas transformaciones, se inducen amplias áreas de interfaz, al igual que pequeños enfrentamientos estéricos. Además, se garantiza, una interfaz que incluye diversas características a nivel local de las proteínas en acoplamiento, de forma complementaria. Este algoritmo, por consiguiente, divide la representación del punto de Conolly (superficie excluida al solvente) (Conolly, 1983) de las moléculas en cóncavas, convexas y parches de tipo planar (Schneidman-Duhovny et al. 2005). A continuación, los parches complementarios se emparejan, generando transformaciones que se puedan considerar posibles candidatas. Si digo posible, es porque cada una de estas transformaciones se evalúa mediante funciones de puntuación considerando el ajuste de la transformación a nivel geométrico y la energía de desolvatación atómica (referencia 28 de PatchDock). Por último, un análisis de desviación cuadrática media (RMSD), nos permite discriminar y descartar las soluciones que resultan redundantes (Schneidman-Duhovny et al. 2005). Como se puede observar, este algoritmo no es de fuerza bruta, sino que impulsado por las características locales coincidentes, busca las transformaciones de forma rápida, y eficiente, destinando menos tiempo en sus cálculos y menos recursos (Schneidman-Duhovny et al. 2005).

En este punto, el programa nos devuelve una serie de posibles interacciones entre ambas proteínas, formando un complejo. Estas diversas formas de interacción que poseen ambas proteínas, están ordenadas desde un punto de vista teórico, siguiendo premisas como el acoplamiento geométrico, o la energía de desolvatación. Ahora bien, el resultado (complejo) que posea mayor puntuación del resto, no tiene por qué coincidir con la respuesta biológica, esto es, lo que ocurre realmente, ya que por ejemplo se podría dar el caso de la existencia de mayor interfase entre ambas proteínas, pero fuera de los

aminoácidos que realmente poseen un dominio de interacción. Para poder encontrar la respuesta biológica entre los complejos formados por el programa, hemos tenido que crear una serie de filtros que nos fueran restringiendo cada vez más, la zona y el modo de interacción entre los productos proteicos. El primer filtro, es el más general. Básicamente, consiste en ser coherentes con la propia distribución de la proteína, es decir, haciendo uso de diversas bases de datos como SMART, buscamos pistas que nos ayuden a esclarecer en que zona de la proteína podría ser la interacción, por ejemplo, si la proteína posee algún dominio de interacción, como pudiera ser del estilo CH₃ o CH₂. El siguiente filtro, lo encontramos en la utilización del programa Fpocket o análogos como Pocketquery. Este programa es capaz de hallar bolsillos o hendiduras dentro de la proteína. Para entender el funcionamiento del programa, debemos explicar un concepto muy importante, esferas α . A grandes rasgos, se pueden definir como esferas que entran en íntimo contacto con 4 átomos en su zona límite, sin poseer átomo alguno en el interior, y reflejando el radio de las mismas, la curvatura local definida por esos cuatro átomos (Le Guilloux et al. 2009). Por lo que a nosotros respecta, las esferas α de pequeño tamaño, se encuentran dentro de la proteína, al contrario de las grandes, en la superficie. Las que nos interesan a nosotros son las de tamaño intermedio y muy juntas, siendo éstas las que establecen los bolsillos o hendiduras en la proteína. De esta forma, el programa de forma rápida, filtra el conjunto de esferas α óptimas para la formación de un bolsillo, conforme al establecimiento de un radio mínimo y máximo. Además, el identificar las esferas α que nos interesa, viene acompañado de la identificación de los átomos que las forman, por lo que en la descripción de las esferas podemos apreciar ciertas características en base al tipo atómico que las forman, como pudiera ser la hidrofobicidad, dándonos a entender por tanto como podría ser el comportamiento de nuestro bolsillo átomos (Le Guilloux et al. 2009), y llevándonos al último paso de filtrado, la interacción específica a nivel atómico, entre ambas proteínas.

A este nivel de especificidad, utilizamos la herramienta PIC (*Protein Interactions Calculator*), un servidor que es capaz de reconocer diversos tipos de interacciones y con diversos criterios de elección para cada tipo como describiré a continuación: En primer lugar, tenemos las interacciones hidrofóbicas, considerándose residuos aptos para tales interacciones, ALA, VAL, LEU, ILE, MET, PHE, TRP, PRO y TYR, siempre que no disten entre ellos más de 5 Å (Tina et al. 2007). A continuación encontramos puentes disulfuro, enlace producido entre dos átomos de azufre de dos cisteínas, dentro de un espacio de 2,2 Å. Otra interacción arquetípica, son los enlaces de hidrógeno, cuyo criterio aparece diferencial según sea para oxígeno y nitrógeno, con un corte distal entre el donante y el receptor de 3,5, o para el azufre con un valor de 4. En cuanto a las interacciones iónicas, las encontramos entre pares de los residuos ARG, LYS, HIS, ASP y GLU, siempre que la distancia sea menor de 6 Å. Otro tipo de interacción a estudio serían las interacciones entre anillos aromáticos, entre un par de centroides de anillos fenilo cuya distancia preferencial la encontramos entre 4,5 y 7 Å. Además, también encontramos interacciones sulfuro-aromáticas entre los átomos de azufre tanto de cisteína como de metionina, con los anillos aromáticos de PHE, TYR y TRP, a una distancia menor de 5,3 Å. El último tipo de interacción que el servidor es capaz de reportar, es la interacción catión- π , dando lugar cuando una cadena lateral catiónica como la de la lisina

o arginina, se encuentra próxima de una cadena lateral aromática como la de Phe, Tyr o Trp, dentro de una distancia menor de 6 Å (Tina et al. 2007).

Gracias a estos pasos consecutivos en la forma de proceder, llegamos a una predicción fiable de como ocurren las interacciones entre las proteínas, aunque expongo de manera clara que esto es solo una predicción, donde tenemos en cuenta las características de los productos a analizar aislados, cosa que no ocurre a nivel biológico, donde existen efectos distorsionadores, que podrían resultar en una solución bien diferente a la teórica aquí expuesta.

Resultados: Realización del screening.

El número total de líneas RNA_i examinadas en el screen ascienden a 800. Ahora bien, aquellas que presentaron cambios en el fenotipo debido a diversas causas fueron el 6,37% del total de líneas, lo que equivale a 51 genes inhibidos por RNA_i. De este porcentaje, debemos desechar aquellas líneas que mostraron cambios en el fenotipo per sé, es decir, la acción de la inhibición del gen en la línea RNA_i repercute intrínsecamente en la modificación del fenotipo sin la presencia de *psq* y/o *esg* (el control negativo presenta modificación del fenotipo). Estas líneas representan el 43% (del 6,37 % que muestran cambios), asumiendo un número igual a 22 líneas. Añadiendo a todo esto, tenemos que, 15 líneas de esas 51 líneas, es decir, el 29,41%, aun no las hemos comprobado de forma tal que podamos clasificarlas como óptimas en nuestro trabajo. Por último, tenemos que 14 líneas, es decir, el 27,45%, del total de líneas con cambios fenotípicos (o el 1,75% del total de líneas analizadas), muestran el patrón alterado debido a la acción de *psq* y/o *esg*. La siguiente subdivisión que se nos presenta, viene por la determinación de aquellas líneas que muestran un cambio fenotípico gracias a la acción sinérgica con *psq*, o por el contrario, es debido a la acción coordinada con *esg*.

A continuación, se expone en la siguiente tabla (tabla1), 29 de los 51 genes que presentaron en un principio cambio en el fenotipo con respecto al control. Entre ellos, destacamos los 14 genes (en rojo) que a día de hoy hemos podido corroborar que existe un efecto cooperativo con *psq* o *esg*, observando la variopinta funcionalidad de cada uno de los productos correspondientes, con las patologías asociadas en el ser humano en el caso de los ortólogos, cuando la expresión no es la adecuada. Encontramos también, genes cuya cooperación ha sido descartada, y algunos otros que aún a día de hoy no hemos podido corroborar su sinergia con *psq* y/o *esg*. Una observación crucial que podemos manifestar a través de estos datos expuestos, es el tipo de cambio de fenotipo que los genes/productos realizan. Estos cambios pueden darnos a entender que la ausencia del gen a estudio, potencia la invasión y metástasis tumoral, como pudiera ser por ejemplo, el gen *br* (*broad*), presente en un alto porcentaje de cánceres de ovario en humanos. La parte antagónica la tenemos en una situación cuya ausencia génica normaliza el patrón, existiendo un estado de represión, y dándonos a entender que el gen en cuestión, esta intrínsecamente relacionado con promulgar dicha invasión celular y metástasis. Un ejemplo de ello, sería el gen *Nrx-IV*, presente en patologías como esquizofrenia, y el síndrome de Guilles de la Tourette.

Algo pendiente por demostrar a día de hoy, es si la magnitud de los efectos a nivel individual (de cada gen reprimido), es debido al efecto conjunto con *pipsqueak* o bien con *escargot*, siendo esta discriminación, uno de los objetivos de esta investigación en el futuro.

Nº exp.	Nombre	Simbol o	Nº CG	Enfermedad humana asociada al ortólogo	Cambio de fenotipo observado	esg/psq
310	Subunidad 11 Complejo Mediador.	MED11	CG6884	Catarata congénita de Volkmann. Indicador en cáncer de endometrio.	↑ apoptosis.	¿?
63	Sub. 29kD proteasoma	Pros29	CG9327	-	↑ apoptosis.	
380	Requerido para dif. celular 1	Rcd-1	CG14213	Ninguna enfermedad asociada.	↑ apoptosis.	¿?
375	-	CG2256	CG2256	-	↑ apoptosis.	¿?
505	Sub. 4 Comp. Med.*	MED4	CG8609	Ninguna enfermedad asociada.	Patrón dpp muy desorganizado.	¿?
706	Sub.31 Comp. Med.	MED31	CG1057	Ninguna enfermedad asociada.	Patrón dpp desorganizado.	¿?
205	Rho1	Rho1	CG8416	-	↑ apoptosis.	
		UAS-APP		Alzheimer, esquizofrenia, autismo, hemorragia cerebral.	Normaliza dpp.	¿?
	Pten	Pten	CG5671	Cáncer de mama y próstata.	dpp desorg.	¿?
136	escargot	esg	CG3758		Normaliza dpp.	¿?
	pixie	pix	CG5651	-		¿?
158	Receptor hormona 39	Hr39	CG 8676	-	Patrón dpp desorganizado.	¿?
141	Canal Ear K ⁺	Elk	CG5076	Cáncer de próstata, vitreorretinopatía exudativa 1.	Normaliza dpp.	¿?
246	Neurexina IV	Nrx-IV	CG6827	Esquizofrenia, Guilles de la Tourette, epilepsia, autismo, etc	Normaliza dpp.	¿?
165	Like Sm 1	LSm1	CG4279	Presente en más del 80% de tumores pancreáticos.	Normaliza dpp con apoptosis.	¿?
364	lethal (3) malignant brain tumor	l(3)mbt	CG5954	Cáncer de mama.	Patrón anormal de dpp. Invasivo.	¿?
384	cropped	crp	CG7664	Ataxia espinocerebelosa tipo 12	Patrón dpp muy expandido.	¿?
119	Sintaxina 18	Syx18	CG13626	-	↑ apoptosis.	
182	broad	br	CG11491	Cáncer de ovario	↑ apoptosis. Sobrecrecimiento.	¿?
		Rps3		-		¿?
143	Canal-b cloro	CIC-b	CG8594	osteopretosis	↑ apoptosis.	¿?
		gho		-		¿?
		hay		-		¿?

	Heat shock protein cognate 4	Hsc-70-4	CG4264	-	↑ apoptosis.	
		Uba1		-		
		cas		-		¿?
		Timp		Marcador pronóstico en cáncer colorectal y mama.	↑ apoptosis.	¿?
		Cenp-C		-		
	Proteína Rab 11	Rab11	CG5771	-	dpp desorganizado. ↑ apoptosis.	¿?

Tabla 1: Genes que mostraron cambios en el fenotipo. En rojo, aquellos genes que hemos corroborado su cooperación con *psq* y/o *esg*. En amarillo aquellos genes aun no testados. En verde, presentamos algunos genes que han sido descartados. En azul, aquellos genes que en un estado de inhibición, muestran características represoras en cuánto al fenotipo.

Debido a la limitada extensión legal del trabajo nos ceñiremos a uno de los genes comprobados hasta ahora, el gen *dL(3)mbt*, cuyo análisis nos reportará la actuación de terceros genes hasta llegar a la explicación que buscamos.

En lo concerniente al gen *dL(3)mbt* observamos en la figura 10B, el disco imaginal de ala del sujeto con genotipo *412 101M2 GS(2)101M2; dpp-Gal4 UAS GFP/ dL(3)mbt*, un patrón de Dpp altamente anómalo sobretudo en la parte próxima al nótum. Discernimos también una elevada invasión celular, como se puede observar en la difuminación de las células que expresan fluorescencia pertenecientes al patrón dpp. Por el contrario, en el control negativo, con genotipo *dL(3)mbt/+; dpp-Gal UAS GFP* (figura 10C) observamos un patrón normal (más estrecho) y sin atisbos de presentar un fenómeno de invasión celular. En la imagen 10A podemos ver claramente un patrón mucho más grueso, siendo este nuestro control positivo para ambos genes (*psq* y *esg*).

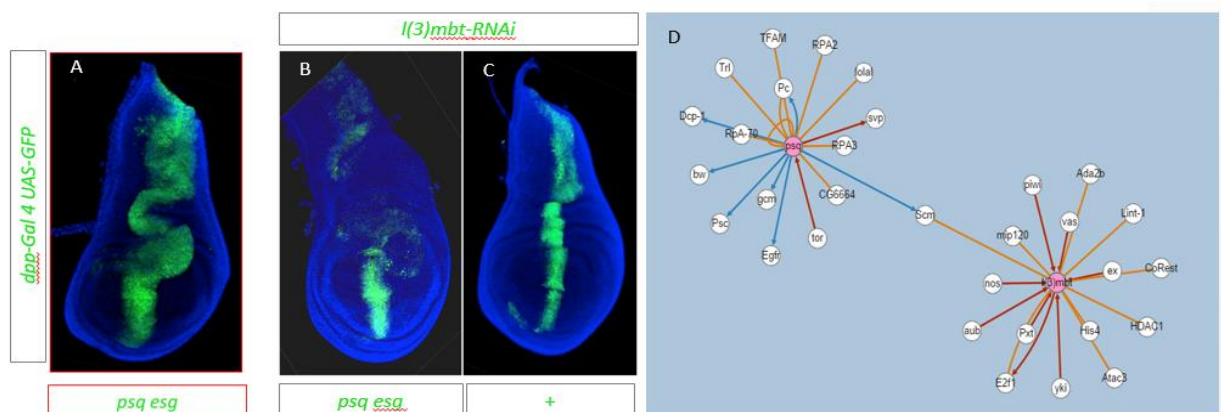


Figura 10. Fenotipo mostrado por *dL(3)mbt*.

A) Control positivo general. B) Genotipo: *412 101M2; dpp-Gal4 UAS GFP/ dL(3)mbt*. Patrón *dpp* anormal e invasivo. C) Control negativo. D) Representación con la herramienta EsyN, acerca de la interacción entre *dL(3)mbt* con *Scm* (unión física), y éste último con *psq* (potenciación génica).

Para intentar explicar cómo ha sido posible el establecimiento de este fenotipo, es decir, el mecanismo subyacente que interrelaciona la inhibición del gen *dL(3)mbt* y la sobreexpresión de *psq/lola* y *esg* con

este fenotipo neoplásico, utilizamos la herramienta ESN. A la luz de los resultados reportados por ESN (figura 10D) pudimos establecer varias conexiones entre los diferentes genes.

A priori, parece ser que la interacción ocurre entre *psq* y *dL(3)mbt* a través de un gen intermediario, *Scm* (explicado en la discusión), dejando en un plano inexistente a *esg*. Este será nuestro marco de referencia a partir del cual, desenmascarar un hipotético mecanismo de acción que vaya a dar lugar a un fenotipo tumoral, planteándose estos genes como “actores” de primer orden.

Discusión

***dL(3)mbt*, el aliado perfecto**

Drosophila Lethal (3) Malignant Brain Tumor (dL(3)MBT) es considerada una proteína supresora de tumores, cuya atribución destacable es la regulación de la proliferación celular en el cerebro, concretamente en los lóbulos ópticos. Para ello, realiza una serie de funciones como la compactación de histonas, establecimiento de aislantes de cromatina, regulación de la ruta de Hippo, y la represión del programa de la línea germinal en células somáticas (flybase.org).

A nivel génico, *dL(3)mbt*, está constituido por dos intrones (de 769 bp y 69 bp) y tres exones, con una extensión global de 4968 pb, a lo que hay que añadir una cola poli (A) predicha de 200 pb (Whale et al. 1992), formando un transcrito de aproximadamente 5,25 kb. Aun siendo esto así, los tres exones forman un marco de lectura abierto de 4431 pb, lo que da pie a una proteína de 1477 aminoácidos, cuyo peso establecido es 163 kDa (Wismar et al. 1995). Navegando un poco por las características del producto proteico de este gen, encontramos que la proteína se caracteriza por un alto contenido en el iminoácido prolina (9,3%), así como de los aminoácidos serina (8,8%) y lisina (8,6%). Entre los 200 primeros aas (desde el extremo N-terminal), discernimos una elevada proporción de aminoácidos de carácter hidrofóbico. Junto a esta región, entre las posiciones 240 y 510, observamos una región rica en lisina (15,5%) (Wismar et al. 1995). Elevándonos en la jerarquía estructural de la proteína, dilucidamos tres motivos con extensiones entre 99 y 105 aas, denominados repeticiones MBT (“*malignant brain tumor*”), enmarcados entre las posiciones 819 y 1133 de la proteína, siendo éstos muy importantes como veremos, ya que pueden formar homodímeros. En cuanto a la región C-terminal, observamos un alto porcentaje del aminoácido aspartato (36%). Por lo que se refiere a la estructura tridimensional de la proteína, encontramos un 35% de α -hélices y un 36% de láminas β , intercaladas con colas y giros (Wismar et al. 1995). Ahora bien, lo que le ofrece la posibilidad de actuar como un factor de transcripción, es un dominio de dedo de zinc del tipo C_2/C_2 entre los aminoácidos 730 y 752, interrumpido por el segundo intrón (Wismar et al. 1995). Pero sin duda esta proteína presenta un dominio vital entre los aminoácidos 1278 y 1367, el dominio PNT-SAM. Este dominio altamente conservado, se encuentra dentro de un subconjunto de los factores de transcripción Ets, como Pnt-P2 o Yan en *Drosophila*. Además el dominio PNT se encuentra relacionado a nivel estructural con el grupo de dominios SAM más grande, ofreciendo una

disposición tridimensional de cuatro α -hélices normalmente. Para este dominio se ha establecido una función asociativa entre proteínas, como analizaremos más adelante.

Cuando se encuentra mutado *dL(3)mbt^{ts}* (mutación sensible a temperatura, al llegar a una temperatura restrictiva de 29°C) da lugar a la sobre-proliferación de células neuroepiteliales en los lóbulos ópticos causadas por la desregulación de genes diana en la vía Salvador-Warts-Hippo (Richter et al. 2011), así como un crecimiento excesivo en las células madre ganglionares cerebrales, en larvas de tercer estadio, resultando en un fenotipo tumoral con potencial metastásico.

LINT, un complejo capaz de unirse a SCM

Estudios recientes ponen de manifiesto que existen una serie de genes desregulados en presencia de la mutación *dL(3)mbt^{ts}*, específicamente podemos encontrar 102 genes desregulados, los que contribuyen a la denominada firma del tumor cerebral maligno (MBTS) (Janic et al. 2011). 32 de estos genes dan lugar a proteínas de gran relevancia en la función germinal, pudiendo la inhibición de alguno de ellos rescatar el fenotipo tumoral típico que observamos en individuos *dL(3)mbt^{ts}* (Janic et al. 2011). Éstos no son los únicos afectados, pues se ha reportado que un grupo de 7 genes, blancos importantes de la vía de señalización de Salvador-Warts-Hippo (SWH), se encuentran desreprimidos en el tejido tumoral, dando por tanto la sobreexpresión de alguno de ellos a una réplica fenotípica de *dL(3)mbt^{ts}* (Richter et al. 2011). Ello nos sugiere que estos genes son objetivos de forma directa de *dL(3)mbt*.

Algo importante que no podemos dejar pasar para realizar un análisis óptimo sobre la funcionalidad de este gen es, la capacidad asociativa que posee, esto es, la posibilidad de formar complejos. *dL(3)MBT* se asocia con el complejo MybMuvB/dREAM a niveles muy bajos (Lewis et al. 2004). Este complejo es similar al que forma el ortólogo humano L3MBTL1, el cual forma parte de un complejo con pRb, HP1y y H1b, sugiriendo un mecanismo para reprimir la transcripción mediante el aumento de la compactación de nucleosomas en los genes diana (Trojer et al. 2007). Al realizar mutaciones en los componentes centrales del complejo MybMuvB/dREAM, se observó que no se daba lugar al establecimiento de tumores en el cerebro de las larvas, planteándose la posibilidad de que la represión de los genes importantes en la mantención del tumor se podría dar con un complejo diferente con *dL(3)MBT* (Meier et al. 2012)

El complejo identificado y caracterizado que reprime un conjunto de genes específicos de la línea germinal, desregulado en condiciones tumorales en el cerebro de larvas, se ha denominado LINT (Meier et al. 2012). Este complejo posee tres subunidades principales, *dL(3)MBT*, *dLint-1* y *dCoREST* (isoforma de 95 kDa), siendo los tres participes, necesarios para mantener el estado represivo de los genes MBTS (Meier et al. 2012). Para determinar la presencia del complejo en el cerebro de las larvas, se prepararon extractos de las mismas, realizando una inmunoprecipitación con anticuerpo *dLint-1*, resultando en la coprecipitación de *dL(3)MBT* y *dCoREST* (Meier et al. 2012). Algo curioso es que en el análisis de purificación por inmutafinidad, también se detectaron la interacción de la desmetilasa *dLsd1*, y de la histona desacetilasa *dRpd3* (HDAC1) con la proteína *Lint-1*, por lo que analizaron si el modo de proceder del complejo se basaba en la arquetípica función de modificación de histonas para silenciar los genes diana.

Esto en principio se descartó ya que la desrepresión de los genes diana no se producía al agotar las dos modificadoras de histonas, aunque se podía apreciar H4K20me2 cerca de los genes diana del complejo. Según argumentan los investigadores de tal experimento, esta modificación es abundante a lo largo de todo el cromosoma, incluso en regiones control, a las que LINT no se une, por lo que no sería contradictorio con lo expuesto.

En cuanto a la forma de proceder, parece ser que el modo de represión de LINT es al menos en gran medida independiente de las actividades represivas en la modificación de histonas, sugiriendo pues, que LINT podría inhibir el proceso de transcripción al no permitir el acceso de la RNA polimerasa tipo II (o factores de transcripción a sus promotores), gracias a la observación de que las subunidades del complejo se unen de forma predominante cerca de los sitios de inicio transcripcional (TSS) (Meier et al. 2012). Mediante un análisis BioTAP-XL (un enfoque de entrecruzamiento de la cromatina que utiliza la espectrometría de masas para identificar componentes de complejos proteicos, junto con la secuenciación de alto rendimiento para identificar los componentes de RNA y sitios de unión al DNA (Artyom et al. 2016)) se co-purificaron tanto dL(3)MBT, como Lint-1 y CoRest junto con SCM, sugiriendo que existe una interacción funcional entre el complejo LINT y el gen *Scm* (*Sex Comb on Midleg*).

Sex comb on midleg es un gen que codifica una proteína que se enmarca dentro del grupo Polycomb (PcG), cuya función reside en formar parte de complejos multiméricos, necesarios para mantener la represión de diversos genes, como los homeóticos durante la fase de desarrollo. *Scm* es una proteína con una extensión de 877 aminoácidos (94 kDa) (Peterson et al. 2004). Contiene varios motivos de interacción de cromatina, incluyendo dos repeticiones MBT entre los residuos 175-273 y 281-382, y un dominio de dedos de zinc (Bornemann et al. 1998). Además, contiene un motivo SAM ("sterile α motif") en su extremo C-terminal (residuos 806-876) implicado en homopolimerización y heteropolimerización (Peterson et al. 1997).

Para profundizar con la idea de la unión física entre dL(3)MBT y SCM, realizamos un análisis bioinformático exhaustivo, intentando establecer como estas dos proteínas interaccionan. En primer lugar debemos ser coherentes con la elección de las zonas dentro de ambas proteínas, con predisposición a formar homo/heterodímeros. En ambas proteínas encontramos la existencia de dominios de interacción proteína-proteína. En el caso de dL(3)MBT, encontramos el dominio PNT/SAM, y en el caso de SCM, el dominio SAM. Para este último dominio, si encontramos la estructura tridimensional caracterizada cristalográficamente por difracción de rayos X, con la entrada PDB "5J8Y" (aunque viene formando un tetrámero, eliminamos las cadenas que no nos sirven para nuestras predicciones), no siendo el caso con el dominio PNT/SAM de dL(3)MBT.

Para subsanar esta situación, empleamos el modelado por homología con la herramienta bioinformática SWISS-MODEL. Como plantilla, utilizamos parte de la estructura tridimensional de la proteína caracterizada "*polyhomeotic-proximal chromatin protein*", establecida bajo la entrada "1PK1", fragmento de la cual comparte un 28,38% de identidad en la secuencia con respecto a nuestro modelo. Para

establecer la calidad del modelo obtenido existen una serie de parámetros como torsiones entre residuos, C β , QMEAN y GMQE, aunque solo me fijaré en este caso en los parámetros QMEAN y GMQE. QMEAN (Benkert et al. 2009) es un estimador compuesto, basado en diferentes propiedades geométricas que proporciona estimaciones de la calidad absoluta tanto a nivel global (como sería el modelo per sé), como a nivel local (por residuo) sobre la base de un solo modelo. Las puntuaciones cercanas a cero, indican una buena calidad en el modelo, mientras que puntuaciones por debajo de -4 indican una baja calidad del modelo. En este caso, la puntuación es de -0,95.

Por otro lado, GMQE (Estimación global de la calidad del modelo) es una estimación de la calidad que combina propiedades de la alineación objetivo-plantilla y el método de búsqueda de la plantilla. La puntuación GMQE resultante se expresa como un número entre 0 y 1, lo que refleja la precisión esperada de un modelo creado con esa alineación y plantilla. Los números más altos indican una mayor fiabilidad. Una vez que se construye un modelo, el GMQE ((1) en la figura anterior) se actualiza para este caso específico al tener también en cuenta el puntaje QMEAN del modelo obtenido para aumentar la confiabilidad de la estimación de calidad. La puntuación obtenida es 0,54. Con esto, podemos decir que hemos obtenido un modelo bastante fiable sobre la estructura tridimensional del dominio PNT/SAM de dL(3)MBT (Figura 11A).

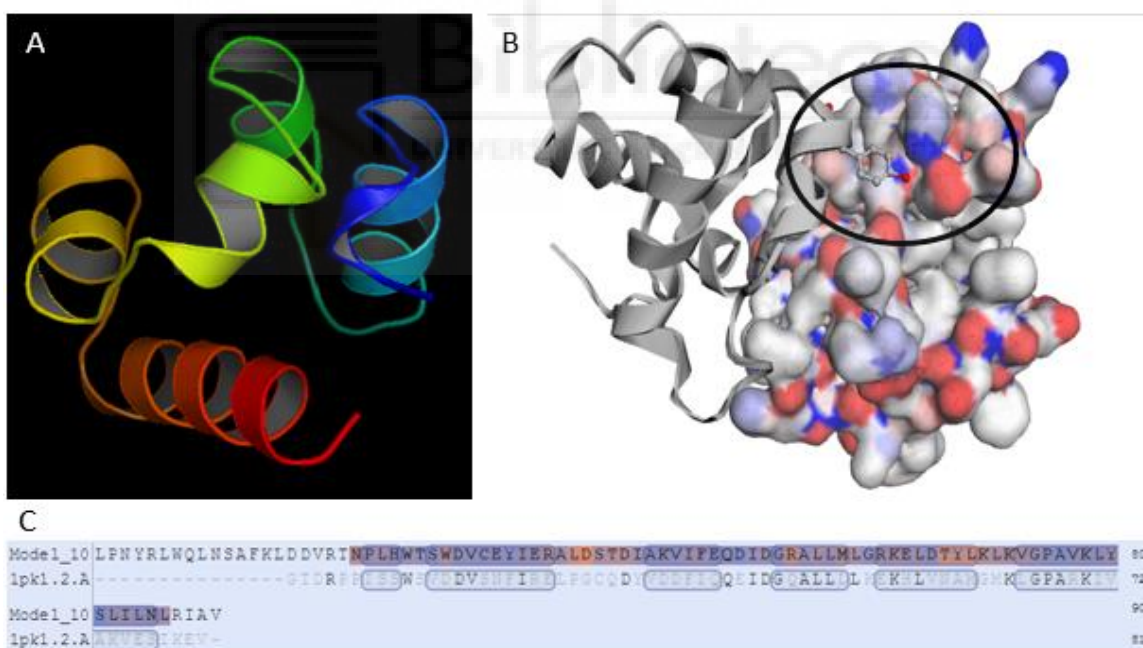


Figura 11. Modelo tridimensional del dominio PNT de dL(3)MBT.

A) Representación gráfica del dominio, usando el programa PyMOL. B) Ejemplo de “hot spot” (TYR850) aportado por el programa FPocket. C) Comparación de identidad de secuencia entre el modelo y la plantilla utilizada.

El siguiente paso viene establecido por la realización del “docking” entre ambos dominios, utilizando el servidor PatchDock. Él mismo nos aportó una serie de posibles soluciones, analizando entre las diez primeras puntuaciones más elevadas. De forma análoga, se utilizó el servidor SwarmDock (Torchala et al. 2013) para comparar las soluciones. Por otra parte, encontramos realizando una busca bibliográfica, una

posible forma de unión entre los dominios, en base a los resultados cristalográficos obtenidos por Frey et al., con los dominios SAM de las proteínas Sfmtb y SCM, lo que nos proporcionó un punto de apoyo importante para poder discriminar soluciones sin valor biológico. A continuación, para poder discriminar con mayor precisión aquellas soluciones no factibles, se utilizaron los programas Fpocket y Pocketquery, en busca de bolsillos y/o puntos calientes (“Hot Spots”) en los dominios que nos dieran pistas acerca de cómo podría darse la interacción biológica, intentando discernir la realidad, de los resultados posibles. Se obtuvieron un total de 7 “puntos calientes”. Por parte del dominio SAM: LYS837 ($\Delta G = -8,01$ Kcal/mol), LEU841 ($\Delta G = -4,26$ kcal/mol) y TYR850 ($\Delta G = -3,05$ kcal/mol). Por parte del dominio PNT: LEU1363 ($\Delta G = -4,12$ kcal/mol), GLU1341 ($\Delta G = -3,95$ kcal/mol), LEU1361 ($\Delta G = -3,17$ kcal/mol) y LEU1335 ($\Delta G = -3,08$ kcal/mol). La solución final, mostrada en la figura 12A, corresponde con la solución 74b del servidor swarmDock, poniendo de manifiesto, que no siempre la/s primera/s posición/es, albergan la solución biológica. Se utilizó una RMSD (medida de la distancia promedio entre los átomos de las proteínas superpuestas) de 4 Å, un valor óptimo y aconsejado para el acoplamiento entre proteínas (desaconsejado a nivel intraprotéico).

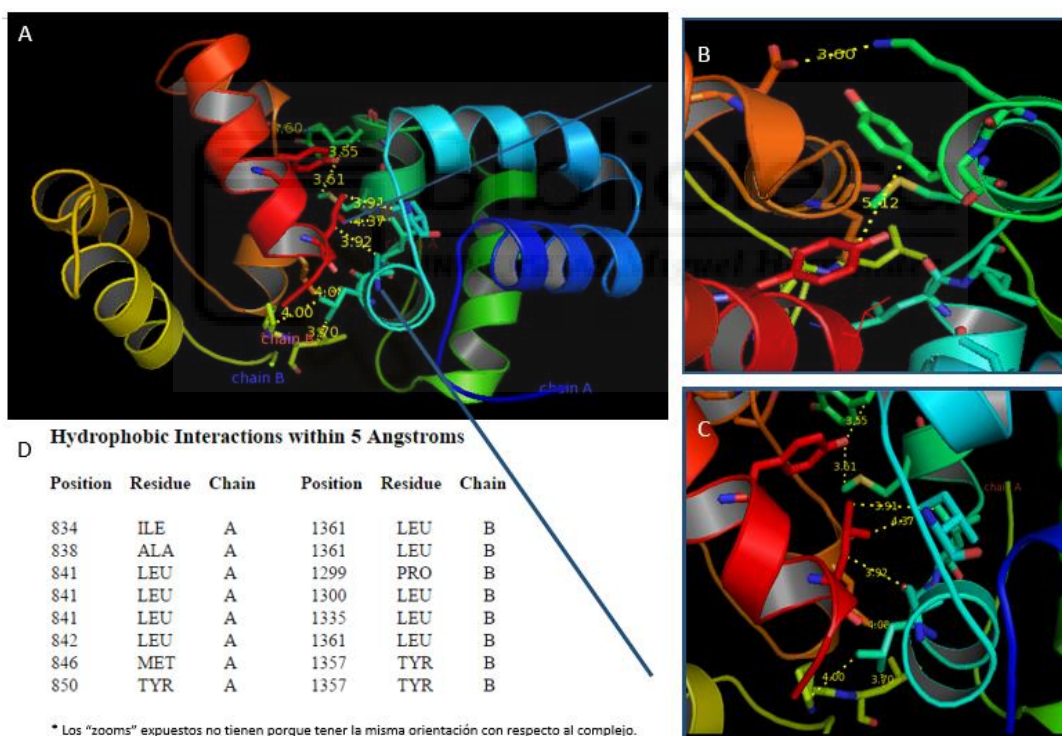


Figura 12. “Docking” entre los dominios PNT y SAM.

A) Representación gráfica del acoplamiento entre los dominios. Visualizado con PyMOL. B) Zoom de las interacciones iónicas con sus respectivas distancias. C) Zoom de las interacciones hidrofóbicas ocurridas entre los dominios. D) Tabla de las interacciones hidrofóbicas.

Para establecer las interacciones que se dan en nuestro modelo predictivo, utilizamos el servidor PIC. En la figura 12D, podemos ver las interacciones hidrofóbicas que ocurren entre los dos dominios. Este tipo de interacción, se establece entre residuos apolares, siendo éstos aquellos residuos que carecen de enlaces con una distribución desigual de electrones entre los núcleos de los átomos que forman el enlace, interaccionando por tanto, a través de fuerzas de Van der Waals (atracciones electrostáticas débiles que

se dan como consecuencia de los dipolos instantáneos creados por la fluctuación de la nube electrónica de los átomos). Aunque estas fuerzas tienen una fuerza de atracción débil (entre -0,5 y -3 kcal/mol) se pueden considerar las fuerzas más importantes en cuanto a establecer el plegamiento adecuado entre los dominios. Observamos que la LEU841 del dominio SAM de SCM (representado bajo la cadena A) forma un bolsillo hidrofóbico importante interactuando con múltiples residuos del dominio PNT de dL(3)mbt (cadena B). En la figura 12C observamos las interacciones hidrofóbicas con sus respectivas distancias.

El servidor PIC además nos aportó enlaces de hidrógeno. Los puentes de Hidrógeno, se forman por átomos de hidrógeno localizados entre átomos electronegativos. Cuando un átomo de hidrógeno se encuentra unido de forma covalente, a un átomo electronegativo (Oxígeno o Nitrógeno por ejemplo), asume una densidad (δ) de carga positiva, debido a la elevada electronegatividad del átomo vecino. Esta deficiencia parcial en electrones, hace a los átomos de hidrógeno susceptibles de atracción por los electrones no compartidos en los átomos de oxígeno o nitrógeno. Encontramos enlaces entre las cadenas laterales de los aminoácidos, así como entre el esqueleto polipeptídico y las cadenas laterales.

Añadiendo, encontramos una interacción iónica. Esta es debida a que en condiciones fisiológicas, el residuo LYS849 (Cadena A) se encuentra cargado positivamente al tratarse de un aminoácido básico, y el GLU1341 (cadena B) se encuentra cargado negativamente. Al poseer carga opuesta claro está que se atraen por fuerzas electrostáticas. Por último, encontramos un enlace entre residuos aromáticos, el aminoácido TYR850 (cadena A), forma enlace con TYR1357 a una distancia de 5,12 Å (figura 12B). De esta forma, manifiesto teóricamente el modo de unión entre los genes *Scm* y *dL(3)mbt* a través de sus respectivos dominios SAM y PNT.

Unión física y funcional entre SCM y PSQ

No debemos olvidar, que las líneas utilizadas en este proyecto de investigación sobreexpresan el gen *psq* (en realidad *psq/lola* y también *esg*), y que por tanto, en el establecimiento del fenotipo tumoral este gen ha colaborado. Numerosas son las evidencias que manifiestan la interrelación entre *Scm* y *psq*. En primer lugar, como observamos en la imagen 13A (María Domínguez, no publicado), el individuo con genotipo *ey>Dl>88A8; Scm/+*, (recordamos que *88A8* es el la línea llamada *Eye-ful* cuyo genotipo se debe a la inserción del transposón GS en un intrón del gen *lola* (*longitudinal lacking*) y que a la vez induce la expresión del gen vecino *psq*) muestra un crecimiento hiperplásico brutal a nivel ocular. En este ejemplar se está sobreexpresando el gen *psq* (también *lola*), mientras que el gen *Scm* se encuentra en un estado de deficiencia funcional, al estar en heterocigosis y por ende, expresar la mitad de la dosis génica con respecto al estado normal.

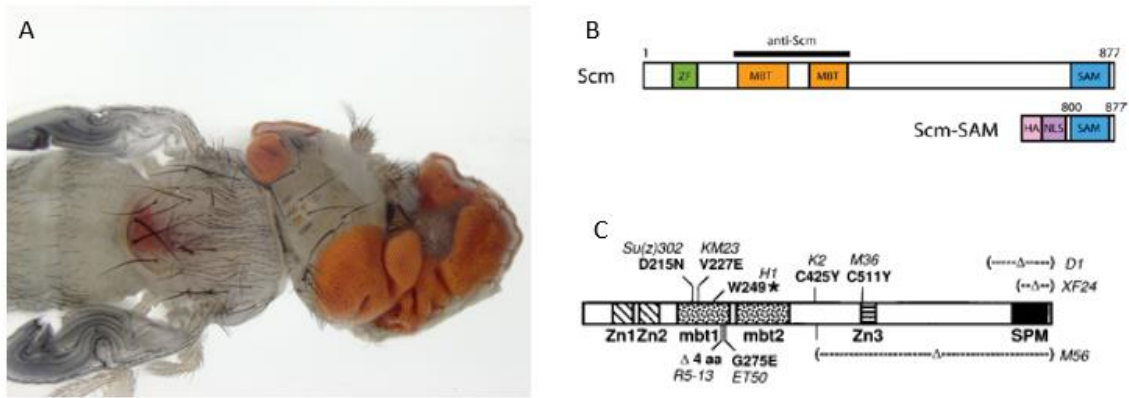


Figura 13. Interacción entre SCM y PSQ.

A) Individuo *ey>Dl>88A8; Scm/+* con fenotipo tumoral exagerado en el ojo (María Domínguez, no publicado). B) Conformación del gen *Scm*. C) Diversos mutantes del gen *Scm*.

Otro argumento que evidencia la relación génica y funcional entre ambos genes es un experimento que se realizó cruzando hembras heterocigóticas para varios alelos mutantes del gen *psq*, con machos heterocigóticos para el alelo mutante *Scm^{D1}* con la finalidad de examinar cambios en el fenotipo causado por la desregulación de genes homeóticos en las piernas transformadas, por ejemplo la producción de peines sexuales ectópicos en el segundo y tercer par de patas en la F₁ (Bornemann et al. 1998). Los machos con la mutación *Scm^{D1}* dieron un aumento de dos o tres veces más peines sexuales ectópicos con respecto al control, lo que sugiere un nivel intermedio de interacción entre los genes *Scm* y *psq* (Bornemann et al. 1998).

Como se vislumbra a partir de la figura 13C, el mutante *Scm^{D1}*, posee una delección génica que abarca todo el dominio SAM (en la imagen pone SPM), por lo tanto, ello nos lleva a plantearnos la idea de que dicho dominio es el que puede que interactúe con la proteína PSQ (ya que en los individuos de la F₁ con la mutación *Scm^{D1}* hay desrepresión de ciertos genes homeóticos al no poder unirse con *psq*). Con respecto a *psq*, encontramos que el mejor dominio candidato para unirse a SCM, es su dominio BTB.

A nivel molecular, el dominio BTB puede auto-asociarse para formar homodímeros, así como interactuar con proteínas de dominio no-BTB (Farid et al. 1998), lo que lo convierte en el candidato idóneo para formar la interacción que deseamos resolver. Con la información disponible no puedo descartar la posibilidad de que dicha unión venga promovida por alguno de los dedos de zinc que posee PSQ, pero debido a que preferencialmente guardan unión con DNA, y que el dominio BTB es un dominio de interacción proteína-proteína, me inclino por la posibilidad de que la interacción venga establecida entre el dominio SAM de SCM y el dominio BTB de PSQ.

Al igual que en el caso del dominio PNT de dL(3)MBT, no encontramos en las bases de datos, la estructura tridimensional del dominio BTB de PSQ, por lo que procedemos del mismo modo que en el caso anterior, es decir, obtenemos un modelo teórico del fragmento deseado, mediante modelado por homología. En este caso, la plantilla que utilizamos bajo la entrada PDB "4HXI" que viene a representar un complejo hetero-oligomérico, aunque solo obtenemos como plantilla para nuestro modelo, la parte concerniente

a “*Kelch-like protein 3*”, guardando una identidad con nuestra secuencia problema del 35,38% (figura 14C). En cuanto a la calidad obtenida para este modelo, obtenemos una puntuación de -2,28 con el estimador QMEAN, lo que nos indicaría en principio que la calidad del modelo es bastante inferior con respecto al primer modelo (PNT de dL(3)MBT), pero debido a que los otros estimadores que analizan diversas variables, como por ejemplo las torsiones, han obtenido puntuaciones más óptimas que en el caso anterior, ponen de manifiesto una mejor calidad global del modelo (GMQE) con respecto al otro modelo, otorgándole una puntuación de 0,68. Si tenemos en cuenta que la GMQE, establece valores entre 0 y 1, podemos decir que hemos obtenido un modelo bastante óptimo (figura 14B), para la información que encontramos disponible.

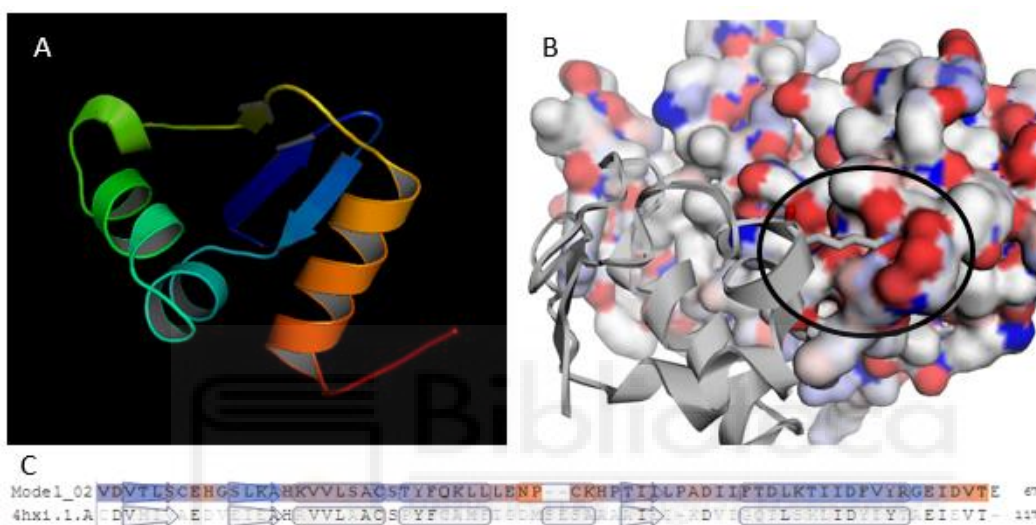


Figura 14. Modelo tridimensional del dominio BTB de PSQ-A.

A) Representación gráfica del dominio, usando PyMOL. B) Ejemplo de “hot spot” (LYS831).

C) Comparación de identidad de secuencia entre el modelo y la plantilla utilizada.

Para la realización del emparejamiento inter-dominio, ahora debemos tener en cuenta un valor añadido, es decir, según la hipótesis que barajamos, el dominio SAM de SCM se une al dominio PNT de dL(3)MBT, por lo que realizo el “*docking*” en base a los resultados obtenidos en el anterior análisis, es decir, teniendo en cuenta también la disposición espacial del dominio PNT.

Para la simulación teórica del emparejamiento, además del servidor PatchDock, se utilizó de forma análoga el servidor ClusPro (Kozakov et al. 2017). Como en el caso anterior, se utilizó una RMSD de 4 Å. Curiosamente, tras el análisis, entre las 15 primeras opciones, se ponía de manifiesto en la mayoría de las hipotéticas soluciones una interacción global bastante importante entre los dominios PNT de dL(3)MBT y BTB de PSQ, dejando en un segundo y débil plano (en algunos casos incluso inexistente) al dominio SAM de SCM. Utilizando como argumento que existe una interacción al menos de nivel intermedio entre el dominio SAM de SCM y PSQ (Huang et al. 2002), junto con el añadido de que no se ha reportado la existencia de interacción alguna entre PSQ y dL(3)MBT, al menos de forma directa hasta el momento, decidí descartar estas opciones, sosteniendo tres posibles resultados. Por consiguiente, se evaluaron las tres posibles soluciones con los programas Fpocket y Pocketquery, en busca de puntos potenciales que

podieran dar lugar a interacciones viables entre los dominios. Solo una, la correspondiente a la solución 8 (reportada por ClusPro), nos aportaba un número lo suficientemente alto de puntos calientes, hasta 7, que pudiera darnos a entender una buena interacción. Estos residuos fueron, por parte del dominio SAM: LYS831 (con una energía libre (ΔG) de -12,59 kcal/mol (cuanto más negativo sea el escalar, más influyente es el residuo en el complejo)), LYS854 (ΔG = -9,22 kcal/mol), GLU816 (ΔG = -6,88 kcal/mol) y VAL824 (ΔG = -3,9 kcal/mol). Por parte del dominio BTB encontramos: GLU96 (ΔG = -10,72 kcal/mol), ASP84 (ΔG = -10,68 kcal/mol) y LYS62 (ΔG = -7,81 kcal/mol). En cuanto al modelo general, se establecieron 4 cálculos de acoplamiento, con 4 conjuntos diferentes de parámetros energéticos, estableciendo un coeficiente ponderado en términos energéticos. Dicho coeficiente fue de -627,8 kcal/mol (Kozakov et al. 2017).

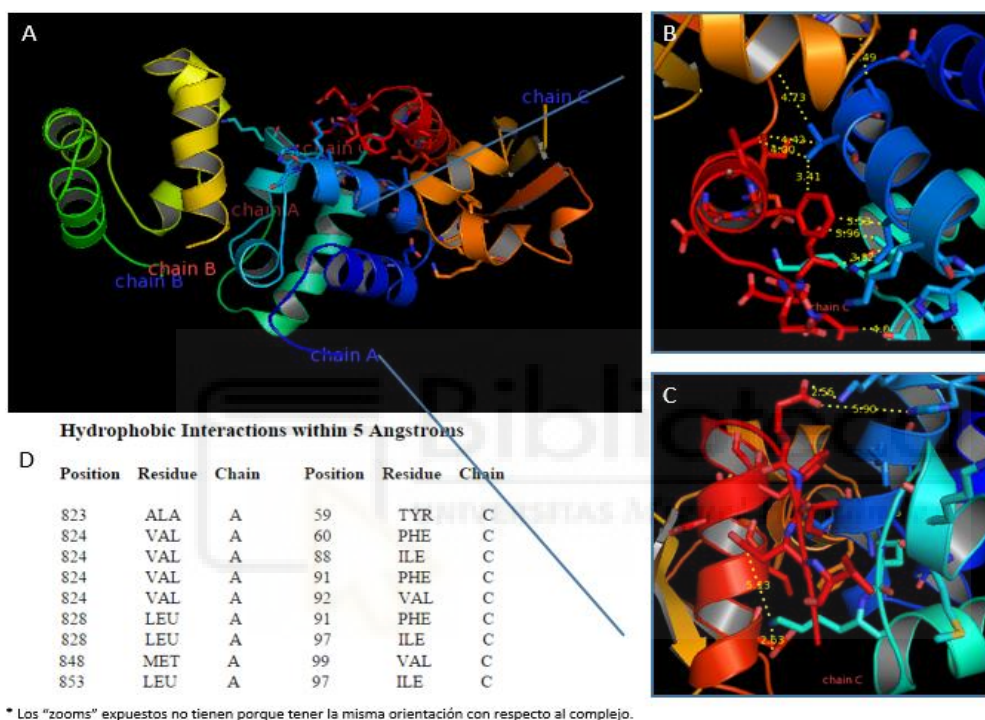


Figura 15. "Docking" entre los dominios BTB y SAM.

A) Representación gráfica del acoplamiento global entre los dominios (teniendo en cuenta al dominio PNT de dL(3)MBT). B) Zoom de las interacciones hidrofóbicas dadas. C) Zoom de las interacciones iónicas dadas entre los dominios. D) Tabla con las interacciones hidrofóbicas.

En la imagen 15A podemos apreciar gráficamente el resultado final del acoplamiento. Especificando las interacciones, resaltamos en las de tipo hidrofóbico (figura 15B), la formación de un buen bolsillo hidrofóbico por parte de VAL824 del dominio SAM de SCM (cadena A), con varios residuos del dominio BTB de PSQ (cadena C). Otras fuerzas que actúan en este caso son; enlaces de hidrógeno (véase el anexo) en todas sus vertientes (cadena lateral-cadena lateral, cadena lateral-cadena principal, cadena principal-cadena principal). También se describen interacciones iónicas (Figura 5678 C), destacando: GLU816-LYS62, LYS831-GLU96, HIS832-GLU96, LYS854-ASP84 y LYS854-ASP98 (siendo siempre el primer miembro de la pareja de la cadena A y el segundo, de la cadena C) encontrándose todas ellas a una distancia menor de 6 Å.

Si establecemos un plano comparativo entre los dos acoplamientos realizados, observamos (según los datos obtenidos a partir de Fpocket y Pocketquery) que las energías libres de ruptura por residuo y en valor absoluto, en promedio son más elevadas en el segundo acoplamiento (entre los dominios SAM y BTB), dándonos a entender que se tendría que imprimir más fuerza para romper estas interacciones, llegando a la conclusión que estos dos dominios se unen de forma más estable que en el primer caso (entre SAM y PNT). Reforzando esta idea, encontramos además un mayor número de enlaces de hidrógeno en el segundo análisis, enlaces que proporcionan estabilidad al complejo.

LINT se une cerca de los TTS.

Según ha aportado Meier et al., el complejo LINT se une a una distancia cercana de los sitios de inicio de transcripción (TTS), como así lo sugiere un mapeo de una región de 250 pb rodeando los TTS, llegando incluso, en muchas ocasiones, a superponerse la proteína dLint-1 sobre dichos TTS.

Por otro lado, encontramos que PSQ, se une a elementos de respuesta a Pc-G (PRE), sitios de unión de proteínas perteneciente al grupo Polycomb (Huang et al. 2002). Más concretamente, la isoforma A de este gen, que cuenta con una longitud de 1046 aminoácidos, se une a motivos (GA)_n, un motivo de unión crítico dentro de los PRE en muchos complejos como CHRASCH (*“chromatin-associated silencing complex for homeotics”*), como así lo demuestran ensayos de competición con los fragmentos bxd-3 y bxd-4 (fragmentos de B-151, el cual posee una longitud de 440 pb, y cuenta con haber la secuencia GAGA) (Huang et al. 2002). Esto que a priori puede hacer tambalear la credibilidad de la hipótesis planteada en este trabajo (ya que LINT se une cerca de TTS de muchos genes), produce el efecto antagónico, reforzándola, ya que se tiene constancia de que muchos de estos elementos PRE se encuentran en puntos cercanos a TTS (Ringrose et al. 2003).

Debido a esto último, de considerarse que existen las relaciones planteadas con anterioridad (El dominio PNT de dL(3)MBT se une SCM mediante su dominio SAM, y que éste a su vez se une con el dominio BTB de PSQ-A), una posible unión del complejo LINT al DNA sería a través del motivo (GA)_n (o en su defecto, algún análogo), unión promovida por PSQ-A, cerca de los sitios de inicio de transcripción (TTS).

LINT reprime por un mecanismo independiente de Grupos Polycomb.

Las proteínas del grupo Polycomb (PcG), son reguladores transcripcionales que mantienen reprimidos ciertos genes reguladores del desarrollo, en células donde estos genes deberían permanecer inactivos. En *Drosophila*, se han caracterizado a día de hoy hasta 18 diferentes proteínas (incluyendo a SCM), que forman diferentes complejos como, PRC1 (Complejo Represivo Polycomb 1), PRC2 (Complejo Represivo Polycomb 2), PhoRC (Complejo Represivo Pho) y PR-DUB, (desubiquitinasa represora Polycomb) entre otros. (Shao et al. 1999; Czermin et al. 2002; Klymenko et al. 2006; Scheuermann et al. 2010).

En cuanto a su modo de acción, cuentan con actividad de unión a DNA (Huang et al. 2002), así como actividad modificadora de histonas, encontrando entre las modificaciones más comunes metilaciones y deacetilaciones (Chang et al. 2001). Una de las propiedades más interesantes en la modificación de la

cromatina, la encontramos bajo ciertas circunstancias, en la propagación en “cis” independientemente de la secuencia (Gelbart & Kuroda 2009). La existencia de dicha capacidad puede ser relevante para la herencia de los estados de cromatina establecidos en su inicio a través de interacciones en sitios de nucleación como PRE. Por lo tanto, la polimerización mediada por el dominio SAM (en la figura 6789 el dominio SAM de Ph, un componente de PRC1, puede ser el responsable de la polimerización), es un modelo apetecible para exponer la propagación del silenciamiento por PcG (Kang et al. 2015).

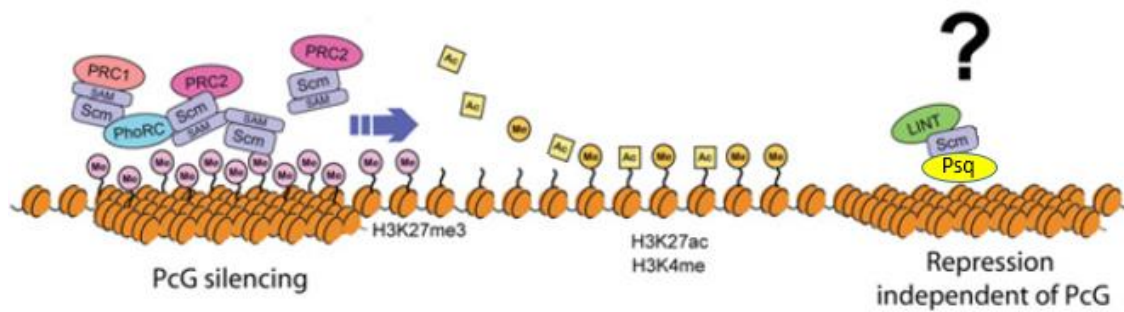


Figura 16. Representación del silenciamiento promovido por PcG (a la izquierda) y el mecanismo de silenciamiento propuesto en este trabajo (a la derecha).

Por el contrario, el mecanismo aquí propuesto postula que es independiente del silenciamiento por complejos Polycomb, basándose en dos argumentos: 1) dL(3)MBT (componente de LINT), se co-localiza con proteínas aislantes, como por ejemplo CP190 y mod (mdg4), en lugar de co-localizarse con proteínas PcG (a excepción de SCM) (Richter et al. 2011). Sin embargo, siendo plenamente consciente de que el silenciamiento PcG requiere de interacciones dinámicas durante las fases de desarrollo, no puedo descartar la posibilidad de que el complejo LINT intervenga de forma transitoria en dicho silenciamiento, situación que explicaría los resultados obtenidos en el análisis bioinformático, donde encontramos una menor estabilidad en el acoplamiento entre los dominios PNT y SAM (de dL(3)MBT y SCM respectivamente), con respecto al acoplamiento entre los dominios SAM y BTB (de SCM y PSQ-A respectivamente); y 2) para observar el grado de implicación de las modificaciones de histonas en el sistema promovido por LINT, se agotaron con RNAi, subunidades de diversos complejos modificadores de histonas cuyo papel se atribuye a represión transcripcional (incluyendo PcG). No se observó la desrepresión de ciertos genes diana de LINT, en cambio al mostrar una deficiencia de cualquier subunidad del complejo LINT, se observó una desrepresión de los mismos (Meier et al. 2012).

Conclusiones

Conclusiones generales del trabajo:

Se ha realizado un “*screening*” con el objetivo de encontrar genes que bajo la regulación de los genes *pipsqueak* y/o *escargot* contribuyan a distintos aspectos de los procesos de invasión y metástasis tumoral. Hemos testado un número total de 800 genes, encontrando respuesta positiva alguna en 51 genes, y corroborada en 14 de esos genes (en este momento concreto), inhibición de los cuales, favorece la deslocalización, invasividad celular y apoptosis, factores intrínsecos en la enfermedad del cáncer.

Conclusiones específicas del trabajo:

Se ha postulado una hipótesis acerca de cuál es el mecanismo que sigue el gen *dL(3)mbt* (uno de los genes testados y corroborados), para desencadenar el fenotipo mostrado en nuestras observaciones, es decir, un patrón anormal de *dpp* junto con invasión celular. La hipótesis establecida es: El complejo LINT, cuyas subunidades son *dLint-1*, *CoRest* y *dL(3)MBT*, mantiene una interacción a través del dominio PNT/SAM de *dL(3)MBT* y el dominio SAM de *SCM*, y que éste último forma una interacción con el dominio BTB de *PSQ-A*, estableciendo un modo de represión transcripcional independiente de complejos de proteínas del grupo Polycomb, o al menos de forma parcial, no dejando cabida en este mecanismo de silenciamiento a la modificación de histonas, actuando por tanto, mediante la ocupación de estos partícipes cerca de los sitios de inicio transcripcional (TTS) de los genes diana (encontrándose entre ellos genes de la línea germinal), posiblemente en motivos de unión (GA)_n a los que se une *PSQ-A*. En células tumorales, este mecanismo estaría alterado, resultando en la desrepresión de una serie de genes, que darían lugar al establecimiento del tumor, así como promulgar una serie de eventos de invasión celular.

Perspectivas futuras

Debido a que una investigación de estas características asume mucho tiempo tanto a nivel experimental en el laboratorio, como a nivel “*in silico*”, por lo que llegado este momento, debemos continuar con una serie de “frentes abiertos” que aquí se han establecido.

En primer lugar, debemos clasificar de forma experimental que los genes confirmados en nuestro screen, cooperan bien con *pipsqueak* o bien con *escargot*. Esto lo podemos determinar gracias a la utilización de los controles positivos con genotipo *GS(2)101; dpp-Gal4 UAS GFP/ RNAi* para el gen *escargot*, o bien con *GS(2)412; dpp-Gal4 UAS GFP/ RNAi* para *pipsqueak*. Así, podemos corroborar de forma tangible la predicción bioinformática con *EsyN*, cuyo interactoma nos sirvió como punto de referencia para establecer las diversas conexiones comentadas en el trabajo.

Por otro lado, una vez somos conocedores de la acción sinérgica de diversos genes, una propuesta interesante sería establecer dobles o triples mutantes silenciados por *RNAi* y observar el fenotipo resultante, dilucidando la magnitud de esos efectos cooperativos entre los genes seleccionados.

Como bien resaltado ha quedado a lo largo de todo el trabajo, estamos trabajando con líneas RNAi, esto es, con líneas que poseen reprimidos un gen, el gen de estudio particular. La pregunta obvia que nos podemos formular es: ¿Qué pasaría si en vez de inhibir al gen, lo sobreexpresáramos? Alguna persona de pensamiento simplista podría opinar que lo lógico sería esperar un efecto antagónico al constatado en esta investigación, pero si algo podemos apreciar de la realidad, es que no vaga de forma unilateral, ofreciéndonos sorpresas a cada paso realizado, por lo que realizar la sobreexpresión de los genes candidatos sería una buena perspectiva de trabajo futura.

Bibliografía

1. **Powell, J.R., et al. (1997).** "Progress and Prospects in Evolutionary Biology: the Drosophila Model." Oxford University Press, New York.
2. **Kornberg, T.B., et al. (2000).** "The Drosophila Genome Sequence: Implications for Biology and Medicine." Science.
3. **Gateff, E. & Schneiderman, H.A. (1967).** "Developmental studies of a new mutant of Drosophila Melanogaster. Lethal malignant brain tumor (l(2)gl 4)." Am. Zool.
4. **Gateff, E. (1978).** "Malignant neoplasms of genetic origin in Drosophila Melanogaster." Science.
5. **Bridges et al. (1916).** "Non-Disjunction as Proof of the Chromosome Theory of Heredity." Genetics.
6. **Artavanis-Tsakonas, S. et al. (1999).** "Notch signaling: cell fate control and signal integration in development." Science.
7. **Taipale, J. et al. (2001).** "The hedgehog and Wnt signalling pathways in cáncer." Nature.
8. **Wang, L., et al. (2014).** "Drosophila as a model to study the role of blood cells in inflammation, innate immunity and cancer." Frontiers in Cellular and Infection Microbiology 3.
9. **Martin-Belmonte, F. et al. (2011).** "Epithelial cell polarity, stem cells and cáncer." Nat. Rev. Cancer.
10. **Levayer, R. & Moreno, E. (2013).** "Mechanisms of cell competition: Themes and variations." JCB Home.
11. **Fan, Y. & Bergmann, A. (2008).** "Apoptosis-induced compensatory proliferation. The cell is dead. Long live the Cell!" Trends Cell Biol.
12. **Previs, R. A., et al. (2014).** "Molecular Pathways: Translational and Therapeutic Implications of the Notch Signaling Pathway in Cancer." Clinical Cancer Research.
13. **Ntziachristos, P., et al. (2014).** "From Fly Wings to Targeted Cancer Therapies: A Centennial for Notch Signaling." Cancer Cell.
14. **Ferres-Marco, D., et al. (2006).** "Epigenetic silencers and Notch collaborate to promote malignant tumours by Rb silencing." Nature.
15. **Toba, G., et al. (1999).** "The gene search system. A method for efficient detection and rapid molecular identification of genes in Drosophila melanogaster." Genetics.
16. **Brand, A.H. & Perrimon, N. (1993).** "Targeted gene expression as a means of altering cell fates and generating dominant phenotypes". Development.
17. **Reece, R.J. & Ptashne, M. (1993).** "Determinants of binding-site specificity among yeast C6 zinc cluster proteins." Science.
18. **Domínguez, M. & de Celis, J.F. (1998).** "A dorsal/ventral boundary established by Notch controls growth and polarity in the Drosophila eye." Nature.
19. **Martín, F. & Morata, G. (2009).** "Cell competition growth and size control in the Drosophila wing imaginal disc." Development.
20. **García-Bellido, A., et al. (2006).** "Eversión y cierre de los discos imaginales." Investigación Y Ciencia.

21. Beira, J.V. & Paro, R. (2016). "The legacy of Drosophila imaginal discs." Chromosoma.
22. Goeke, S. et al. (2003). "Alternative splicing of lola generates 19 transcription factors controlling axon guidance in Drosophila." Nature Neurosci.
23. Bardwell, V.J. & Treisman, R. (1994). "The POZ domain: a conserved protein-protein interaction motif." Genes. Dev.
24. Siegmund, T. & Lehmann, M. (2002). "The Drosophila Pipsqueak protein defines a new family of hélix-turn-helix DNA-binding proteins." Dev. Genes Evol.
25. Schwendemann, A. & Lehmann, M. (2002). "Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci" Proc. Natl Acad. Sci. USA.
26. Weber, U. et al. (1995) "pipsqueak encodes a novel nuclear protein required downstream of seven-up for the development of photoreceptors R3 and R4" EMBO J.
27. Ringrose, L. & Paro, R. (2004). "Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins" Annu. Rev. Genet.
28. Acloque, H. et al. (2009). "Epithelial-Mesenchymal Transitions: The Importance of Changing Cell State in Development and Disease." The Journal of Clinical Investigation.
29. Nisticò, P. et al. (2012). "Epithelial-mesenchymal transition: general principles and pathological relevance with special emphasis on the role of matrix metalloproteinases." Cold Spring Harb Perspect. Biol.
30. Yang, M.H. et al. 2017. "Revisiting epithelial-mesenchymal transition in cancer metastasis: the connection between epithelial plasticity and stemness." Mol Oncol.
31. Lamouille, S. et al. (2014). "Molecular mechanisms of epithelial-mesenchymal transition." Nat. Rev. Mol. Cell Biol.
32. Webster, L.C. et al. (1991). "Conversion of the E1A Cys4 zinc finger to a nonfunctional His2,Cys2 zinc finger by a single point mutation." Proc. Natl. Acad. Sci. USA.
33. Hayashi, S. et al. (1993). "Control of imaginal cell development by the escargot gene of Drosophila." Development.
34. Peinado, H. et al. (2007). "Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype?" Nat. Rev. Cancer.
35. Abrams, J.M. et al. (1999). "An emerging blueprint for apoptosis in Drosophila." Trends in Cell Biology.
36. Farkas, G. et al. (1994). "The Trithorax-like gene encodes the Drosophila GAGA factor." Nature.
37. Wu, Y. et al. (1996). "Quantitation of putative activator-target affinities predicts transcriptional activating potentials." The EMBO Journal.
38. Carrozza, M.J. et al. (2002). "Gal80 confers specificity on HAT complex interactions with activators." J Biol. Chem.
39. Zenke, F.T. et al. (1996). "Activation of Gal4p by galactose-dependent interaction of galactokinase and Gal80p." Science.
40. Leuther, K.K. & Johnston, S.A. (1992) "Nondissociation of GAL4 and GAL80 in vivo after galactose induction." Science.
41. Bhaumik, S.R. et al. (2004). "In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer." Genes Dev.
42. Peng, G. & Hopper, J.E. (2000). "Evidence for Gal3p's cytoplasmic location and Gal80p's dual cytoplasmic-nuclear location implicates new mechanisms for controlling Gal4p activity in Saccharomyces cerevisiae." Mol. Cell Biol.
43. Traven, A., et al. (2006). "Yeast Gal4: a transcriptional paradigm revisited." EMBO reports
44. Dietzl, G. et al. (2007). "A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila." Nature.
45. Horn, T. & Boutros, M. (2010). "E-RNAi: a web application for the multi-species design of RNAi reagents—2010 update." Nucleic Acids Res.

46. Groth, A.C. et al. (2004). "Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31." Genetics.
47. King, T. & Reiss, M. (2001). "Practical Advanced Biology," Nelson Thornes.
48. Lawrence, P.A. (1992). "The making of a fly: the genetics of animal design." Blackwell Science.
49. Sullivan, M. et al. (2000). "Drosophila Protocols." Cold Spring Harbor Laboratory Press.
50. Baker, D. et al. (2001). "Protein structure prediction and structural genomics." Science.
51. Fiser, A. (2010) "Template-based protein structure modeling." Methods Mol. Biol.
52. Bordoli, L. et al. (2012). "Automated Protein Structure Modeling with SWISS-MODEL Workspace and the Protein Model Portal" Methods in molecular biology.
- Battey, J.N. et al. (2007). "Automated server predictions in CASP7." Proteins.
53. Koh, I.Y. et al. (2003). "EVA: Evaluation of protein structure predictions server." Nucleic Acids Res.
54. Zhang, Y. & Skolnick, J. (2005). "The protein structure prediction problema could be solved using the current PDB library." Proc. Natl Acad. Sci. USA.
55. Peitsch, M.C. (1995). "Protein modelling by E-Mail." BioTechnology.
56. Van Gunsteren, W.F. et al. (1995). "Biomolecular Simulations: The GROMOS96 Manual and User Guide." Zürich.
57. Benkert, P. et al. (2009). "QMEAN server for protein model quality estimation." Nucleic Acids Res.
58. Bean, D.M. et al. (2014). "EsyN: Network Building, Sharing and Publishing." PLoS ONE.
59. Inbar, Y. et al. (2005). "Approaching the CAPRI challenge with efficient geometry based docking." Proteins.
60. Conolly, M.L. (1983). "Analytical molecular Surface calculation." J. Appl. Crystallogr.
61. Schneidman-Duhovny, D. et al. (2005). "PatchDock and SymmDock: servers for rigid and symmetric docking." Nucleic Acids Research.
62. Le Guilloux, V. et al. (2009). "Fpocket: An open source platform for ligand pocket detection." BMC Bioinformatics.
63. Tina, K.G. et al. (2007). "PIC: Protein Interactions Calculator." Nucleic Acids Research.
64. Whale, E. et al. (1992). BioEssays.
65. Wismar, J. et al. (1995). "The Drosophila melanogaster tumor suppressor gene lethal(3)malignant brain tumor encodes a proline-rich protein with a novel zinc finger." Mech. Dev.
66. Richter, C. et al. (2011). "L(3)mbt inhibits neuroepithelial proliferation and acts on insulator elements" Nature Cell Biology.
67. Janic, A. et al. (2011). "Ectopic expression of germline genes drives malignant brain tumor growth in Drosophila" Science.
68. Lewis, P.W. et al. (2004). "Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex." Genes Dev.
69. Trojer, P. et al. (2007). "L3MBTL1, a histone-methylation-dependent chromatin lock." Oncogene.
70. Meier, K. et al. (2012). "LINT, a novel dL(3)mbt-containing complex, represses malignant brain tumour signature genes." PLoS Genet.
71. Peterson, A.J. et al. (2004). "Requirement for sex comb on midleg protein interactions in Drosophila polycomb group repression." Genetics.
72. Artyom, A. et al. (2016). "BioTAP-XL - crosslinking/tandem affinity purification to study DNA targets, RNA and protein components of chromatin associated complexes." Curr. Protoc. Mol. Biol.
73. Bornemann, D. et al. (1998). "Expression and Properties of Wild-Type and Mutant Forms of the Drosophila Sex Comb on Midleg (SCM) Repressor Protein." Genetics.

74. Peterson, A.J. et al. (1997). "A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions." Mol Cell Biol.
75. Torchala, M. et al. (2013). "SwarmDock: a server for flexible protein-protein docking." Bioinformatics.
76. Frey, F. et al. (2016). "Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC." Genes Dev.
77. Farid, K. et al. (1998). "Crystal structure of the BTB domain from PLZF." Proc. Natl. Acad. Sci.
78. Huang, D. et al. (2002). "pipsqueak Encodes a Factor Essential for Sequence-Specific Targeting of a Polycomb Group Protein Complex." Mol. Cell. Biol.
79. Ringrose, L et al. (2003). Genome-Wide Prediction of Polycomb/Trithorax Response Elements in Drosophila melanogaster." Developmental Cell.
80. Shao, Z. et al. (1999). "Stabilization of chromatin structure by PRC1, a Polycomb complex." Cell.
81. Czermin, B. et al. (2002). "Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites." Cell.
82. Klymenko, T. et al. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities." Genes Dev.
83. Scheuermann, J.C. et al. (2010). "Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB." Nature.
84. Gelbart, M.E. & Kuroda, M.I. (2009). "Drosophila dosage compensation: a complex voyage to the X chromosome." Development.
85. Kozakov, D. et al. (2017). "The ClusPro web server for protein-protein docking." Nat. Protoc.
86. Chang, Y. et al. (2001). "Essential role of Drosophila Hdac1 in homeotic gene silencing." Natl. Acad. Sci. Usa.
87. Kang, H. et al. (2015). "Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in Drosophila." Genes Dev.

Tesis doctorales:

M^a Dolores Ferrés Marcó. (2010). "IDENTIFICACIÓN DE GENES Y MECANISMOS CAUSATIVOS DE CÁNCER Y METÁSTASIS."

Antonio José Montes Ruiz. (2016). "Control del crecimiento en los discos imaginales de Drosophila Melanogaster."

